

Wasserstein HOG: Local Directionality Extraction via Optimal Transport

Jiening Zhu, Harini Veeraraghavan, Jue Jiang, Jung Hun Oh,
Larry Norton, Joseph O. Deasy, *Member, IEEE*, Allen Tannenbaum, *Fellow, IEEE*

Abstract—Directionally sensitive radiomic features including the histogram of oriented gradient (HOG) have been shown to provide objective and quantitative measures for predicting disease outcomes in multiple cancers. However, radiomic features are sensitive to imaging variabilities including acquisition differences, imaging artifacts and noise, making them impractical for using in the clinic to inform patient care. We treat the problem of extracting robust local directionality features by mapping via optimal transport a given local image patch to an iso-intense patch of its mean. We decompose the transport map into sub-work costs each transporting in different directions. To test our approach, we evaluated the ability of the proposed approach to quantify tumor heterogeneity from magnetic resonance imaging (MRI) scans of brain glioblastoma multiforme, computed tomography (CT) scans of head and neck squamous cell carcinoma as well as longitudinal CT scans in lung cancer patients treated with immunotherapy. By considering the entropy difference of the extracted local directionality within tumor regions, we found that patients with higher entropy in their images, had significantly worse overall survival for all three datasets, which indicates that tumors that have images exhibiting flows in many directions may be more malignant. This may seem to reflect high tumor histologic grade or disorganization. Furthermore, by comparing the changes in entropy longitudinally using two imaging time points, we found patients with reduction in entropy from baseline CT are associated with longer overall survival (hazard ratio = 1.95, 95% confidence interval of 1.4-2.8, $p = 1.65e-5$). The proposed method provides a robust, training free approach to quantify the local directionality contained in images.

Index Terms—Optimal transport, MRI, CT, radiomic features, imaging processing

I. INTRODUCTION

Computerized feature extraction has remained a key topic ever since image data was first digitalized. Local directionality is an important feature in image processing. Over the past

This study was supported by AFOSR grants (FA9550-20-1-0029, FA9550-23-1-0096), Army Research Office grant (W911NF-22-1-0292), NIH grant (R01-AG048769), a grant from the Breast Cancer Research Foundation (BCRF-17-193), a grant from the Cure Alzheimer Foundation, and the MSK Cancer Center Support Grant (P30 CA008748).

J. Zhu and A. Tannenbaum are with the Department of Applied Mathematics & Statistics, Stony Brook University, NY; Tannenbaum is with the Department of Computer Science, Stony Brook University, NY; email: {jiening.zhu, allen.tannenbaum}@stonybrook.edu

H. Veeraraghavan, J. Jiang, J.H. Oh, and J. Deasy are with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY; email: {veerarah.jiangj1, ohj, nortonl, deasyj}@mskcc.org

L. Norton is with the Department of Medicine, Memorial Sloan Kettering Cancer Center, NY; email: nortonl@mskcc.org

several decades, multiple different image feature descriptors have been proposed to extract such information. Steerable filters are synthesized filters of arbitrary orientations from linear combinations of basis filters [1]. Scale-Invariant Feature Transform (SIFT) and Histogram of Orientated Gradient (HOG) are two feature descriptors based on histogram of gradients [2], [3]. Image feature describing directionality are widely used for object detection, object matching and network pretraining [4]–[6].

More specifically, many directional related features are used in radiomics analysis. The Gabor filter is a linear filter based on the Gaussian kernel and sinusoidal plane wave. It gives an orientation based representation of the image. Wavelets are time and frequency localized functions that can be used to decompose signals into different scales or resolutions, which may also represent directionality [7]. The Gray-Level Co-Occurrence Matrix (GLCM) is a method used to calculate the number of occurrences of neighboring pixels with respect to each gray level value [8]. It utilizes two parameters: the distance between the pixels and the angle at which the co-occurrences are counted.

The Wasserstein distance is a powerful metric for distributions, or more generally measures, which is known to be robust compared to other metrics (or divergences) given its property of stability to small perturbations in the input data (via weak continuity). It is naturally defined by the optimal cost of transporting one distribution to another, which was first motivated by the civil engineering problem of relocating a pile of soil to an excavation site by Gaspard Monge in 1781 [9]–[12]. A relaxation was proposed by the Russian mathematician Leonid Kantorovich in 1942 [13], and so the optimal transport problem is many times called the *Monge-Kantorovich problem*. It gives a distance of the space of probability distributions.

Optimal mass transport methods are widely used in signal processing, machine learning, computer vision, meteorology, statistical physics, quantum mechanics, and network theory [14]–[19]. Some examples include Zhu et al. and Pouryahya et al. employed distances derived from the optimal mass transport theory to study multiomics networks for cancer subtype clustering [20], [21]. A regularized version of optimal transport is utilized to visualize fluid flows in the glymphatic system [22], as well as an unbalanced version of the Wasserstein distance is utilized to identify high-risk normal tissue regions associated with worse mortality from spillover radiation in radiation treatments [23].

In this work, we make use of the robust property of the

Wasserstein distance for image local analysis in order to extract directional information from the optimal transport map instead of gradient. The optimal transport based computation increases robustness to noise, making it preferable for a variety of medical image analysis applications. In the following sections, we sketch some of the background on commonly used optimal transport models. Then we first detail our proposed methodology in terms of the continuous case in general and then introduce the discrete case with an additional multi-scale integration step, which we specifically use for image feature extraction. Quantitative comparison and illustrative examples of our method are included in the Experiments section. We conclude this note with some discussion about future work on this topic. The contributions of this work include:

- Novel features characterizing local directionality are developed, using an optimal mass transport methodology, which is robust to local image variations and reconstructions required for reproducible and reliable radiomic analysis.
- Our multi-scale recombination which utilize the unique structure of optimal transport further increases the robustness of the method and is able to extract directional features at different scales.
- The parallelized implementation allows for faster computation of the feature calculations.
- We demonstrate ability of our features to prognosticate patient outcomes using magnetic resonance imaging (MRI) and computed tomography (CT) images for brain, head & neck and lung cancers, respectively.

II. BACKGROUND

Optimal transport was originally formulated by Gaspar Monge in 1781 and given a relaxed formulation by Leonid Kantorovich in 1942, and hence is known as the *Monge-Kantorovich problem* [9], [10]. We will use the Kantorovich version in this paper which may expressed as follows:

$$\mathcal{W}_K(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^n} c(x, y) d\pi(x, y), \quad (1)$$

where μ_0, μ_1 are two absolutely continuous measures on \mathbb{R}^n , $\Pi(\mu_0, \mu_1)$ denotes the set of all the couplings between μ_0 and μ_1 (measures on $\mathbb{R}^n \times \mathbb{R}^n$ whose two marginals are μ_0 and μ_1 : $\pi(\cdot \times \mathbb{R}^n) = \mu_0(\cdot), \pi(\mathbb{R}^n \times \cdot) = \mu_1(\cdot)$). In the present work, we will take $c(x, y) = \|x - y\|$. More generally, one may employ any lower semi-continuous function such that $c(x, y) \geq a(x) + b(y)$ for all x, y , where $a \in L^1(\mu_0), b \in L^1(\mu_1)$. In terms of discrete probability densities, the Kantorovich model may be expressed as a standard linear programming problem, which may be solved very efficiently. The discrete form of the Kantorovich optimal transport may be formulated as follows:

$$d_K(\rho_0, \rho_1) = \min_{\Pi \in \mathbb{R}^{M \times N}} \sum_{i=1}^M \sum_{j=1}^N C(i, j) \cdot \Pi(i, j) \quad (2a)$$

$$\text{subject to: } \Pi \vec{\mathbf{1}}_N = \rho_0, \quad (2b)$$

$$\Pi^T \vec{\mathbf{1}}_M = \rho_1, \quad (2c)$$

where $\rho_0 \in \mathbb{R}_+^M$ and $\rho_1 \in \mathbb{R}_+^N$ are densities of M/N discrete locations of initial/target distributions and their total

mass needs to be preserved as a prerequisite ($\sum_{i=1}^M \rho_0(i) = \sum_{j=1}^N \rho_1(j)$). C is a point to point cost matrix and $\vec{\mathbf{1}}_N(\vec{\mathbf{1}}_M)$ is an all-1 vector of length $N(M)$.

