

Wasserstein-based texture analysis in radiomic studies

Zehor Belkhatir^{a,*}, Raúl San José Estépar^{b,c}, Allen R. Tannenbaum^{d,e}

^a School of Engineering and Sustainable Development, De Montfort University, Leicester, United Kingdom

^b Brigham and Women's Hospital, Boston, United States of America

^c Harvard Medical School, Boston, United States of America

^d Computer Science and Applied Mathematics & Statistics, Stony Brook University, NY, United States of America

^e Department of Medical Physics, Memorial Sloan Kettering Cancer Center, NY, United States of America

ARTICLE INFO

Keywords:

Radiomics texture
Optimal mass transport
Reference samples
Spatial texture features
Supervised classification
Wasserstein metric
Bayesian optimization

ABSTRACT

The emerging field of radiomics that transforms standard-of-care images to quantifiable scalar statistics endeavors to reveal the information hidden in these macroscopic images. The concept of texture is widely used and essential in many radiomic-based studies. Practice usually reduces spatial multidimensional texture matrices, e.g., gray-level co-occurrence matrices (GLCMs), to summary scalar features. These statistical features have been demonstrated to be strongly correlated and tend to contribute redundant information; and does not account for the spatial information hidden in the multivariate texture matrices. This study proposes a novel pipeline to deal with spatial texture features in radiomic studies. A new set of textural features that preserve the spatial information inherent in GLCMs is proposed and used for classification purposes. The set of the new features uses the Wasserstein metric from optimal mass transport theory (OMT) to quantify the spatial similarity between samples within a given label class. In particular, based on a selected subset of texture GLCMs from the training cohort, we propose new representative spatial texture features, which we incorporate into a supervised image classification pipeline. The pipeline relies on the support vector machine (SVM) algorithm along with Bayesian optimization and the Wasserstein metric. The selection of the best GLCM references is considered for each classification label and is performed during the training phase of the SVM classifier using a Bayesian optimizer. We assume that *sample fitness* is defined based on *closeness* (in the sense of the Wasserstein metric) and *high correlation* (Spearman's rank sense) with other samples in the same class. Moreover, the newly defined spatial texture features consist of the Wasserstein distance between the optimally selected references and the remaining samples. We assessed the performance of the proposed classification pipeline in diagnosing the coronavirus disease 2019 (COVID-19) from computed tomographic (CT) images. To evaluate the proposed spatial features' added value, we compared the performance of the proposed classification pipeline with other SVM-based classifiers that account for different texture features, namely: statistical features only, optimized spatial features using Euclidean metric, non-optimized spatial features with Wasserstein metric. The proposed technique, which accounts for the optimized spatial texture feature with Wasserstein metric, shows great potential in classifying new COVID CT images that the algorithm has not seen in the training step. The MATLAB code of the proposed classification pipeline is made available. It can be used to find the best reference samples in other data cohorts, which can then be employed to build different prediction models.

1. Introduction

The importance of medical imaging has expanded to all major areas of health care. Motivated by the goal of unveiling the information hidden in standard-of-care images, the research field of radiomics has emerged (Lambin et al., 2012). 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. For instance, in outcome prediction, tumor

classification, treatment planning, and personalized therapy (Morin et al., 2018; Peeken et al., 2018; Lambin et al., 2017; Shiradkar et al., 2016) 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., Meng et al. (2018), Yan et al. (2019), Scalco and Rizzo (2016). A variety of mathematical models have been proposed to evaluate the gray level and the position of the pixels within an image (Hung et al., 2019). The

* Corresponding author.

E-mail address: zehor.belkhatir@dmu.ac.uk (Z. Belkhatir).

<https://doi.org/10.1016/j.compmedimag.2022.102129>

Received 5 February 2022; Received in revised form 11 August 2022; Accepted 3 October 2022

Available online 19 October 2022

0895-6111/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

predominant process used for texture analysis involves mapping the image domain into a gray-level co-occurrence matrix (GLCM), which is known as the *second-order histogram method* and which was introduced in 1973 by Haralick et al. (1973). GLCM is a matrix defined over an image domain with cells comprised of counts that describe the distribution of co-occurring gray-scale valued pixels (or voxels for 3D images) at a given offset and angle. More specifically, the (i, j) th entry of a GLCM represents the frequency with which a pixel with gray-level i was present in a spatial location either horizontally, vertically, or diagonally to adjacent pixel with gray-level j . GLCMs are often analyzed through summary statistics which effectively reduce the lattice count data to sets of GLCM-based textural features. Several authors have noted that the resulting summary statistics are strongly correlated with each other, resulting in over-fitting, and may also lead to a loss of the spatial information that is inherent in these texture matrices (Vickers and Modestino, 1982; Li et al., 2019; Belkhatir et al., 2020). Consequently, current practices invite scrutiny, challenge reproducibility, and encourage innovation.

