Optimal Mass Transport for Robust Texture Analysis

Zehor Belkhatir[1,¶], Aditi Iyer[1], James C. Mathews[1], Maryam Pouryahya[1], Saad Nadeem[1], Joseph O. Deasy[1], Aditya P. Apte[1,*], and Allen R. Tannenbaum[2,*]

[1]Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY, USA

[2]Departments of Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY, USA

Abstract

The emerging field of radiomics, which consists of transforming standard-of-care images to quantifiable scalar statistics, endeavors to reveal the information hidden in these macroscopic images. This field of research has found different applications ranging from phenotyping and tumor classification to outcome prediction and treatment planning. Texture analysis, which often consists of reducing spatial texture matrices to summary scalar features, has been shown to be important in many of the latter applications. However, as pointed out in many studies, some of the derived texture statistics are strongly correlated and tend to contribute redundant information; and are also sensitive to the parameters used in their computation, e.g., the number of gray intensity levels. In the present study, we propose first to consider texture matrices, with an emphasis on gray-level co-occurrence matrix (GLCM), as a non-parametric multivariate objects. The proposed modeling approach avoids evaluating redundant and strongly correlated features and also prevents the feature processing steps. Then, via the Wasserstein distance from optimal mass transport theory, we propose to compare these spatial objects to identify computerized tomography slices with dental artifacts in head and neck cancer. We demonstrate the robustness of the proposed classification approach with respect to the GLCM extraction parameters and the size of the training set. Comparisons with the random forest classifier, which is constructed on scalar texture features, demonstrates the efficiency and robustness of the proposed algorithm.

¶ email: belkhatz@mskcc.org

* These authors share the senior authorship.

I. INTRODUCTION

The importance of medical imaging has expanded to all major area of health care, including oncology. Motivated by the goal of unveiling the information hidden in standard-of-care images, the research field of radiomics has emerged [1]. Radiomics, which consists of the high-throughput extraction of quantitative features from medical images using automated quantification algorithms, is becoming a powerful tool in different applications, e.g., outcome prediction, tumor classification, treatment planning, and personalized therapy [2], [3], [4], [5] to cite few examples. Texture analysis, which refers to the quantification of spatial variations in gray levels within an image or region of interest (ROI), is widely used and has been successful in different studies, e.g., [6], [7], [8].

Usually in the radiomics research field, characterizing texture consists first of computing multivariate texture matrices such as gray-level co-occurrence matrix (GLCM), run-length matrix (RLM), size zone matrix (SZM), and neighborhood gray-tone difference matrix (NGTDM), and then reducing these multivariate objects to scalar summary statistics. This reduction step may lead to a loss of the spatial information that is inherent in texture matrices. In fact, Vickers and Modestino [9] noted in their work that using scalar features of GLCM for a classification task is suboptimal and that better results may be obtained by using the spatial GLCM directly in their proposed maximum likelihood classifier. Moreover, it has been pointed out in several studies that some of the resulting summary statistics (i) are highly correlated which may lead to overfitting [10], [11], [12], and (ii) also depend on gray-level discretization [13]. In [14], the authors propose to model GLCM as a multivariate object using a latent Gaussian Markov random field model in order to avoid the loss of spatial information. The method has shown promising results in classifying benign and malignant adrenal lesions. Moreover, considering the GLCM as a density function resulted in texture features that are less sensitive to the gray-level quantization, as shown in [15], [16].

Motivated by these recent works, we propose in this study: (i) to consider texture matrices in general, and GLCM in particular, as non-parametric multivariate objects; and (ii) use Wasserstein distance, from Optimal Mass Transport (OMT) theory, as a metric to compare them for the purpose of classifying computed tomographic (CT) slices with dental artifacts in head and neck cancer (H&NC). The OMT problem seeks the most efficient way to transform one distribution

of mass to another given a cost function [17]. It origins go back in 1781, when Gaspar Monge formulated the problem of finding the minimal transportation cost to redistribute earth for building fortifications [18]. Leonid Kantorovich in 1942, relaxed Monge's formulation to find an optimal coupling of distributions using linear programming [19]. Since then, OMT has played a crucial role in many fields of science and engineering; see [20], [21]and references therein. The increased interest in the use of OMT-based metrics, known as Wasserstein distance or Earth-Mover's-Distance (EMD) in the engineering field, is mainly due to their natural ability to capture spatial information when comparing signals, images, or other types of data. This allows to provide various data distributions with different geometric interpretations, which we are seeking to capture from spatial texture matrices in the present work.