III. METHODS

A. Continuous case

In the Monge-Kantorovich problem, Π is a generalized transport map between two distributions [9], [10], which specifies the amount of mass that is transported from location x in the initial distribution to location y in the target distribution, from which one derives a transport direction. To extract the local directionality, Kantorovich optimal transport between the original distribution and a smoothed one on a given local image patch \mathcal{X} is employed. Namely, we consider

$$\mathcal{W}_{K, \mathcal{X}}^p(\rho) = \inf_{\Pi} \int_{\mathcal{X} \times \mathcal{X}} c(x, y) \cdot \Pi(x, y) dx dy, \quad (3a)$$

$$\text{subject to: } \int_{\mathcal{X}} \Pi(x, y) dy = \rho(x) \mathbf{1}_{\mathcal{X}}(x), \quad (3b)$$

$$\int_{\mathcal{X}} \Pi(x, y) dx = \mathcal{K}(\rho \mathbf{1}_{\mathcal{X}})(y), \quad (3c)$$

where $\mathbf{1}_{\mathcal{X}}$ is the indicator function, and \mathcal{K} denotes any mass preserving smoothing filter on \mathcal{X} ,

$$\int_{\mathcal{X}} \rho(x) dx = \int_{\mathcal{X}} \mathcal{K}(\rho)(y) dy. \quad (4)$$

In the present work we will employ the following simple filter:

$$\mathcal{K}(\rho)(y) = \frac{\int_{\mathcal{X}} \rho(x) dx}{|\mathcal{X}|} = \text{constant}. \quad (5)$$

By considering the transport distance (3), we can quantitatively measure how different the patch is from its feature removal/smoothed version. Notice that the optimal transport also gives an optimal transport map Π , from which one derives the probability $\Pi(x, y)$ of each point $x \in \mathcal{X}$ to be transported to any other point $y \in \mathcal{X}$. The displacement vector $y - x$ together with the probability gives a natural directionality at each point. By collecting all the sub-transport work in the same direction, we can define a distribution ρ_{Θ} on \mathcal{S}^n in terms of directionality of the patch \mathcal{X} as follows:

$$\mathcal{W}_{\Theta, \mathcal{X}}^p(\theta) = \int_0^{\infty} \int_{\mathcal{X}} \lambda \cdot \Pi(x, x + \lambda\theta) dx d\lambda, \quad (6)$$

which is a decomposition of the total transport work into all possible directions.

Proposition 1: $\mathcal{W}_{\Theta, \mathcal{X}}(\theta)$ is decomposition of the total transport in (3):

$$\mathcal{W}_{K, \mathcal{X}}(\rho) = \int_{\mathcal{S}^n} \mathcal{W}_{\Theta, \mathcal{X}}^p(\theta) d\theta, \quad (7)$$

Proof:

$$\begin{aligned} & \int_{\mathcal{S}^n} \mathcal{W}_{\Theta, \mathcal{X}}(\theta) d\theta \\ &= \int_{\mathcal{S}^n} \int_{\{(x, y) | (x, y) \parallel \theta\}} \|x - y\| \cdot \Pi(x, y) (dx dy) d\theta \\ &= \int_{\mathcal{X}} c(x, y) \cdot \Pi(x, y) dx dy = \mathcal{W}_{K, \mathcal{X}}(\rho). \end{aligned}$$

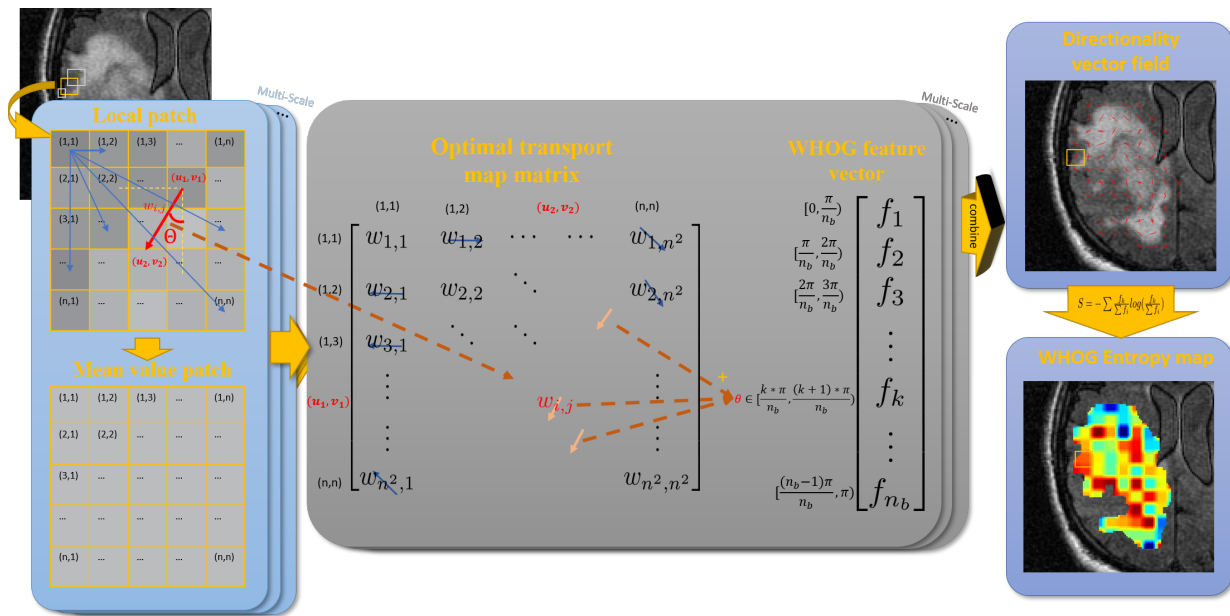


Fig. 1: An illustration of the proposed WHOG method. Local patches of varying sizes and positions are extracted from within the tumor region of an image. For each one, the optimal transport problem of transporting the local patch to a mean value vector is computed. The resulting optimal transport matrix gives a map matrix, each entry of which represents a sub-transport work value with a clearly defined direction. The values of sub-transport work in the matrix are decomposed into an n_b binned vector, which represents the directionality of the patch. Multiscale feature recombination is employed to gather information from different scales, and directionality vectors are used to compute entropy on each patch. Finally, the output WHOG feature is the average entropy of all the patches within the tumor region.

The latter result indicates that $\mathcal{W}_{K,\mathcal{X}}$ decomposes the transport distance into different directions.

B. Discrete case

For application of our directionality distribution, we employ a discrete version of (6) to extract the local directionality of images. We choose to use the same output format as the HOG because they extract the same type of local information. Because of the similarity, we call our method *Wasserstein HOG* (WHOG). Note that we only borrow the name HOG because of the similar information extracted. Our method does not depend on the gradient. We consider the following discrete version of Kantorovich optimal transport:

$$d_{K,\mathcal{X}_n}(\rho_0) = \min_{\Pi \in \mathbb{R}^{n^2 \times n^2}} \sum_{i=1}^{n^2} \sum_{j=1}^{n^2} C(i,j) \cdot \Pi(i,j) \quad (8a)$$

$$\text{subject to: } \Pi \vec{\mathbf{1}}_{n^2} = \rho_0, \quad (8b)$$

$$\Pi^T \vec{\mathbf{1}}_{n^2} = \rho_1 = \frac{\sum \rho_0}{n^2} \vec{\mathbf{1}}, \quad (8c)$$

where \mathcal{X}_n is a local $n \times n$ image patch and $\rho_0 \in \mathbb{R}_+^{n^2}$ is the flattened image intensity on that patch. An iso-intense patch with the mean value of ρ_0 as intensity of all pixels on that patch is used as $\rho_1 \in \mathbb{R}_+^{n^2}$. Further, the cost matrix $C \in \mathbb{R}^{n^2 \times n^2}$ is taken as the standard pairwise distance matrix.

Local image directions are extracted from the matrix $W = C \dots \Pi^*$ by regrouping the entries into bins where Π^* is the

- optimal solution to the optimization (Fig. 1). Each entry of W has a corresponding orientation given by the transport starting point and end point locations. For the (i,j) entry of W ($i \neq j$), the i^{th} location in the density vector corresponds to a coordinate (u_1, v_1) in the original $n \times n$ grid and the j^{th} location in the density vector corresponds to another coordinate (u_2, v_2) in the original $n \times n$ grid. The *directionality* of the sub-transport route is given as:

$$\theta = \arctan\left(\frac{v_1 - v_2}{u_1 - u_2}\right). \quad (9)$$

We evenly divide $[0, \pi]$ into n_b bins and add up all the w_{ij} values in their corresponding bins. So each bin contains all the sub-work of the optimal transport in its corresponding direction. The n_b -vector is a representation of local directionality. The sum of n_b values in the feature vector coincides with the sum of all w_{ij} , which is the Wasserstein distance (total work) by the same reasoning as Proposition 1. On the other hand, we decompose the distance into components in the n_b directions. This gives a similar feature vector that characterizes local directionality as in HOG, but is more robust. From a different perspective, optimal transport has been used in registration [16]. The proposed feature may be understood as a summary of the directions of corresponding points of a generalized local registration from each local patch to an iso-intense patch, where the "registration" based on the optimal transport method (8) is not 1-to-1, but 1-to-n.