The increased interest in using 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 for providing various data distributions with different geometric interpretations, which we seek to capture from multidimensional texture matrices in the present work. In particular, the OMT problem seeks the most efficient way to transform one distribution of mass to another given a cost function (Villani, 2009). Its origin goes back to 1781 when Gaspar Monge formulated the problem of finding the minimal transportation cost to redistribute earth for building fortifications (Monge, 1781). Leonid Kantorovich, in 1942, relaxed Monge's formulation to find an optimal coupling of distributions using linear programming (Kantorovich, 2006). Since then, OMT has played a crucial role in many fields of science and engineering; see Evans (1999), Kolouri et al. (2017) and references therein.

Our proof-of-concept study presented in Belkhatir et al. (2020) shows that considering GLCM as a non-parametric multivariate object improves the robustness of classifying even small datasets of CT images when combined with the Wasserstein metric. Although simple to implement, the pipeline in Belkhatir et al. (2020) does not provide insight into the classification domain as it does not explicitly train and learn a prediction model and relies on local decisions that emerge from comparing a new sample's features with stored feature samples. Such a local decision pipeline can be inefficient and insufficient to distinguish between different classes with overlapping or duplicate feature values. In the present work, however, we learn a predictive supervised machine learning model based on the combination of statistical texture features and newly defined optimized spatial (multidimensional) texture features. The study is based on the assumption that *representative* samples exist, which we refer to as *references* as well, i.e., "good" or "bad" samples representing each class/data cohort in a given classification/regression problem. Therefore, we propose here an optimized set of new spatial features, which are represented as the distance between all remaining samples in the cohort and the selected reference samples. In Belkhatir et al. (2020), however, we consider the Wasserstein distance between a new sample and all other stored samples, which is computationally heavy, especially when dealing with high-dimensional data.

The proposed spatial features are captured by the Wasserstein distance, from optimal mass transport (OMT) theory (Villani, 2009), between the optimally selected reference GLCMs and the texture matrices of other samples in the given class. The selection of the reference samples is conducted in the training phase using a Bayesian optimization algorithm along with a support vector machine (SVM) classifier. A natural application of this advanced textural classification approach is in diseases with specific patterns. As an example, coronavirus disease 2019 (COVID-19) related pneumonia, an infectious disease caused by

the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has a prototypical radiological appearance that can be amenable to illustrate the virtues of the proposed methodology (Kaufman et al., 2020). In this study, we tested the performance of the supervised classification algorithm to diagnose COVID-19 using computed tomographic (CT) images effectively and showed the added value of the newly defined optimized spatial features by comparing the performance of the proposed pipeline to different SVM-based classifiers namely: statistical features only, optimized spatial features using Euclidean metric, non-optimized spatial features with Wasserstein metric. The pipeline can be applied to other radiomic-based applications where texture is an important feature to consider.

The remainder of the present paper is organized as follows. In Section 2, the different components of the proposed classification pipeline are explained along with the data cohort and its pre-processing. Section 3 displays and discusses the results of the proposed classification algorithm and compares them to other machine learning classification pipelines. The paper is concluded in Section 4.

2. Materials and method

In this section, we introduce the proposed classification algorithm and the data used to test its performance. The algorithm is tested for diagnosing COVID-19 cases from CT images. At the time this paper is being written, the disease has infected more than 40 million individuals worldwide and caused more than 1 million deaths.

2.1. Data preprocessing and lung segmentation

The COVID-19 CT images used in this study are publicly available.¹ The original dataset includes 349 CT scans that are COVID-19 positive and 397 that are COVID-19 negative, which are either normal or contain other types of diseases. The images were collected from 760 preprints about COVID-19, where the COVID19-positive scans are from 143 patient cases. The average, maximum, and minimum number of CT scans that a patient has is 1.6, 16.0, and 1.0 respectively. These CT images have different sizes. The average, maximum and minimum height are 491, 1853, and 153 respectively. The average, maximum, and minimum width are 383, 1485, and 124 respectively. Regarding the COVID19-negative scans, 202 are selected from the PubMed Central (PMC) search engine. The rest 195 come from MedPix, which is a publicly-open online medical image database that contains CT scans with various diseases. For further details about the collection process, we refer the interested reader to He et al. (2020).