II. MATERIALS AND METHODS

Dental artifacts can dramatically attenuate the quality of CT images and make their analysis challenging. It is essential, therefore, to develop an automated algorithm to detect CT slices with dental artifacts in H&NC. This preliminary preprocessing phase will allow taking measures to reduce the effect of the noisy slices when using the images for other modeling purposes. In this section, we present an automatic classification algorithm, which is based on the idea of considering GLCM as a multivariate object, not a summary of scalar statistics, and employs the Wasserstein distance to compare pairs of GLCMs.

A. Data characteristics

The proposed method was applied to two datasets: (i) an internal dataset of 44 H&N CT scans from our institution, resulting in 1165 axial slices with voxel sizes ranging from $0.0914 - 0.1367$ cm (median: 0.0977 cm) and slice thickness $0.1 - 0.3$ cm (median: 0.25 cm), and (ii) an external dataset of 24 H&N CT scans from the open-source archive TCIA [22], resulting in 679 slices with voxel sizes ranging from $0.0352 - 0.0977$ cm (median: 0.0488 cm) and slice thickness $0.0450 - 0.5009$ cm (median: 0.25 cm). Both datasets were resampled to 0.0977 cm $\times 0.0977$ cm $\times 0.25$ cm prior to analysis, resulting in 1235 and 620 slices from internal and external datasets, respectively. Each CT slice was labeled as noisy or clean based on the presence of dental artifacts by a medical imaging expert, resulting in $334 \, (27.04\%)$ noisy and $901 \, (72.96\%)$ clean slices from MSK cohort, and $53 \, (8.55\%)$ noisy and $567 \, (91.45\%)$ clean slices from TCIA cohort.

B. Radiomic texture features

CT images were thresholded (at the 5th percentile) to exclude regions of air, followed by morphological processing to extract the patient's outline. GLCMs were extracted within the resulting mask using the radiomics extension of the Computational Environment for Radiological Research (CERR) [23]. GLCMs were computed by combining contributions from all 2-D neighbors, using 4 different gray levels from $16 - 128$ and 5 different sampling rates between $\pm 15\%$ from the original resolution. The resulting parameters are listed in Table I. Additionally, 25 scalar features, which are listed in Table II, were extracted from the GLCMs to be used in a machine learning classifier for comparison purpose.

Voxel resolution	$0.0830\,\text{cm} \times 0.0830\,\text{cm}$, $0.0903\,\text{cm} \times 0.0903\,\text{cm}$, $0.0977\,\text{cm} \times 0.0977\,\text{cm}$, $0.1050\,\text{cm} \times 0.1050\,\text{cm}$, $0.1123\,\text{cm} \times 0.1123\,\text{cm}$
No. gray levels	16, 32, 64, 128

TABLE I: GLCM extraction parameters.

1 : Energy	10 : Inverse difference normalized	18 : Cluster tendency
2 : Joint entropy	11 Inverse variance	19 : Cluster shade
3 : Joint max	12 : Difference entropy	20 : Cluster prominence
4 : Joint average	13 : Difference variance	21 : Haralick correlation
5 : Joint variance	14 : Sum average	22 : Auto-correlation
6 : Contrast	15 : Sum entropy	23 : First measure of information correlation
7 : Inverse difference moment	16 : Sum variance	24 : Second measure of information correlation
8 : Inverse difference moment normalized	17 : Correlation	25 : Dissimilarity
9 : Inverse difference		

TABLE II: GLCM-based scalar features.

C. Image classification algorithm using spatial GLCMs and Wasserstein distance

We use the theory of OMT [17] to define distances between spatial GLCMs, which we consider as 2-D probability distributions.

Let ρ_0, ρ_1 be two d-dimensional probability distributions defined on $\Omega \subset \mathbb{R}^d$. The ***Wasserstein-1 distance*** is defined as follows:

$$\inf_{\pi:\Omega\times\Omega\to\mathbb{R}} \int_{x,y\in\Omega} \|x-y\|_p \, \pi(x,y) \, dxdy,$$

$$\text{subject to } \int_{y\in\Omega} \pi(x,y)dy = \rho_0(x), \; \forall x \in \Omega$$

$$\int_{x\in\Omega} \pi(x,y)dx = \rho_1(y), \;\; \forall y \in \Omega \tag{1}$$

$$\pi(x,y) \geq 0, \;\; \forall x,y \in \Omega$$

where $\|.\|_p$, $1 \leq p \leq \infty$, is the ground metric of the Wasserstein distance. The variable π is the set of joint distributions $\pi : \Omega \times \Omega \to \mathbb{R}$ whose marginal distributions are ρ_0, ρ_1.