C. Discrete case with multi-scale recombination

Multi-scale is a commonly used technique in image processing. In this current work, we find it extremely useful to use different patch sizes. To deal with the problem of aligning patches of various sizes and centered in different locations, we first extract directionality vectors on the level of pixel and then recombine together, making a minor change from the method of the last section (III-B), which has the same form of output as before. But instead of using one fixed patch size n , we use multiple patch sizes and multiple patch starting positions: $(n_1, p_1), (n_2, p_2), \dots, (n_i, p_i), \dots, (n_k, p_k)$, where $n_i \in \mathbb{Z}$ is a patch size and $p_i \in \{1, 2, \dots, n_i\} \times \{1, 2, \dots, n_i\} \subset \mathbb{Z}^2$ is to shift the grid of all patches. The Kantorovich optimal transport problem in the same form of (8) with different patch parameters is solved. Instead of deriving just one n_b directionality vector from the entire matrix W for that patch, n_i^2 of n_b -vectors for each pixel are extracted from each row of the matrix W . Noting that each row of the matrix W specifies the optimal transport plan for a pixel on that patch, the n_i^2 vectors represent the directionality on all of the n_i^2 pixels of that patch. For that purpose, we further decompose the matrix W in an even finer level. To get the same output format as before, for a given output patch size n , we simply sum all $n^2 \times k$ n_b -vectors on each output patch. So the final output contains information of a neighborhood in the perspective of multi-scale.

D. Implementation of discrete case

Because of the translational invariance of the pairwise distance within a given patch, the cost matrix C is the same for a fixed n_i . It is the pairwise distance matrix of all the lattice points, and it is constructed beforehand to save computational cost. In addition, the n_b mask matrices M_j 's for each directional bin are also constructed. They are also the same for a fixed value of n_i . The main step is solving a linear programming problem (2) for each patch from an input image I . The 2D intensity is vectorized and reformulated as a standard linear programming problem, which is solved via the dual-simplex algorithm [24]. The output solution of the transport matrix is then decomposed into n_b directions using mask matrices, and the pixelwise transport cost is computed by summing each row of the transport matrix, which is stored in the n_b matrices $(I_{PV_i}(:, :, 1), \dots, I_{PV_i}(:, :, n_b))$ of the same size as the original image I . I_{PV_i} 's from different patch parameters (n_i, p_i) sum up to be the multiscale pixelwise directionality matrices I_{PV} . Finally, the features are output as the sum on each output patch. We implemented our proposed method in MATLAB. For the linear programming part, the built-in *linprog* function is used. Parallelization was utilized for speed-up by taking advantage of the independence of computations on different scales and different patches (See Algorithm 1).

Algorithm 1 Wasserstein HOG

Inputs: Gray scale image I , patch size n , number of bins n_b
Parameters: Multi-scale parameters $(n_1, p_1), \dots, (n_k, p_k)$
Outputs: Directionality vector field

- 1: **for** $1, \dots, i, \dots, k$ **do**
- 2: Construct n_i^2 by n_i^2 cost matrix C
- 3: **for** $1, \dots, j, \dots, n_b$ (in parallel) **do**
- 4: Construct an indicator matrix $M_j \in \mathbb{R}^{n_i^2 \times n_b}$ for all transport directions within the range of j 'th directionality bin
- 5: **end for**
- 6: **for** each patch \mathcal{X} in I starting from pixel p_i with step = n_i (in parallel) **do**
- 7: Get intensity ρ on patch \mathcal{X}
- 8: Construct mean value patch ρ_m
- 9: Solve (8) via *linprog* for $\Pi_{\mathcal{X}}$ using C, ρ, ρ_m
- 10: $W_{\mathcal{X}} = \Pi_{\mathcal{X}} * C$
- 11: **for** $1, \dots, j, \dots, n_b$ **do**
- 12: Assign j 'th pixel-wise directional work portion to each pixel on \mathcal{X} :
 $I_{PV_i}(\mathcal{X}, j) \leftarrow \text{sum}(W_{\mathcal{X}} * M_j, 2)$
- 13: **end for**
- 14: **end for**
- 15: **end for**
- 16: $I_{PV} = \sum_{i=1}^k I_{PV_i}$
- 17: **for** each output patch \mathcal{X}_o **do**
- 18: Sum to get final directionality vector $f_{\mathcal{X}_o}$ on patch \mathcal{X}_o :
 $f_{\mathcal{X}_o}(1 : n_b) = \sum_{p \in \mathcal{X}_o} I_{PV}(p, 1 : n_b)$
- 19: **end for**

IV. EXPERIMENTS

A. Comparison with conventional HOG

A brief overview of extracting the standard histogram of oriented gradients (HOG) features is provided for completeness. The first step involves computation of the image gradients, which typically uses the Sobel image filters, described as:

$$[1 \quad 0 \quad -1], \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$$

to compute the x and y coordinate derivatives (g_x, g_y) for each pixel. The directionality of a pixel located at $\{x, y\}$ is computed using the direction of the gradient vector:

$$\theta = \arctan\left(\frac{g_y}{g_x}\right). \quad (10)$$

Second, patch histograms of oriented gradients are computed by binning the pixels within the patch using θ values, with the number of bins corresponding to the chosen discretization for the gradient θ . The sums of gradient vector norms $g = \sqrt{g_x^2 + g_y^2}$ present in all bins are then combined to form a feature vector of that patch. Feature vectors of neighboring patches are normalized together to alleviate artifacts caused by light variations. Patch sizes used to construct the HOG features were determined empirically and was set to 8 by 8.

Of note, the computation of directionality (9) using WHOG is very different from HOG (10). Specifically, the directionality of features using HOG is based on image gradient, which is sensitive to image noise. WHOG, on the other hand, considers constituents of all the possible sub-transport routes traversing the grid points between all pairs of pixels to compute the direction, thus making WHOG more robust to image noise than the HOG approach. The robustness of WHOG to noise is illustrated by the toy examples in Fig. 2, wherein both WHOG and HOG perform similarly well without noise. However, when more noise is added (Gaussian noise with mean= 0, $\sigma = \{0, 0.01, 0.01, 0.1, 0.15\}$), WHOG computed directions show very little change. HOG on the other hand, gradually loses directionality until the histogram is evenly distributed to all the bins with the highest noise levels.

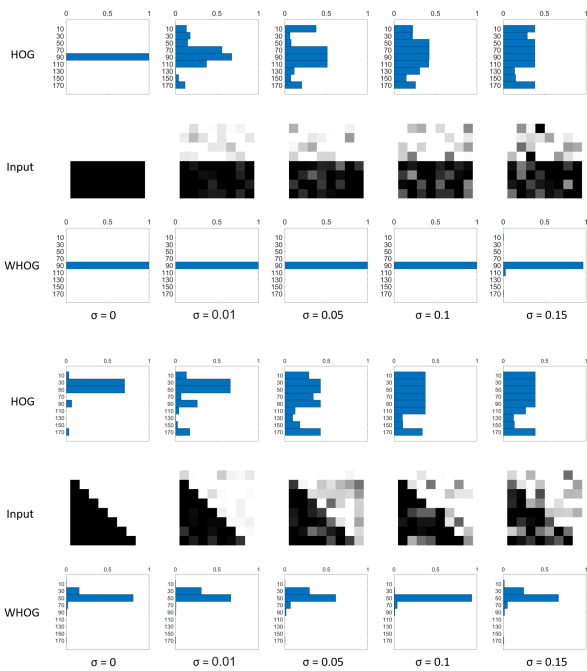


Fig. 2: HOG vs WHOG under different noise levels: for both methods, we use 8 by 8 patch and 9 bins.

The qualitative performance of the two methods for extracting the gradient orientations using rose plots within the tumor and at its boundary (yellow box in Fig. 3) shows that WHOG (Fig. 3b) extracts both the large and small differences in edge orientations with stronger resolution of differences compared to the HOG method (Fig. 3a). Furthermore, clear edge orientations are lost with HOG within the tumor, but continue to be extracted for the WHOG, indicating that WHOG may be better at extracting clear directionality even inside tumors.

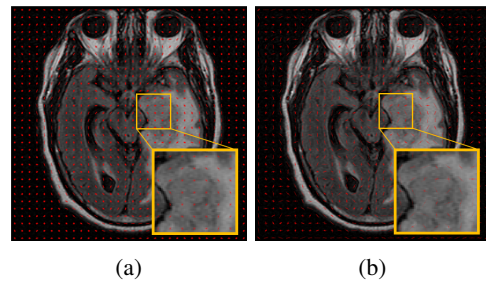


Fig. 3: An example of a brain GBM slice from a TCIA dataset. (a): Rose plot of the HOG with a zoom-in view of the tumor region, (b): Rose plot of the WHOG with a zoom-in view of the same tumor region.