The CT images were resampled to a fixed grid of 512×512 and also equalized to adapt their contrast. Moreover, the lung masks have been derived based on a contrast stretching with an output range between $(-1024, 226)$. The lung fields were segmented using a **modified UNet architecture** as described in Moreta-Martinez et al. (2020). This network consists of encoding and decoding convolution blocks using a DenseNet design (Huang et al., 2017). To avoid representation bottlenecks, the network adopts an Efficient Net (ENet) representation (Paszke et al., 2016) that combines both max-pooling and strided convolutions on the same input. The architecture also uses a squeeze and excitation block in the encoder and decoder blocks to perform activation recalibration to enhance relevant features while suppressing irrelevant ones (Roy et al., 2018). The network was trained with 2D slices from a randomly selected set of CT scans spanning a range of pathologies, including interstitial abnormalities that resembled some of the radiological patterns that can be found in patients with COVID-19 pneumonia. The pre-trained network is available in the Chest Imaging Platform open source toolkit (Estépar et al., 2015). This method was specifically designed to work without the need of volumetric data.

¹ <https://github.com/UCSD-AI4H/COVID-CT>

Table 1
GLCM-based scalar statistical features.

1 : Energy	10 : Auto-correlation	18 : Inverse difference moment
2 : Joint entropy	11 Inverse variance	19 : Inverse difference moment normalized
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 : Inverse difference normalized
6 : Contrast	15 : Sum entropy	23 : First measure of information correlation
7 : Cluster tendency	16 : Sum variance	24 : Second measure of information correlation
8 : Cluster shade	17 : Correlation	25 : Dissimilarity
9 : Inverse difference		

The code and model of the employed segmentation technique are fully public and available in.² After applying the lung segmentation pipeline in all images, non-valid segmentations were filtered out and removed, resulting in 150 images for non-COVID patients and 174 images for COVID patients.

2.2. Texture extraction

The GLCM is a two-dimensional matrix in which each element represents the frequency of occurrences of a pair of pixels in a spatial relation separated by a distance d and an angle θ . In our study, GLCM texture matrices were extracted automatically from the segmented lungs using the radiomics extension of the Computational Environment for Radiological Research (CERR) (Apte et al., 2018). They were computed by combining contributions from all 2-D neighbors (i.e., $d = 1$ and $\theta_i = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$), using a gray quantization level value of 32. Additionally, 25 scalar statistical features, which are listed in Table 1, were extracted from the GLCMs to be used in the proposed supervised classifier in addition to the newly defined spatial texture features.

2.3. Optimal mass transport and Wasserstein distance

We use the Wasserstein-1 distance from OMT theory (Villani, 2009) to capture the geometrical properties that are inherent in the GLCM texture matrices. The Wasserstein-1 distance between two d -dimensional probability distributions ρ_0, ρ_1 defined on $\Omega \subset \mathbb{R}^d$ is defined as follows:

$$\begin{aligned} & \inf_{\pi: \Omega \times \Omega \rightarrow \mathbb{R}} \int_{x,y \in \Omega} \|x - y\|_p \pi(x, y) dx dy, \\ & \text{subject to } \int_{y \in \Omega} \pi(x, y) dy = \rho_0(x), \quad \forall x \in \Omega \\ & \int_{x \in \Omega} \pi(x, y) dx = \rho_1(y), \quad \forall y \in \Omega \\ & \pi(x, y) \geq 0, \quad \forall x, y \in \Omega \end{aligned} \quad (1)$$

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

$$\begin{aligned} & \inf_{m: \Omega \rightarrow \mathbb{R}^d} \int_{x \in \Omega} \|m(x)\|_1 dx, \\ & \text{subject to } \nabla \cdot (m(x)) = \rho_0(x) - \rho_1(x), \quad \forall x \in \Omega \\ & m(x) \cdot n(x) = 0, \quad \forall x \in \partial\Omega \end{aligned} \quad (2)$$

where $n(x)$ denotes the normal to the boundary $\partial\Omega$, and m is a d -dimensional field satisfying the zero flux boundary condition (Beckmann, 1952). A fast numerical scheme that relies on multilevel primal-dual optimization algorithms was proposed in Liu et al. (2018) to solve (2). This latter numerical scheme is adopted in the present study.