An equivalent alternative formulation of the Wasserstein-1 distance, which is simpler and computationally more efficient, is defined by the following optimization problem:

$$\inf_{m:\Omega\to\mathbb{R}^d} \int_{x\in\Omega} \|m(x)\|_p \, dx,$$

$$\text{subject to } \nabla \cdot (m(x)) = \rho_0(x) - \rho_1(x), \forall x \in \Omega \tag{2}$$

$$m(x).n(x) = 0, \;\; \forall x \in \partial\Omega$$

where "$\nabla\cdot$" denotes the divergence operator, $n(x)$ is the normal to the boundary $\partial\Omega$, and m is a d-dimensional field satisfying the zero flux boundary condition [24]. A fast numerical scheme that relies on multilevel primal-dual optimization algorithms was proposed in [25] to solve (2). This latter numerical scheme is adopted in the present study.

• **Proposed classification algorithm:** We start by normalizing the computed GLCMs to be 2-D probability distributions. This step, essential to apply the balanced OMT theory introduced earlier, doesn't affect the textural pattern. However, interestingly, it was found in recent studies that this step offers some advantages. It adjusts for heterogeneity in lesion size [14], and produce texture features that are invariant to the quantization gray-levels [15], [16]. We then partition the available data to training and test sets, where the training dataset consists of a subset of MSK slices, and the test dataset is comprised of the whole TCIA slices. The proposed supervised classifier mainly relies on comparing 2-D normalized GLCMs of the test CT slices with the ones of the training set using Wasserstein-1 distance. We point out that the comparison between

the GLCMs extracted from MSK and TCIA cohorts is performed using the same values of the extraction parameters (voxel resolution and number of gray levels). The main steps of the proposed classification algorithm, which are depicted in Figure 1, are given as follows

(i) We randomly select a subset, $N_1 = 50$, of noisy and clean slices, denoted ns_j and cs_j, for $j = 1 : N_1$, respectively, from the training cohort.

(ii) Then, for a given TCIA slice denoted $s_i, i = 1 : N_2$, with $N_2 = 620$, we compute the Wasserstein distance between its corresponding GLCM and the ones of all noisy and clean MSK slices. This step is repeated 5 times, resulting two matrices $M_c^r = W_1^r(s_i, cs_j) \in \mathbb{R}^{N_2 \times N_1}$ and $M_n^r = W_1^r(s_i, ns_j) \in \mathbb{R}^{N_2 \times N_1}$, for $r = 1 : 5$.

(iii) Based on the averaged comparison with respect to the training dataset between $W_1^r(s_i, cs_j)$ and $W_1^r(s_i, ns_j)$, the class for slice s_i is concluded.

(iv) Finally, Youden's J index = (sensitivity + specificity − 1), which is a way of summarizing sensitivity and specificity metrics into a single numeric value, was used to assess the accuracy of the classification results.

III.   RESULTS

We evaluated the performance of the proposed algorithm for different parameter combinations involved in the computation of GLCM. We also compared the results of the proposed classifier with the random forest classifier, which uses the scalar features shown in Table II. Robustness of both algorithms with respect to the GLCM computation settings and size of the training set was carried out as well.