B. Quantifying image tumor heterogeneity using WHOG

Imaging tumor heterogeneity has been shown to be associated with biological heterogeneity by multiple radiomics studies [25]–[27]. In particular, radiomic features constructed with oriented edges from CT as well as MRI scans were shown to be associated with survival and treatment outcomes in solid cancers [28], [29]. Hence, we asked whether the imaging tumor heterogeneity quantified as average entropy of the local feature directionality (described in the subsequent subsection) is associated with survival and treatment outcomes in three different solid cancers, namely glioblastoma multiforme (GBM) in the brain, oro/nasopharynx cancer in the head and neck, and non-small cell lung cancer (NSCLC) using three separate datasets. Two of the three datasets were sourced from the publicly available the Cancer Imaging Archive (TCIA) and previously used in radiomic studies for brain [30], head and neck cancer [6]. The third dataset in NSCLC patients treated with immunotherapy was previously used for evaluating deep learning segmentation of lung tumors [31]. Description of the analyzed datasets and the outcomes are provided in the subsequent subsection.

1) *Entropy formulation*: Because we are only interested in the flows within the tumor, we utilized the tumor masks and considered transport problem only within the tumor region:

$$\tilde{d}_{K, \mathcal{X}_n}(\rho_0) = \min_{\Pi \in \mathbb{R}^{n^2 \times n^2}} \sum_{i=1}^{n^2} \sum_{j=1}^{n^2} C(i, j) \cdot \Pi(i, j) \quad (11a)$$

$$\text{subject to: } \Pi \vec{1}_{n^2} = \rho_0 \cdot * \vec{mask}, \quad (11b)$$

$$\Pi^T \vec{1}_{n^2} = \rho_1 = \frac{\sum \rho_0 \cdot * mask}{\sum mask} \times \vec{mask}, \quad (11c)$$

where both the original patch and the iso-intense patch are taken to be within the tumor region. The following is the same set-up as in the section III-C. By considering the sub-transport route direction derived from (9), a vector $v_p \in \mathbb{R}^{n_b}$ which represents the directionality of patch q is computed using multi-scale recombination. Notice that

$$\sum_{c=1}^{n_b} v_q(c) = \tilde{d}_{K, \mathcal{X}_n}(\rho_0),$$

which gives a measure of the strength of directionality of a given patch, which may not sum up to be 1. The local

directional distribution is defined as the normalized v_q :

$$P_q(c) = \frac{v_q(c)}{\sum_{c=1}^{n_b} v_q(c)}. \quad (12)$$

Finally, entropy is defined as the average of all patches within the tumor region (the patch intersects more than 80% of mask area):

$$S = \frac{1}{|Q|} \sum_{q \in Q} \sum_{c=1}^{n_b} -P_q(c) \log(P_q(c)), \quad (13)$$

where Q is the set of all feasible patch q .

Finally, CoLIAGe [28], a recently developed radiomic measure quantifying imaging heterogeneity using histogram of oriented gradients together with cooccurrence matrix formulation was evaluated as a benchmarking metric to assess the gains achieved using WHOG as a novel radiomic measure of imaging tumor heterogeneity.

We used $n = 8$ and $n_b = 9$ for both WHOG and HOG in the following tests, which is the default setting of HOG from MATLAB. For the latent multi-scale parameters, we used $(4, (1, 1))$, $(4, (3, 3))$, $(8, (1, 1))$ and $(8, (5, 5))$ – four different sets of patches, which are patches starting from origin or the first patch center for both patch sizes equal to 4 or 8.

We used the implementation of CoLIAGe from CERR with all default parameters [32].

C. Patient datasets description

1) *Brain GBM MRI dataset*: This cohort consisted of 50 patients diagnosed with GBM and arose from a multi-institutional open-source dataset sourced from 28 different institutions and provided by the TCIA [33] and was used in our prior work on radiomics analysis [30]. In order to limit the number of analyzed features, we restricted the analysis to fluid attenuation inversion recovery (FLAIR) images provided by the cohort alone as these depict tumors as well as edema outside of the tumor. Tumor features were extracted within segmentations generated using a semi-automatic method and available from prior publication [30].

In addition to the imaging data, TCIA also provides matched overall survival (OS), and relevant clinical data for these patients, which were used in the univariate and multivariable Cox proportional hazards regression model to measure associations of WHOG and HOG features with OS.

2) *Head and neck CT dataset*: The head and neck dataset consisted of 136 patients diagnosed with head and squamous cell carcinoma treated with radiotherapy and was used in a large radiomics analysis paper by Aerts et al. [6] and made available together with the gross tumor volume segmentations in the TCIA. All CT scans were acquired in the MAASTRO clinic, the Netherlands.

The standardized pre-processing steps were applied as described previously in [34]. In order to increase contrast on CT, all image slices were subject to intensity clipping, with the clip range values corresponding to those that covered the mean intensity of the tumor regions. The same clipping upper and lower bounds of 1000 Hounsfield unit (HU) to 1200 HU were applied. The clipping range was set to 200 HU.

Together with the CT scans, volumetric segmentations produced by the radiation oncologist and relevant clinical data such as the disease free survival, OS, biological sex at birth, human papilloma virus (HPV) status, age, and other confounding clinical variables were also provided with the dataset and were used in the univariate and multivariable Cox proportional hazards regression analyses.

3) *Lung immunotherapy longitudinal CT dataset*: 143 stage III-IV NSCLC patients treated with immunotherapy who underwent serial CT imaging at baseline (pre-treatment) and 9 weeks after treatment initiation were analyzed. This dataset was used in our prior segmentation studies and the segmentation results were utilized in this study [31], [35].

Clinical outcomes included progression free survival (PFS), defined as the time from the date of primary surgery to the date of documented first recurrence on the basis of findings on a CT scan, physical examination results, or death prior to recurrence, and OS defined as the time interval between surgery and the date of death or censure. In addition, clinical variables such as patient age, biological sex at birth, and smoking status were also available for measuring associations of WHOG and HOG features using univariate and multivariable Cox proportional hazards regression.

4) *RIDER lung CT dataset*: The RIDER Lung CT collection consists of 31 patients with non-small cell lung cancer [36]. It was constructed as part of a study to evaluate the variability of tumor unidimensional, bidimensional, and volumetric measurements on same-day repeat computed tomographic (CT) scans. Each patient who underwent two chest CT scans within 15 minutes using the same imaging protocol, were included in this study. Three radiologists independently measured the two greatest diameters of each lesion on both scans, and during another session, measured the same tumors of the first scan [37].

D. Comparing methods

We compared our proposed method with CoLIAGe [28], HOG [3], wavelet [6], Gabor [38] and GLCM [8]. We used the implementation HOG of *extractHOGFeatures* function from the Image Processing and Computer Vision package. CoLIAGe, wavelet, Gabor and GLCM features are extracted via the implementation from CERR [32]. We chose the output patch size [8, 8] from the same default setting from both HOG and CoLIAGe (the same for WHOG). For wavelet, Gabor and GLCM features, we used the standard feature extraction pipeline which all parameters are set to in accordance with the IBSI guidelines [39].

E. Statistical analysis

Statistical evaluation of the association of the entropy of WHOG, CoLIAGe, HOG, wavelet, Gabor and GLCM features was performed with the available outcomes for each dataset using univariate Cox proportional hazards regression. Multivariable Cox proportional hazards regression was also done by adjusting for known clinical variables including age, biological sex at birth, smoking status, and HPV status, as appropriate and available for the different datasets. Adjustment

for multiple comparisons between the three features was done using Bonferroni correction. Only p -values with $p < 0.05$ were considered significant. Measurement of association was performed using dichotomized low-risk (\leq median of entropy feature) and high-risk ($>$ median of entropy feature) groups. Kaplan-Meier (KM) survival analysis was performed using the dichotomized patient curves. Population median values were chosen as an unbiased cutoff to dichotomize patients and avoid need for calibrating the best cut-off value from the individual datasets.

V. RESULTS

A. Brain GBM: association of entropy features with overall survival

Results of univariate and multivariable (adjusted for age, tumor volume, biological sex at birth and *MGMT* methylation) Cox proportional hazards regression using the dichotomized high and low-risk groups determined using entropy of WHOG, CoLIAGe, and HOG features are shown in Table I.

	Univariate Analysis		Multivariable Analysis	
	p -value	HR (95% CI)	p -value	HR (95% CI)
WHOG	0.030	2.1 (1.0-4.0)	0.021	2.4 (1.1-4.0)
CoLIAGe	0.049	1.9 (1.0-3.7)	0.059	1.9 (0.9-3.9)
HOG	0.160	1.6 (0.8-3.0)	0.238	1.5 (0.7-2.9)
Wavelet	0.058	1.9 (1.0-3.9)	0.124	1.7 (0.9-3.6)
Gabor	0.071	0.55 (0.3-1.1)	0.139	0.5 (0.2-1.2)
GLCM	0.26	1.4 (0.8-2.7)	0.555	1.2 (0.6-2.4)

TABLE I: Univariate and multivariable Cox proportional hazards regression analysis on brain GBM MRI data. HR: hazard ratio; CI: confidence interval

KM analysis showed that entropy of WHOG resulted in a stronger association with OS ($p = 0.027$) compared with entropy of CoLIAGe ($p = 0.047$) and entropy of HOG ($p = 0.106$) features, the latter of which was not associated with OS. Fig. 4a, Fig. 4f and Fig. 4e show the KM curves for the entropy of WHOG, CoLIAGe, and HOG features, respectively. Fig. 6 shows two representative patients randomly selected from low and high-risk groups as determined using the WHOG entropy feature.