2.4. Selection of representative texture matrices and supervised image classification algorithm

In this paper, we refer to the “good and bad” samples in a given dataset as representative or reference samples. Moreover, the selection of such representative samples is presented for a binary classification task. However, it can be easily adopted in multi-label classification problems or even regression problems.

The main novelty of our approach consists of proposing a new set of features relying on Wasserstein distance, and the 2-D texture matrices which are considered as probability distributions. The main assumption that underlies the proposed set of features is that there may exist reference samples, either good or bad, in a studied dataset cohort (training set) and those references will be used to assess their similarity to the remaining training and testing texture samples through the Wasserstein metric.

Fig. 1 depicts the devised workflow in our study. First, data were split into training (80%, 260 samples) and testing (20%, 65 samples) sets. Then, reference samples are selected from each class of the training set as explained in the following.

Let $T_i^1, i = 1 : m_1$, and $T_i^2, i = 1 : m_2$, denote the GLCM texture matrices from the binary classes C_1 and C_2 , respectively. We initially compute the Wasserstein-1 distance between all pairs T_i^1 from C_1 and all pairs T_j^2 from C_2 in the training set. The similarity measures are presented by the distance matrices $Dt_1 \in \mathbb{R}^{m_1 \times m_1}$ and $Dt_2 \in \mathbb{R}^{m_2 \times m_2}$. For each sample in a given class, we compute the correlation between its distance to all the remaining samples in that class through the Spearman's rank-order correlation given as follows:

$$\text{corr}(Dt_i) = 1 - \frac{6 \sum \Delta_i^2}{m_i(m_i^2 - 1)}, \quad i = 1, 2 \quad (3)$$

where $\Delta_i, i = 1, 2$ is the difference between the ranks of each pair of columns in Dt_1 and Dt_2 , and $m_i, i = 1, 2$ is the length of each column in $Dt_i, i = 1, 2$.

Then, we order the average Spearman's rank-order correlations of the GLCM samples $T_i^1|_{i=1}^{m_1}$ and $T_i^2|_{i=1}^{m_2}$. We refer to the “good” reference samples those having high average correlation ranks, and “bad” reference samples those with low average correlation ranks. Based on the computed averaged correlation ranks, we use Bayesian optimization algorithm to find the optimal number of good and bad samples (n_1^* and n_2^*) that maximizes the classification accuracy of the training set, or minimizes $(1 - \text{accuracy})$, as follows:

$$(n_1^*, n_2^*) = \arg \min_{0 \leq n_1, n_2 \leq U} (1 - f(n_1, n_2)), \quad (4)$$

where $f(n_1, n_2)$ is the training classification accuracy using n_1 higher ranked samples and n_2 lower ranked samples from the two classes C_1

² https://github.com/acil-bwh/ChestImagingPlatform/blob/develop/cip_python/dcn/projects/lung_segmenter/lung_segmenter_dcn.py