Figure 2 depicts Youden's J index that combines the performances of classifying the noisy and clean slices correctly, i.e., sensitivity and specificity, in a single number. The best performance of 77.76% (sensitivity: 87.2%, specificity: 90.57%) is achieved for a gray-level discretization level of 16 and voxel resolution of $0.83\,\text{mm} \times 0.83\,\text{mm}$. It is worth noting, however, that the performance for other combination pairs of GLCM extraction parameters is also good with small variability between the relatively bad and best performances (mean = 76.66 %, std = 1.54 %). Regarding the random forest classifier, the best performance of 76.24% (sensitivity: 89.07%, specificity: 87.17%) is achieved with gray-level quantization of 32 and voxel resolution of $1.05\,\text{mm} \times 1.05\,\text{mm}$ as illustrated in Figure 2. We emphasize, however, that the performance
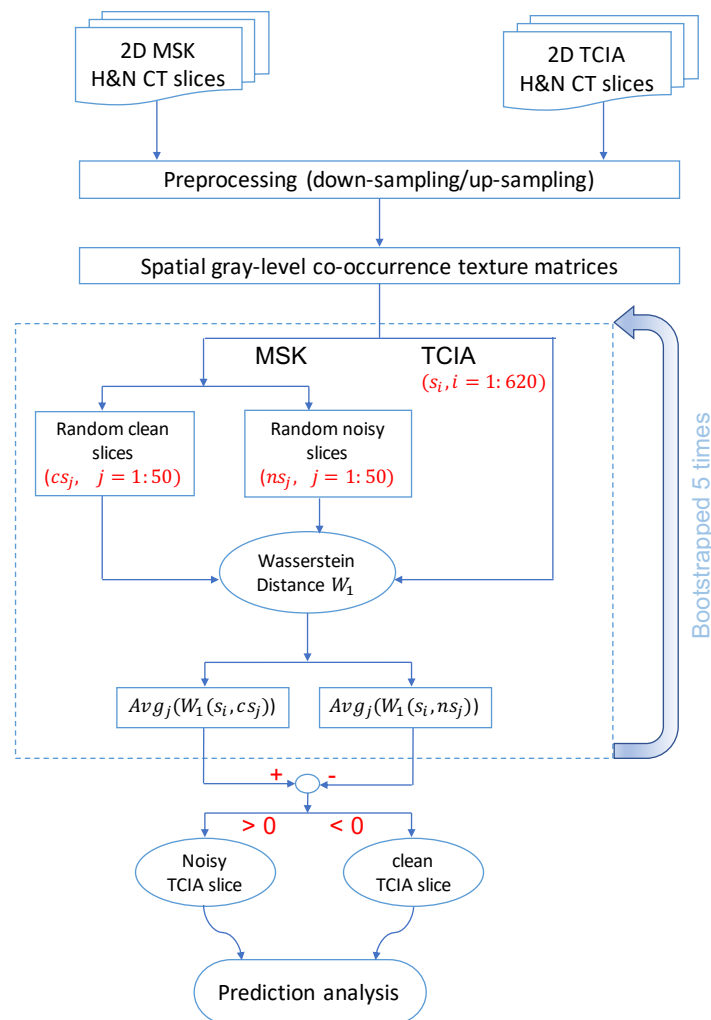
Fig. 1: Flowchart of the proposed classification algorithm for CT slices with dental artifacts in H&NC.

of the random forest technique is more sensitive to the GLCM computation settings (mean = 69.94 %, std = 3.63 %).