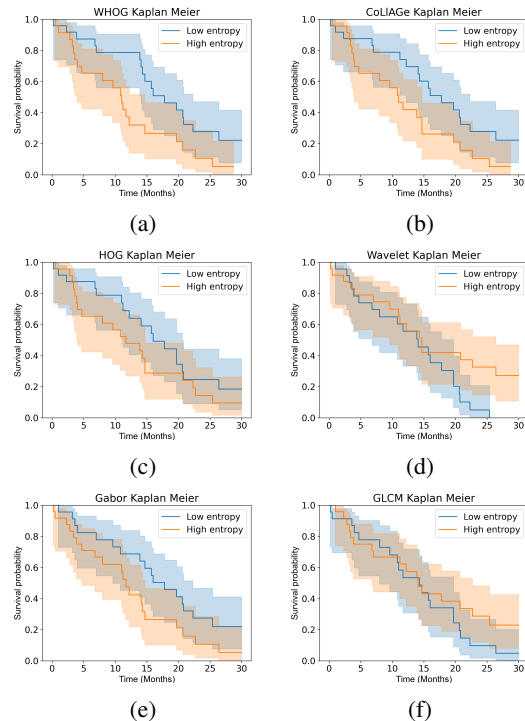


Fig. 4: Kaplan-Meier plots of survival difference between the high entropy group and the low entropy group of brain GBM patients of method WHOG (a), CoLIAGe (b), HOG (c), Wavelet (d), Gabor (e) and GLCM (f)

Robustness of features to noise: We tested the robustness of the extracted features to noise and the impact of added noise in measuring the association with OS by adding zero-centered Gaussian noise as well as Rician noise [40], [41]. WHOG features showed less variability measured using coefficient of variation for the different noise levels. Importantly, the WHOG features extracted at the different noise levels remained associated with OS for larger amounts of added noise when compared to other features, indicating the stability of WHOG to added image noise (see Table ??). On the other hand, CoLIAGe features were not associated with OS with any added noise. Wavelet features gave the smallest p -value of Rician noise $\sigma = 10^{-3}$, with the high risk group and low risk group reversed from the original case. The result could be caused by random effects (see Table ??).

As seen from the results, HOG and wavelet showed a (marginal) dependence on the magnet strengths indicating reduced robustness compared to other methods.

	WHOG	CoLIAGe	HOG
GE and non-GE	0.751	0.622	0.505
1.5 Tesla vs. 3 Tesla	0.977	0.627	0.055
	Wavelet	Gabor	GLCM
GE and non-GE	0.2691	0.5768	0.8353
1.5 Tesla vs. 3 Tesla	0.0176	0.5672	0.8863

TABLE III: Wilcoxon rank sum test on differences of scanner manufacturers and magnetic strength.

Comparison of patients assigned to high and low entropy groups by WHOG and CoLIAGe: We compared patients assigned to different risk groups by WHOG and CoLIAGe methods (see Table IV). Fisher exact test of the patient risk groups showed significant agreement in the assigned risk groups ($p = 7.26e-06$). Spearman rank correlation coefficient of the WHOG and CoLIAGe features was also high (Spearman ρ of 0.85). These results indicate that the two features result in similar grouping of patients, albeit WHOG indicates higher robustness to noise compared to the CoLIAGe method. Fig. 5 shows two examples where WHOG and CoLIAGe resulted in opposite categorization of patients into low and high-risk groups. In particular, CoLIAGe-based dichotomization grouped the image in Fig. 5a into a low-risk group even though the tumor appeared non-homogenous due to the presence of edema and differential contrast in various parts of the tumor. WHOG, on the other hand, placed this patient in the high-risk group and the patient was associated with a shorter OS of 10.8 months. Another representative example where the CoLIAGe categorized the image into the high-risk group, although the presence of homogenous sub-regions is shown in Fig 5(b). WHOG grouped this patient in the low-risk group and the patient was associated with a longer OS of 13.9 months.

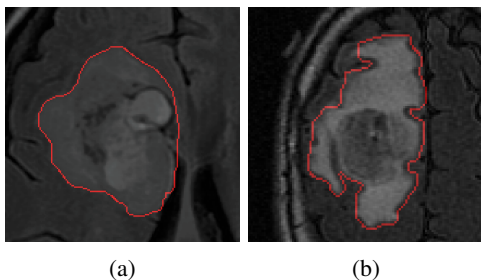


Fig. 5: Two examples of WHOG and CoLIAGe with divergent grouping of patients into low vs. high-risk groups: (a) WHOG based categorization was high-risk but CoLIAGe categorization was low-risk, (b) CoLIAGe categorization was high-risk but WHOG categorization was low-risk.

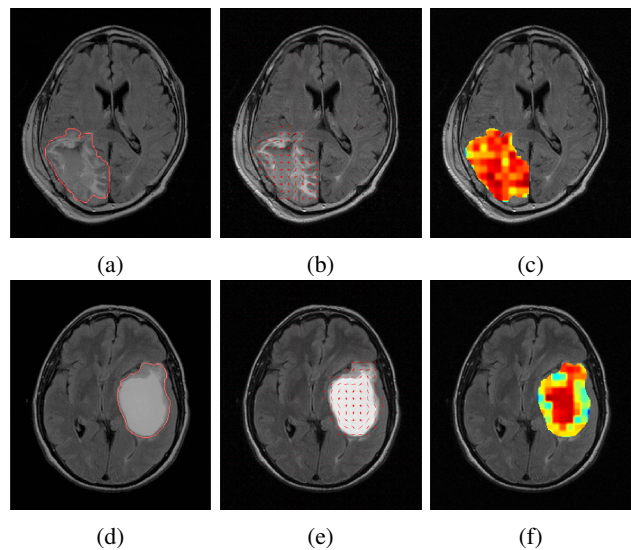


Fig. 6: WHOG categorized high-risk group patients ((a), (b) and (c)) and low-risk group patients ((d), (e) and (f)). Representative central slices are shown. (a),(d): original slice (boundaries of tumors are indicated by the red lines); (b), (e): rose plot of directionality; (c), (f): heatmap of voxel-wise WHOG entropy.

	WHOG	Low entropy group	High entropy group
CoLIAGe			
Low entropy group		20	4
High entropy group		4	19

TABLE IV: Comparison of WHOG low/high entropy groups vs CoLIAGe low/high entropy groups.

B. Head & neck cancer: association with overall survival

Results of univariate and multivariable (adjusted for age, biological sex at birth, tumor volume and overall HPV p16 status) Cox proportional hazards regression using the dichotomized high and low-risk groups determined using entropy of WHOG, CoLIAGe, and HOG features are shown in Table V.

	Univariate Analysis		Multivariable Analysis	
	<i>p</i> -value	HR (95% CI)	<i>p</i> -value	HR (95% CI)
WHOG	0.011	1.9 (1.2-3.1)	0.157	1.5 (0.8-2.8)
CoLIAGe	0.048	1.6 (1.0-2.6)	0.849	1.1 (0.5-2.0)
HOG	0.290	1.3 (0.8-2.1)	0.622	1.1 (0.7-2.0)
Wavelet	0.680	1.1 (0.7-1.8)	0.831	0.9 (0.6-1.6)
Gabor	0.320	1.3 (0.8-2.1)	0.636	0.9 (0.5-1.5)
GLCM	0.041	1.7 (1-2.7)	0.664	1.1 (0.7-2.0)

TABLE V: Univariate and multivariable Cox proportional hazards regression analysis on head & neck CT dataset. HR: hazard ratio; CI: confidence interval

WHOG entropy based patient dichotomization resulted in a significant association with OS ($p = 0.009$) compared to a borderline association when using CoLIAGe ($p = 0.046$) and no association for HOG entropy based patient dichotomization ($p = 0.386$). Fig. 7 and Fig. 9 show KM survival curves for the WHOG and CoLIAGe entropy based patient groups, respectively. Fig. 8 shows two representative patients randomly

selected from low and high-risk groups as determined using the WHOG entropy feature.

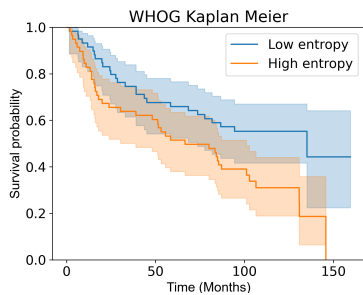


Fig. 7: Kaplan-Meier plots of survival difference between the high entropy group and the low entropy group of head and neck patients of method WHOG.