We also tested the sensitivity of the performance of the classifiers to different sizes of training data. The results of the proposed classifier and the random forest are given in Tables III and IV, respectively. The OMT/spatial GLCM based method performs better than the random forest approach for a small number of training data (10 slices). Even for this small number of training data, the proposed algorithm was not very sensitive to the GLCM computation settings, while
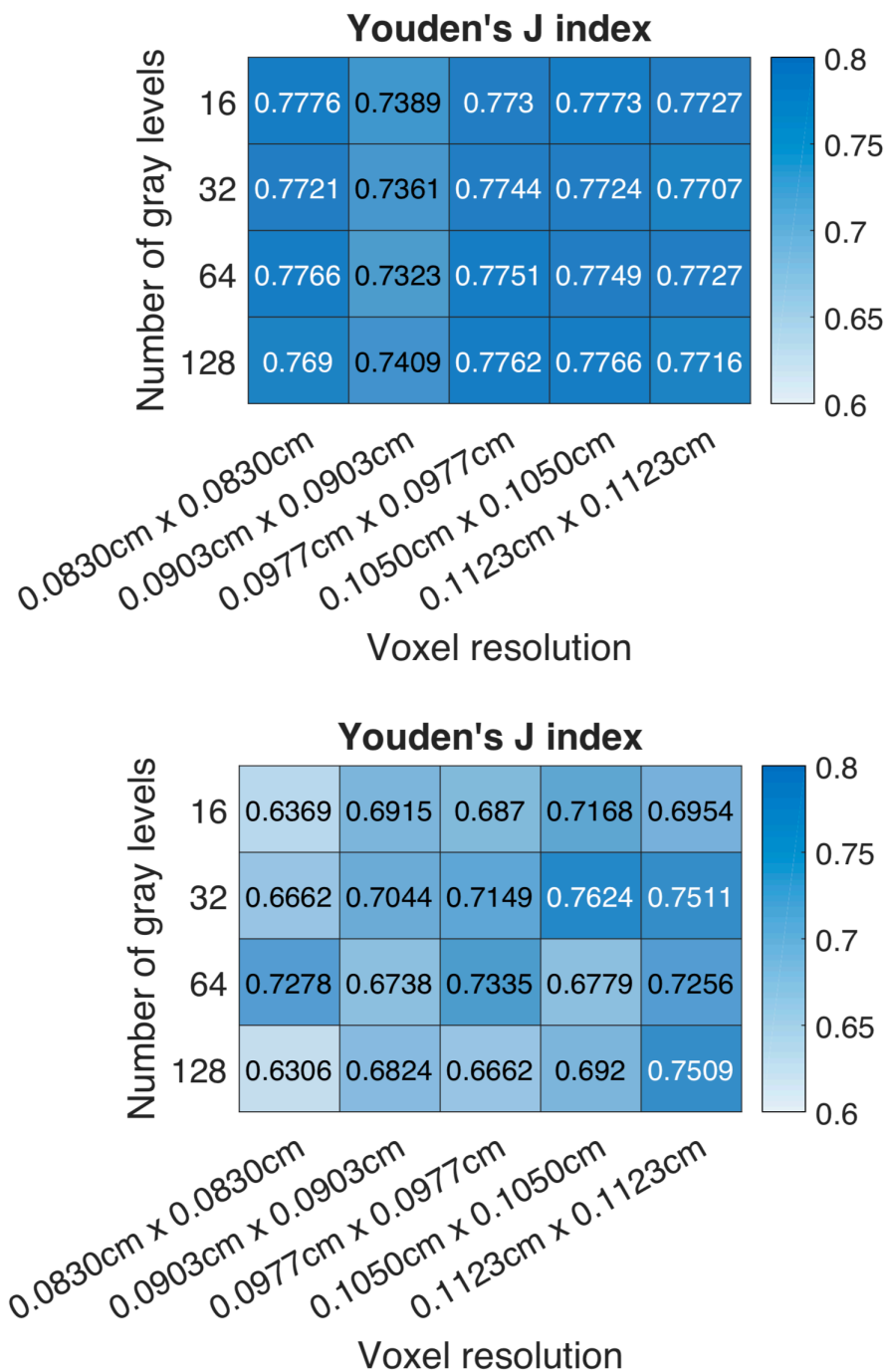
Fig. 2: Classification accuracy results using 100 traning data (50 noisy and 50 clean). Upper panel: Proposed OMT/spatial GLCM classifier. Lower: Random forest classifier.

the random forest method fails to achieve good performance for all the considered extraction parameter pairs, as depicted in Figure 3. Increasing the size of the training set leads to a significant increase in the accuracy of the random forest approach, while the proposed method achieves comparable good classification results (see Tables III and IV).

Size of training set	10 (5 noisy/5 clean) mean (std)	40 (20 noisy/20 clean) mean (std)	100 (50 noisy/50 clean) mean (std)	400 (200 noisy/200 clean) mean (std)
Sensitivity (test data)	0.8594 (0.0043)	0.8699 (0.0028)	0.8703 (0.0023)	0.8704 (0.0024)
Specificity (test data)	0.8840 (0.0148)	0.8949 (0.0172)	0.8962 (0.0166)	0.8975 (0.0153)
Youden's J index (test data)	0.7434 (0.0158)	0.7648 (0.0162)	0.7666 (0.0154)	0.7680 (0.0143)

TABLE III: Robustness of the classification results to the size of training dataset using the proposed OMT/spatial GLCM based classifier. The **mean** and **std** represent the average and standard deviation of the classification accuracy for the different combination pairs of GLCM extraction parameters.

Size of training set	10 (5 noisy/5 clean) mean (std)	40 (20 noisy/20 clean) mean (std)	100 (50 noisy/50 clean) mean (std)	400 (200 noisy/200 clean) mean (std)
Sensitivity (test data)	0.5660 (0.0821)	0.8741 (0.0204)	0.8848 (0.0154)	0.8909 (0.0103)
Specificity (test data)	0.5283 (0.1227)	0.7711 (0.0454)	0.8145 (0.0451)	0.8574 (0.0284)
Youden's J index (test data)	0.0943 (0.1639)	0.6452 (0.0357)	0.6994 (0.0363)	0.7483 (0.0254)

TABLE IV: Robustness of the classification results to the size of training dataset using random forest and GLCM scalar features. The **mean** and **std** represent the average and standard deviation of the classification accuracy for the different combination pairs of GLCM extraction parameters.

IV. DISCUSSION

In this study, we investigated two ideas related to texture analysis for medical images. The first idea is based on directly utilizing texture matrices, not their scalar summary statistics, as multivariate radiomic features. The second idea consists of using a robust metric, the so-called Wasserstein distance, on these high-dimensional objects to classify CT images with dental
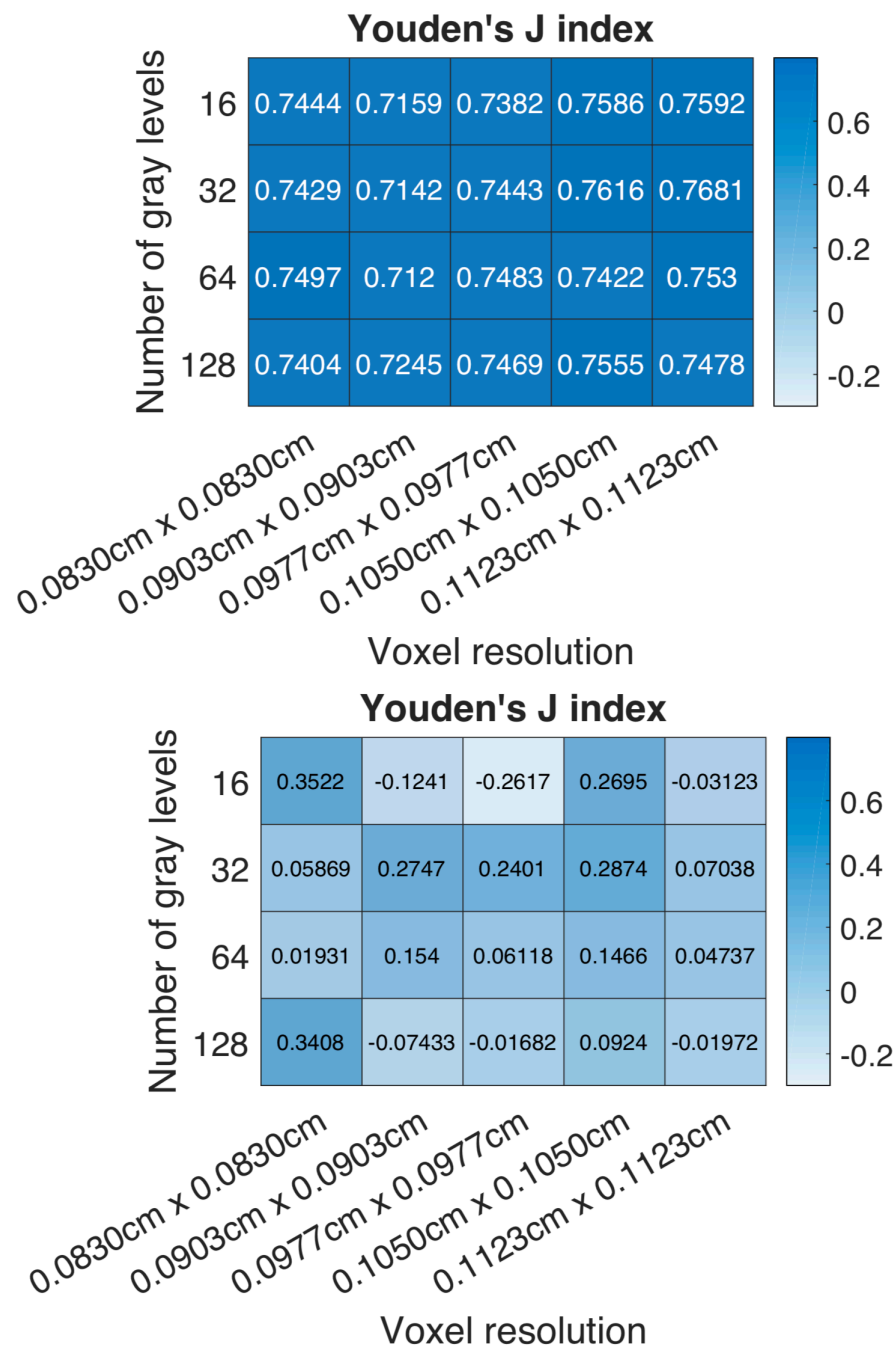
Youden's J index