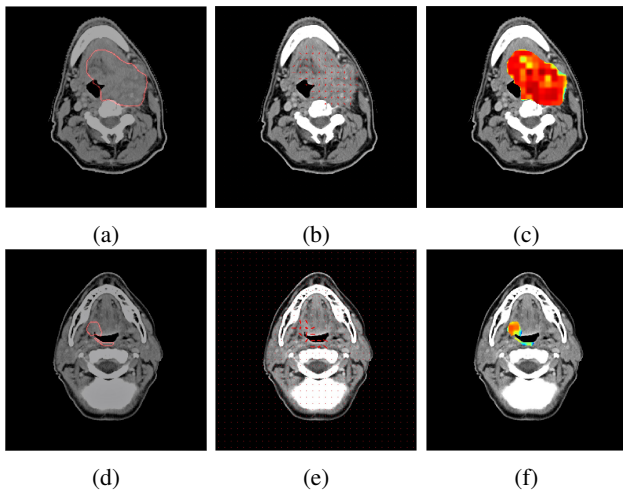


Fig. 8: Central tumor slices from representative patients chosen from high-risk ((a), (b) and (c)) and low-risk ((d), (e) and (f)) groups as selected using WHOG entropy feature. CT image slices with tumor contour are shown in (a) and (d). Rose plots showing the directionality of image gradients as computed using WHOG are shown in (b) and (e). Finally, heatmaps of voxel-wise WHOG entropy for the two patients are shown in (c) and (f).

Comparison of patients assigned to high and low entropy groups by WHOG and CoLIAGe: The WHOG entropy and CoLIAGe entropy features were strongly correlated (Spearman rank correlation coefficient of $\rho = 0.7143$) with a high agreement in the grouping of patients into the high and low-risk groups. However, WHOG entropy was more strongly associated with OS compared to the CoLIAGe entropy as shown in Table V. Fig. 10 shows two examples where WHOG and CoLIAGe resulted in opposite categorization of patients into low and high-risk groups. In particular, CoLIAGe-based dichotomization grouped the image in Fig. 10a into a low-risk group while WHOG placed this patient in the high-risk group and the patient was associated with a shorter OS of 16 months. Another representative example where the CoLIAGe categorized the image into the high-risk group, although the

presence of homogenous sub-regions is shown in Fig 10 (b). WHOG grouped this patient in the low-risk group and the patient was associated with a longer OS of 135 months.

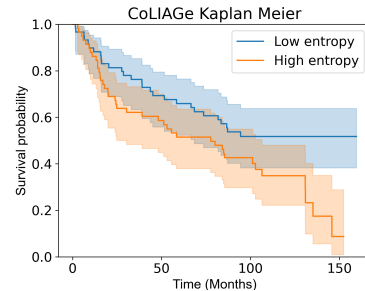


Fig. 9: Kaplan-Meier plots of the high entropy group and the low entropy group from CoLIAGe in head and neck patients.

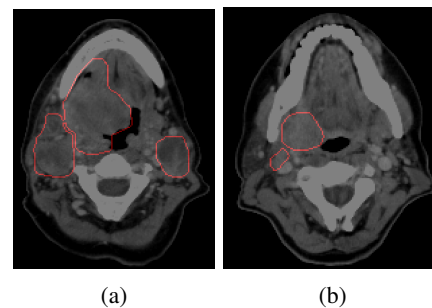


Fig. 10: Two examples of WHOG and CoLIAGe with divergent grouping of patients into low vs. high-risk groups: (a) a patient with a low OS of 15 months and WHOG correctly classified the patient into high-risk but CoLIAGe classified the patient as low-risk, (b) a patient with a longer OS of 135 months, which was correctly categorized as low-risk using WHOG but incorrectly using CoLIAGe as high-risk.

C. Lung cancer treated with immunotherapy: association with overall survival using longitudinal CT scans

Different from prior disease sites, we studied whether longitudinal changes in the entropy features extracted from the CT scans using WHOG and other features are able to predict OS. Longitudinal analysis using the change in the radiomic features requires features that are robust to imaging differences in order to capture real changes resulting from image noise and acquisition differences. Change in entropy was computed between the pre-treatment CT scan (S_0) and the first CT scan acquired 9 weeks after starting treatment (S_1) as the difference in average entropy computed within segmented tumors or $\Delta S = S_1 - S_0$.

As shown in Table VI, only the change in WHOG entropy was associated with OS in both univariate and multivariable (adjusted for age, biological sex at birth, smoking, treatment, tumor volume and tumor volume change) Cox proportional hazards regression analysis. HOG entropy based univariate analysis showed an association with OS, but it was not associated in the multivariable analysis. Furthermore, WHOG entropy resulted in a stronger association, indicated by smaller

p -values than HOG entropy. Different from the other two datasets, CoLIAGe showed no association with OS.

	Univariate Analysis		Multivariable Analysis	
	p -value	HR (95% CI)	p -value	HR (95% CI)
WHOG	2.7e-5	2.2 (1.5-3.2)	0.0004	2.1 (1.4-3.2)
CoLIAGe	0.670	1.1 (0.8-1.5)	0.970	1.0 (0.7-1.5)
HOG	0.001	1.6 (1.1-2.4)	0.470	1.2 (0.8-1.8)
Wavelet	0.51	1.2 (0.8-1.7)	0.789	1.1 (0.7-1.5)
Gabor	0.15	0.8 (0.5-1.1)	0.771	0.9 (0.6-1.4)
GLCM	0.13	1.4 (1.0-1.9)	0.274	1.2 (0.8-1.8)

TABLE VI: Univariate and multivariable Cox proportional hazards regression analysis performed on the lung cancer dataset using longitudinal or delta entropy features. HR: hazard ratio; CI: confidence interval

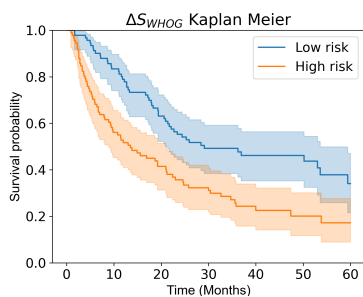


Fig. 11: Kaplan-Meier plots of association with OS computed using dichotomized patients (low-risk as $\Delta S_{WHOG} \leq 0$ vs. high-risk as $\Delta S_{WHOG} > 0$) using WHOG entropy.

The KM curves showing the association of the WHOG entropy change with OS are shown in Fig. 11. Additionally, we studied whether the WHOG entropy features computed from the individual CT (baseline and 9-week) scans showed an association with OS. As shown in Fig. 12a, the features computed from the baseline image did not show an association with OS ($p = 0.444$), but those computed after the first treatment scan (Fig. 12b) did ($p = 0.0172$). On the other hand, the delta WHOG entropy feature showed a stronger association ($p < 0.0001$) compared to the WHOG entropy computed from the 9-week CT.

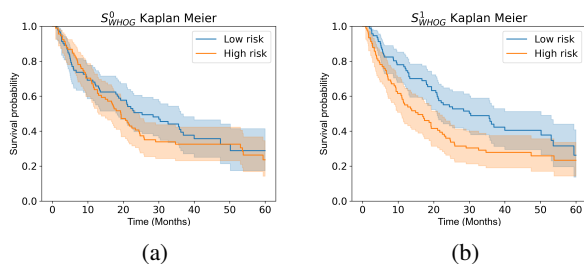


Fig. 12: Kaplan-Meier plots of association with OS computed using dichotomized patients using (a) (low-risk as $\leq S_{WHOG}^0$ vs. high-risk as $> S_{WHOG}^0$) and (b) (low-risk as $\leq S_{WHOG}^1$ vs. high-risk as $> S_{WHOG}^1$), where the computed features are WHOG entropy measures, 0 corresponds to features extracted within tumor from baseline CT and 1 corresponds to features extracted within tumor from 9-week CT scans, respectively.

Of note, the high-risk patients were those in whom the delta entropy increased with the WHOG feature, indicating higher imaging heterogeneity within tumor after treatment (median of 0.0492 and inter-quartile range (IQR) of 0.0450). On the other hand, patients grouped as low-risk using the delta WHOG feature had a decrease in the entropy (median of -0.0632 and IQR of 0.1046).

We next evaluated the association of patients in the high-risk $\Delta S_{WHOG} > 0$ vs. low-risk $\Delta S_{WHOG} \leq 0$ groups with disease response, namely, partial or complete response (PR/CR), stable disease (SD), and progression of disease (POD). Fisher exact test showed a significant difference ($p = 3.78 \times 10^{-6}$) between the two groups, showing a larger proportion of POD in the group with increased tumor heterogeneity than the group with decreased WHOG entropy (Table VII). On the other hand, patients who had a reduction in WHOG entropy, had a higher prevalence of stable disease.

Group	Response		
	POD	SD	PR/CR
$\Delta S_{WHOG} > 0$	38/65.52%	5/8.62%	15/25.86%
$\Delta S_{WHOG} < 0$	33/26.40%	50/40.00%	42/33.60%

TABLE VII: Response counts and percentages of the entropy increase group, the minor change group and the entropy decrease group from the baseline images to the 9-week images after the start of immunotherapy. POD: progression of disease, SD: stable disease, PR: partial response, CR: complete response.

D. RIDER lung CT dataset: Reproducible and robust image feature

Figure 13 shows the feature values (normalized by mean and standard deviation to align different features into the same range) from two CT scans that were taken 15 minutes apart.

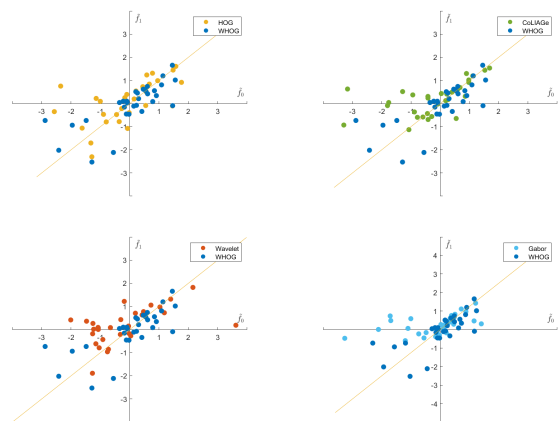


Fig. 13: Scatter plots of normalized features extracted from two CT scans of HOG vs WHOG, CoLIAGe vs WHOG, Wavelet vs WHOG and Gabor vs WHOG.

Robustness as a repeatable feature: The robustness of the proposed WHOG image feature and other features in terms of repeatability was measured by the Pearson correlation coefficients and p -values between the output from two time

points. To mitigate the effect of WHOG using multi-scale, single-scaled features of WHOG are extracted compared to HOG and CoLIAGe features with the same scale (patch size). Table VIII shows that Pearson correlation coefficients and p-values between two feature outputs. We expect robust radiomic features to have similar values in two consecutive CT measures. So higher correlation coefficients of WHOG features than other methods indicates that it is indeed a robust feature.

	WHOG(4)	WHOG(8)	WHOG(4+8)	HOG(4)	HOG(8)
ρ	0.7332	0.7542	0.7568	0.6701	0.6034
p	2.71e-06	9.59e-07	8.37e-07	3.72e-05	3.26e-04

	CoLIAGe(4)	CoLIAGe(8)	Wavelet	Gabor	GLCM
ρ	0.5755	0.3752	0.5437	0.4705	0.1040
p	7.06e-04	0.0375	0.0016	0.0076	0.5777

TABLE VIII: Pearson correlation coefficients and p-values between two feature outputs from CT scans of 15 minutes apart. The numbers in parentheses are the patch sizes used in computation. In particular, WHOG(4+8) is the multi-scale version of WHOG feature used in experiments of other data sets. Wavelet, Gabor and GLCM parameters are defined according to IBSI guidelines.

VI. CONCLUSION AND FUTURE WORK

The present work makes use of optimal transport for the local analysis of general distributions. We explore the expressions under three commonly used optimal transport models. In particular, we utilize the discrete Kantorovich version for image directionality feature extraction. Because of the robust nature of optimal transport, we find our method performs well even in the presence of noise. The main reason for this is that our approach does not use gradients, and the underlying Wasserstein metric is continuous to perturbations in the data.

We reached a similar conclusion about image directionality and tumor heterogeneity as in [28]. They performed classifications based on their HOG based feature vectors. In principle, we can replace their HOG extracted feature by our WHOG feature. By doing this, we hope that the classification accuracy may be improved.

One area of improvement for the proposed algorithm concerns the speed of computation. We noticed in our experiments, that the WHOG method took more time to compute than the other tested methodologies. To test feasibility, our implementation only used standard dual-simplex algorithm of linear programming. There are a number of other algorithms for the step of solving the Monge-Kantorovich problem, which should lead to a better performance [42], [43].

We plan to further evaluate ability of our method to study pathology images [44]–[47]. The microscopic features in these high resolution images contain information for which our proposed method may better capture directionality, and thus distinguish different biological processes occurring in different tissues. Our preliminary results indicate that our WHOG extracted feature may indeed be correlated with tumor stage and recurrence (See an example in Figure 14).

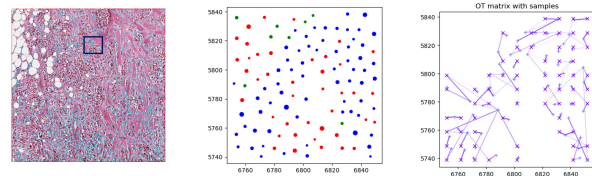


Fig. 14: An example of extracting WHOG feature from a pathology image: Centroids of different cell types are extracted. An optimal transport map is used to compute directionality on a patch following a similar pipeline as done for other image types.

The WHOG feature has the potential to be integrated into complex clinical applications. We have demonstrated its ability to quantify tumor heterogeneity and its applicability to several modalities of medical imaging data. It may be utilized in conjunction with other commonly used image features for patient prognosis prediction. We have indicated that the WHOG directionality feature, extracted from image data, represents an attractive robust alternative to the more standard HOG or CoLIAGe techniques. It is interpretable, carries prognostic information, and may be incorporated into a machine learning framework to enhance performance.

The method can be easily modified to the case of color imagery by looking at the transport map between a local cube and a pure color cube. We also want to test on the 3D extension so that the directionality is within the transverse plane, frontal, and sagittal directions. Finally, we plan to consider certain possible connections between the extracted WHOG features and the fractal dimension of the tumor especially for MRI data where textural information is more clearly exposed.

REFERENCES

- [1] W. Freeman and E. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, 1991.
- [2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004. [Online]. Available: <https://doi.org/10.1023/b:visi.0000029664.99615.94>
- [3] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," pp. 886–893, 6 2005.
- [4] L. Zheng, Y. Yang, and Q. Tian, "Sift meets cnn: A decade survey of instance retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 5, pp. 1224–1244, 2018.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893, 2005.
- [6] H. J. W. L. Aerts, E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebens, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, no. 1, Jun. 2014.
- [7] S. Mallat, "Wavelet tour of signal processing," in *Comprehensive Sampling and Reconstruction*. Elsevier, 2009, pp. 1–52.
- [8] V. BinoSebastian, A. Unnikrishnan, and K. Balakrishnan, "Gray level co-occurrence matrices: Generalisation and some new features," *ArXiv*, vol. abs/1205.4831, 2012.
- [9] C. Villani, *Topics in Optimal Transportation*. American Mathematical Soc., 2003.
- [10] —, *Optimal Transport: Old and New*. Springer Science & Business Media, 2008, vol. 338.
- [11] L. C. Evans and W. Gangbo, "Differential equations methods for the Monge-Kantorovich mass transfer problem," *Memoirs of the American Mathematical Society*, vol. 137, no. 653, 1999.

[12] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*, ser. Probability and its Applications. Berlin: Springer, 1998.

[13] L. V. Kantorovich, "On a problem of monge," *CR (Doklady) Acad. Sci. URSS (NS)*, vol. 3, pp. 225–226, 1948.

[14] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arxiv.org*, vol. 1701.07875, 2017.

[15] E. A. Carlen and J. Maas, "An analog of the 2-Wasserstein metric in non-commutative probability under which the Fermionic Fokker-Planck equation is gradient flow for the entropy," *Communications in Mathematical Physics*, vol. 331, no. 3, pp. 887–926, 2014.

[16] S. Haker, L. Zhu, A. Tannenbaum, and S. Angenent, "Optimal mass transport for registration and warping," *International Journal of Computer Vision*, vol. 60, no. 3, pp. 225–240, Dec. 2004.

[17] M. Mittnenzweig and A. Mielke, "An entropic gradient structure for lindblad equations and coupling of quantum systems to macroscopic models," *J. Stat. Physics*, vol. 167, no. 2, 2017.

[18] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volumes I and II*. Springer Science & Business Media, 1998.

[19] J. C. Matthews, S. Nadeem, M. Pouryahya, Z. Belkhatir, J. O. Deasy, A. J. Levine, and A. R. Tannenbaum, "Functional network analysis reveals an immune tolerance mechanism in cancer," *Proceedings of the National Academy of Sciences*, vol. 117, no. 28, pp. 16 339–16 345, jul 2020.

[20] J. Zhu, J. H. Oh, J. O. Deasy, and A. Tannenbaum, "vwcluster: A network based clustering of multi-omics breast cancer data based on vector-valued optimal transport," *bioRxiv*, 2021. [Online]. Available: <https://www.biorxiv.org/content/early/2021/06/18/2021.06.17.448878>

[21] M. Pouryahya, J. Oh, P. Javanmard, J. C. Matthews, Z. Belkhatir, J. O. Deasy, and A. Tannenbaum, "awcluster: A novel integrative network-based clustering of multiomics for subtype analysis of cancer data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, no. 01, pp. 1–1, nov 5555.