Number of gray levels	0.0830cm × 0.0830cm	0.0903cm × 0.0903cm	0.0977cm × 0.0977cm	0.1050cm × 0.1050cm	0.1123cm × 0.1123cm
16	0.7444	0.7159	0.7382	0.7586	0.7592
32	0.7429	0.7142	0.7443	0.7616	0.7681
64	0.7497	0.712	0.7483	0.7422	0.753
128	0.7404	0.7245	0.7469	0.7555	0.7478

Voxel resolution

Youden's J index

Number of gray levels	0.0830cm × 0.0830cm	0.0903cm × 0.0903cm	0.0977cm × 0.0977cm	0.1050cm × 0.1050cm	0.1123cm × 0.1123cm
16	0.3522	-0.1241	-0.2617	0.2695	-0.03123
32	0.05869	0.2747	0.2401	0.2874	0.07038
64	0.01931	0.154	0.06118	0.1466	0.04737
128	0.3408	-0.07433	-0.01682	0.0924	-0.01972

Voxel resolution

Fig. 3: Classification accuracy results using 10 traning data (5 noisy and 5 clean). Upper panel: Proposed OMT/spatial GLCM classifier. Lower: Random forest classifier.

artifacts in H&NC. Combining these two ideas has resulted in a classification algorithm that is less sensitive to the GLCMs computation parameters. It also does not need a large number of training data to achieve high classification performance.

The potential of using the GLCM as a multivariate object for an oncological application recently introduced in [14]. In this work, the authors used a Gaussian Markov random field model to characterize GLCMs. Their modeling approach was employed to distinguish between benign and malignant adrenal lesions based on CT scans, and the obtained results were promising compared to different machine learning methods that were based on scalar feature quantities. The concept though of using multidimensional parametric models such as Markov random field to capture texture is old [26], [27], but never used in oncology to the best of our knowledge.

Even though there is no precise universal definition for texture [11], it still remains a key element of visual perception to analyze images in different fields, particularly the cancer research field. Therefore, more efforts need to be employed to go beyond the summary texture features by considering the multivariate structure of texture matrices. Doing so, hopefully, will avoid information loss of the functional patterns that describe the perfusion, density, or morphology of tumor microenvironment.

This work is the first proof-of-concept study that offers some new insights in dealing with texture to reveal the intrinsic heterogeneity observed in medical images. Further research investigations need to be conducted in the future to show the applicability and potential use of multivariate texture matrices (GLCM, RLM, SZM, NGTDM) and OMT theory for other purposes besides medical image classification, e.g., in outcome prediction and treatment planning. A potential future direction is to combine the geometrical texture features, which consists of the Wasserstein distance between texture lattices, with other types of radiomic features, e.g., shape, to build appropriate machine learning prediction models.

V. CONCLUSION

We developed a classification algorithm for 2D medical images based on the Wasserstein distance and the multivariate GLCMs. Overall, the proposed classifier approach outperformed the random forest method, which does not leverage the spatial information in GLCMs, in terms of robustness to the GLCM extraction parameters as well as the size of the training set.

ACKNOWLEDGMENTS

REFERENCES

[1] P. Lambin, E. Rios-Velazquez, R. Leijenaar, and et al., "Radiomics: extracting more information from medical images using advanced feature analysis," *European Journal of Cancer*, vol. 48, pp. 441–446, 2012.

[2] O. Morin, M. Vallières, A. Jochems, and et al., "A deep look into the future of quantitative imaging in oncology: A statement of working principles and proposal for change," *International Journal of Radiation Oncology*, vol. 102, pp. 1074–1082, 2018.

[3] J. C. Peeken, M. Bernhofer, B. Wiestler, and et al., "Radiomics in radiooncology-challenging the medical physicist," *European Journal of Medical Physics*, vol. 48, pp. 27–36, 2018.