[22] X. Chen, A. P. Tran, R. Elkin, H. Benveniste, and A. Tannenbaum, "Visualizing fluid flows via regularized optimal mass transport with applications to neuroscience," 01 2022.

[23] M. A. A. O. J. R. A. D. J. T. A. Zhu, J.; Thor, "A non-parametric analysis to identify high-risk vs. low-risk anatomic regions associated with reduced overall survival in the rtog 0617 clinical trial data," 06 2021.

[24] P. M. Pardalos, "Linear programming," in *Encyclopedia of Optimization*. Springer US, 2008, pp. 1883–1886.

[25] S. Roger, J. L. Elaine, V. Maria, D. Laurent, C. Stéphane, R. H. Shan, V. Loïc, B. David, L. Andrea, A. Samy, H. Antoine, S. Jean-Yves, M. Aurélien, M. Christophe, S. Jean-Charles, R. Charlotte, P. Nikos, D. Eric, and F. Charles, "A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-11 immunotherapy: an imaging biomarker, retrospective multicohort study," *Lancet Oncol*, vol. 19, no. 9, pp. 1180–1191, 2018.

[26] G. Olya, E. B. Anders, B. S. Matthew, J. A. Hugo, D. Andre, W. Hua, R. V. Emmanuel, L. Philippe, G. Yuhua, B. Yoganand, E. Edward, A. G. Robert, E. Steven, and J. G. Robert, "Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma," *Plos One*, vol. 10, no. 3, p. e0118261, 2015.

[27] J. S. Elizabeth, O. Jung-Hun, Z. D. Brittany, V. Harini, P. A. Aditya, B. T. Sunitha, O. D. Joseph, and A. M. Elizabeth, "Breast cancer subtype intertumor heterogeneity: MRI-based features predict results of a genomic assay," *Journal of Magnetic Resonance Imaging*, vol. 42, no. 5, pp. 1398–1406, 2015.

[28] P. Prasanna, P. Tiwari, and A. Madabhushi, "Co-occurrence of Local Anisotropic Gradient Orientations (CoLIAGE): A new radiomics descriptor," *Scientific Reports*, vol. 6, no. 1, p. 37241, Dec. 2016. [Online]. Available: <http://www.nature.com/articles/srep37241>

[29] V. Harini, F. F. Claire, F. D. Deborah, N. Josip, H. Yuki, G. B. Silvio, C. Giovanni, P. Iva, N. Stephanie, N. Ines, Z. Ahmet, R. A.-R. Nadeem, A. Carol, Z. Dmitriy, A. C. Karen, A. D. J. Luis, M. L. J. Mario, M. Vicky, A. S. Robert, J. M. Jennifer, W. Britta, and Y. Lakhman, "Machine learning-based prediction of microsatellite instability and high tumor mutation burden from contrast-enhanced computed tomography in endometrial cancers," *Sci Rep*, vol. 10, no. 1, p. 17769, 2020.

[30] U. Hyemin, T. Florent, B. Dalton, R. J. Y. Joseph, O. Deasy, and V. Harini, "Impact of image preprocessing on the scanner dependence of multi-parametric MRI radiomic features and covariate shift in multi-institutional glioblastoma datasets," 2019.

[31] J. Jue, H. Yu-Chi, L. Chia-Ju, H. Darragh, D. H. Matthew, O. D. Joseph, M. Gig, and V. Harini, "Multiple resolution residually connected feature streams for automatic lung tumor segmentation from ct images," *IEEE Trans Med Imaging*, vol. 38, no. 1, pp. 134–144, 2019.

[32] A. Apte, A. Iyer, M. Crispin-Ortuzar, R. Pandya, van Dijk L.S, E. Spezi, M. Thor, H. Um, H. Veeraraghavan, O. J.H, A. Shukla-Dave, and J. Deasy, "Technical note: Extension of CERR for computational radiomics: A comprehensive MATLAB platform for reproducible radiomics research," *Med Phys*, 2018.

[33] C. Kenneth, V. Bruce, S. Kirk, F. John, K. Justin, K. Paul, M. Stephen, P. Stanley, M. David, P. Michael, T. Lawrence, and P. Fred, "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *J. Digit Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.

[34] E. Katsoulakis, Y. Yu, A. P. Apte, J. E. Leeman, N. Katabi, L. Morris, J. O. Deasy, T. A. Chan, N. Y. Lee, N. Riaz, V. Hatzoglou, and J. H. Oh, "Radiomic analysis identifies tumor subtypes associated with distinct molecular and microenvironmental factors in head and neck squamous cell carcinoma," *Oral Oncology*, vol. 110, p. 104877, Nov 2020.

[35] J. Jue, R. Andreas, O. D. Joseph, and V. Harini, "Unpaired cross-modality educed distillation (CMEDL) for medical image segmentation," *IEEE Trans Med Imaging*, vol. 41, no. 5, pp. 1057–1068, 2022.

[36] B. Zhao, L. H. Schwartz, and M. G. Kris, "Data from rider_lung_ct," 2015.

[37] B. Zhao, L. P. James, C. S. Moskowitz, P. Guo, M. S. Ginsberg, R. A. Lefkowitz, Y. Qin, G. J. Riely, M. G. Kris, and L. H. Schwartz, "Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer," *Radiology*, vol. 252, no. 1, pp. 263–272, Jul. 2009.

[38] G. H. Granlund, "In search of a general picture processing operator," *Computer Graphics and Image Processing*, vol. 8, no. 2, pp. 155–173, 1978.

[39] A. Zwanenburg, M. Vallieres, M. A. Abdalah, H. J. W. L. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, M. Bogowicz, L. Boldrini, I. Buvat, G. J. R. Cook, C. Davatzikos, A. Depeursinge, M.-C. Desserot, N. Dinapoli, C. V. Dinh, S. Echegaray, I. E. Naqa, A. Y. Fedorov, R. Gatta, R. J. Gillies, V. Goh, M. Gotz, M. Guckenberger, S. M. Ha, M. Hatt, F. Isensee, P. Lambin, S. Leger, R. T. Leijenaar, J. Lenkiewicz, F. Lippert, A. Losnegard, K. H. Maier-Hein, O. Morin, H. Muller, S. Napel, C. Nioche, F. Orhac, S. Pati, E. A. Pfaehler, A. Rahmim, A. U. Rao, J. Scherer, M. M. Siddique, N. M. Sijtsema, J. S. Fernandez, E. Spezi, R. J. Steenbakkers, S. Tanadini-Lang, D. Thorwarth, E. G. Troost, T. Upadhya, V. Valentini, L. V. van Dijk, J. van Griethuysen, F. H. van Velden, P. Whybra, C. Richter, and S. Lock, "The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, May 2020.

[40] A. K. Boyat and B. K. Joshi, "A Review Paper : Noise Models in Digital Image Processing," *Signal & Image Processing : An International Journal*, vol. 6, no. 2, pp. 63–75, Apr. 2015.

[41] H. Gudbjartsson and S. Patz, "The rician distribution of noisy mri data," *Magnetic Resonance in Medicine*, vol. 34, no. 6, pp. 910–914, Dec. 1995.

[42] N. Bonneel, M. van de Panne, S. Paris, and W. Heidrich, "Displacement interpolation using Lagrangian mass transport," *ACM Transactions on Graphics*, vol. 30, no. 6, p. Article n°158, 2011.

[43] M. Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds., vol. 26. Curran Associates, Inc., 2013.

[44] J. Xu, H. Lu, H. Li, C. Yan, X. Wang, M. Zang, D. G. de Rooij, A. Madabhushi, and E. Y. Xu, "Computerized spermatogenesis staging (css) of mouse testis sections via quantitative histomorphological analysis," *Medical Image Analysis*, vol. 70, p. 101835, 2021.

[45] H. Lu, M. Zang, G. P. L. Marini, X. Wang, Y. Jiao, N. Ao, K. Ong, X. Huo, L. Li, E. Y. Xu, W. W. B. Goh, W. Yu, and J. Xu, "A novel pipeline for computerized mouse spermatogenesis staging," *Bioinformatics*, vol. 38, no. 23, pp. 5307–5314, 10 2022.

[46] H. Chen, C. Li, X. Li, M. M. Rahaman, W. Hu, Y. Li, W. Liu, C. Sun, H. Sun, X. Huang, and M. Grzegorzec, "Il-mcam: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach," *Computers in Biology and Medicine*, vol. 143, p. 105265, 2022.

[47] J. Zhang, C. Li, S. Kosov, M. Grzegorzec, K. Shirahama, T. Jiang, C. Sun, Z. Li, and H. Li, "Lcu-net: A novel low-cost u-net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, p. 107885, 2021.