[4] P. Lambin, R. T. Leijenaar, T. M. Deist, and et al., "Radiomics: the bridge between medical imaging and personalized medicine," *Nature Reviews Clinical Oncology*, vol. 14, pp. 749–762, 2017.

[5] R. Shiradkar, T. K. Podder, A. Algohary, and et al., "Radiomics based targeted radiotherapy planning (Rad-TRaP): a computational framework for prostate cancer treatment planning with MRI," *Radiation Oncology*, vol. 11, pp. 1–14, 2016.

[6] J. Meng, S. Liu, L. Zhu, and et al., "Texture analysis as imaging biomarker for recurrence in advanced cervical cancer treated with CCRT," *Scientific reports*, vol. 8, pp. 1–9, 2018.

[7] Z. Yan, J. Zhang, H. Long, and et al., "Correlation of CT texture changes with treatment response during radiation therapy for esophageal cancer: An exploratory study," *PLoS ONE*, vol. 14, pp. 1–12, 2019.

[8] E. Scalco and G. Rizzo, "Texture analysis of medical images for radiotherapy applications," *British institute of radiology*, vol. 90, pp. 1–15, 2016.

[9] A. L. Vickers and J. W. Modestino, "A maximum likelihood approach to texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-4, pp. 61–68, 1982.

[10] F. Albregtsen, "Statistical texture measures computed from gray level coocurrence matrices," *Department of Informatics, University of Oslo*, 2008.

[11] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019.

[12] A. Kassner and R. E. Thornhill, "Texture analysis: A review of neurologic MR imaging applications," *Journal of digital imaging*, vol. 31, pp. 809–816, 2010.

[13] M. S. ul Hassan and et al., "Intrinsic dependencies of ct radiomic features on voxel size and number of gray levels," *Medical Physics*, vol. 44, pp. 1050–1062, 2017.

[14] X. Li, M. Guindani, C. S. Ng, and B. P. Hobbs, "Spatial bayesian modeling of glcm with application to malignant lesion characterization," *Journal of applied statistics*, vol. 46, pp. 230–246, 2019.

[15] T. Löfstedt, P. Brynolfsson, T. Asklund, T. Nyholm, and A. Garpebring, "Gray-level invariant haralick texture features," *PLoS ONE*, vol. 14, pp. 1–18, 2019.

[16] A. Garpebring, P. Brynolfsson, P. Kuess, D. Georg, T. H. Helbich, T. Nyholm, and T. Loöfstedt, "Density estimation of grey-level co-occurrence matrices for image texture analysis," *Physics in medicine and biology*, vol. 63, pp. 1–14, 2018.

[17] C. Villani, *Optimal transport: old and new*. Springer-Verlag Berlin Heidelberg, 2009.

[18] G. Monge, *Mémoire sur la théorie des déblais et des remblais*. De l'Imprimerie Royale, 1781.

[19] L. V. Kantorovich, "On the translocation of masses," *Journal of mathematical sciences*, vol. 133, 2006.

[20] L. C. Evans, "Partial differential equations and monge-kantorovich mass transfer," in *Current Developments in Mathematics, International Press*, 1999.

[21] S. Kolouri, S. Park, M. Thorpe, and et al., "Optimal mass transport: signal processing and machine-learning applications (survey paper)," *IEEE Signal Processing Magazine*, vol. 34, pp. 43–59, 2017.

[22] K. Clark, B. Vendt, and K. S. et al., "The cancer imaging archive (TCIA): Maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, pp. 1045–1057, 2013.

[23] A. P. Apte and et al., "Extension of CERR for computational radiomics: A comprehensive matlab platform for reproducible radiomics research," *Medical physics*, vol. 45, pp. 3713–3720, 2018.

[24] M. Beckmann, "A continuous model of transportation," *Econometrica: Journal of the Econometric Society*, vol. 129, pp. 643–660, 1952.

[25] J. Liu, W. Yin, W. Li, and Y. T. Chow, "Multilevel optimal transport: A fast approximation of Wasserstein-1 distances," *Available: https://arxiv.org/pdf/1810.00118.pdf*, 2018.

[26] G. R. Cross and A. K. Jain, "Markov random field texture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, pp. 25–39, 1983.

[27] M. Hassner and J. Sklansky, "The use of markov random fields as models of texture," *Computer Graphics and Image Processing*, vol. 12, pp. 357–370, 1980.