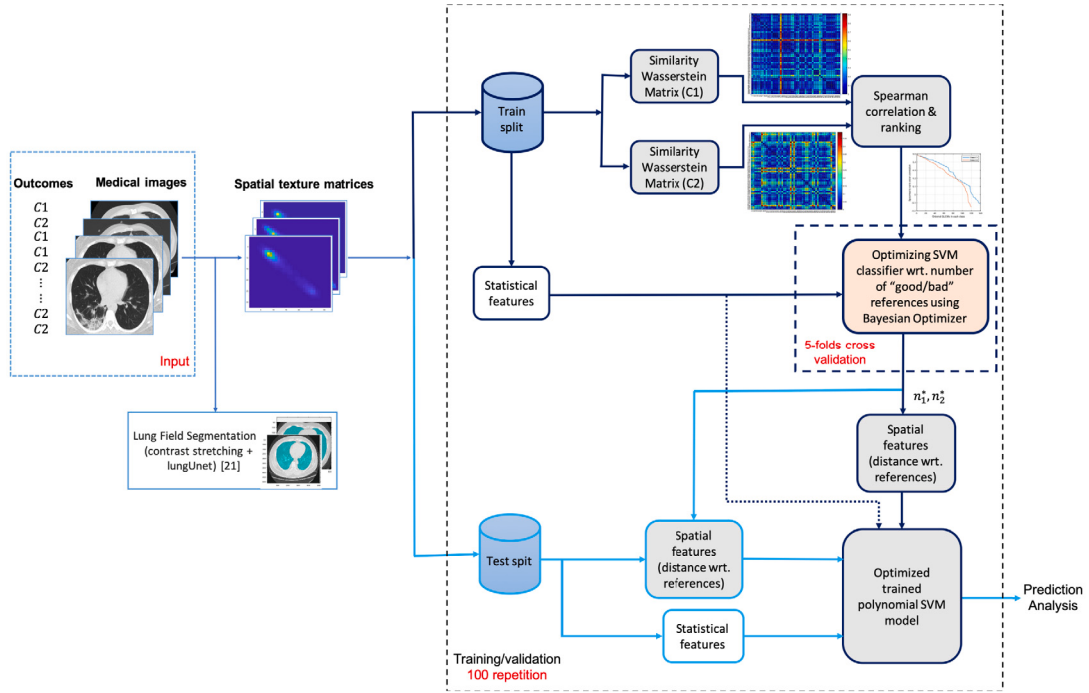


Fig. 1. Flowchart of the proposed supervised classification algorithm using representative spatial texture features.

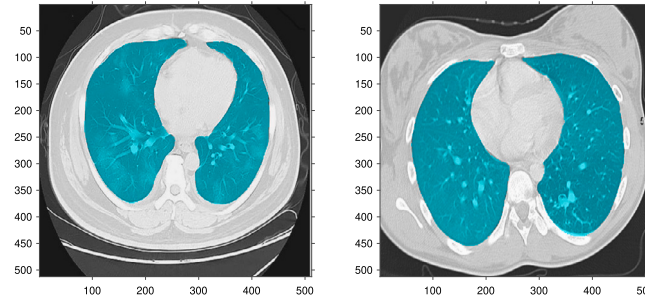


Fig. 2. Example CT images from a COVID (left panel) and Non-COVID (right panel) patients overlaid by the lung segmentation.

and C_2 . The maximum number of reference samples is U . The 5-fold cross-validation (CV) process (Berrar, 2019) is used in the training step to select the best reference samples. The training data is randomly split into five folds. Each time, four folds are used for training and the remaining fold for validation. The CV performance that is optimized in Eq. (4) is measured by the average accuracy of all five folds. Moreover, in addition to considering the validation step using the 5-fold CV the proposed pipeline considers the testing step using unseen data. This adopted strategy reduces the maximization bias problems that may arise from considering only training and/or validation sets.

We note that the selected references are removed from the training set that is used to train the optimal SVM classifier. So implicitly, the proposed optimization approach optimizes, in addition to the training accuracy, the *training efficiency* which is captured through the number of training samples employed to build the optimized prediction model after the reference samples have been removed. The optimized model which achieves the highest accuracy is employed to test the pipeline on new samples that were not seen in the training phase. Moreover, the Wasserstein distance between the selected reference samples and each sample from the remaining training samples is added to the statistical features described previously in Table 1. Finally, the procedure of learning and testing was replicated 100 times as shown in Fig. 1 to study the robustness of the classification model. We should emphasize that the added Wasserstein metric based features between the reference

and other training/testing samples captures the geometric similarity of their GLCM matrices to the good/bad representative GLCMs in each class. This is why we refer to these newly proposed features as *representative spatial texture features*.

A Bayesian optimizer has been chosen to solve the optimization problem (4) and find the optimal number of good/bad references in each class because of its advantages. This class of optimization techniques considered as a sequential design strategy trades off exploration and exploitation mechanisms, unlike traditional active learning, where the focus is often only on exploration. It attempts to find the global optimum of black-box functions; that do not usually assume any functional form; in a minimum number of steps (Frazier, 2018). This type of optimization approach has been adopted in different application fields (Shahriari et al., 2016). It has been also very successful for hyper-parameters optimization when building predictive machine/deep learning-based models (Snoek et al., 2012; Archetti and Candeliari, 2019).

3. Results and discussion

Before assessing the similarity between texture matrices (GLCMs) in the proposed pipeline, we segmented the lung fields in all the images. Examples of the obtained results for a patient diagnosed with COVID-19 and a non-diagnosed patient are shown in Fig. 2. Then, the 2D

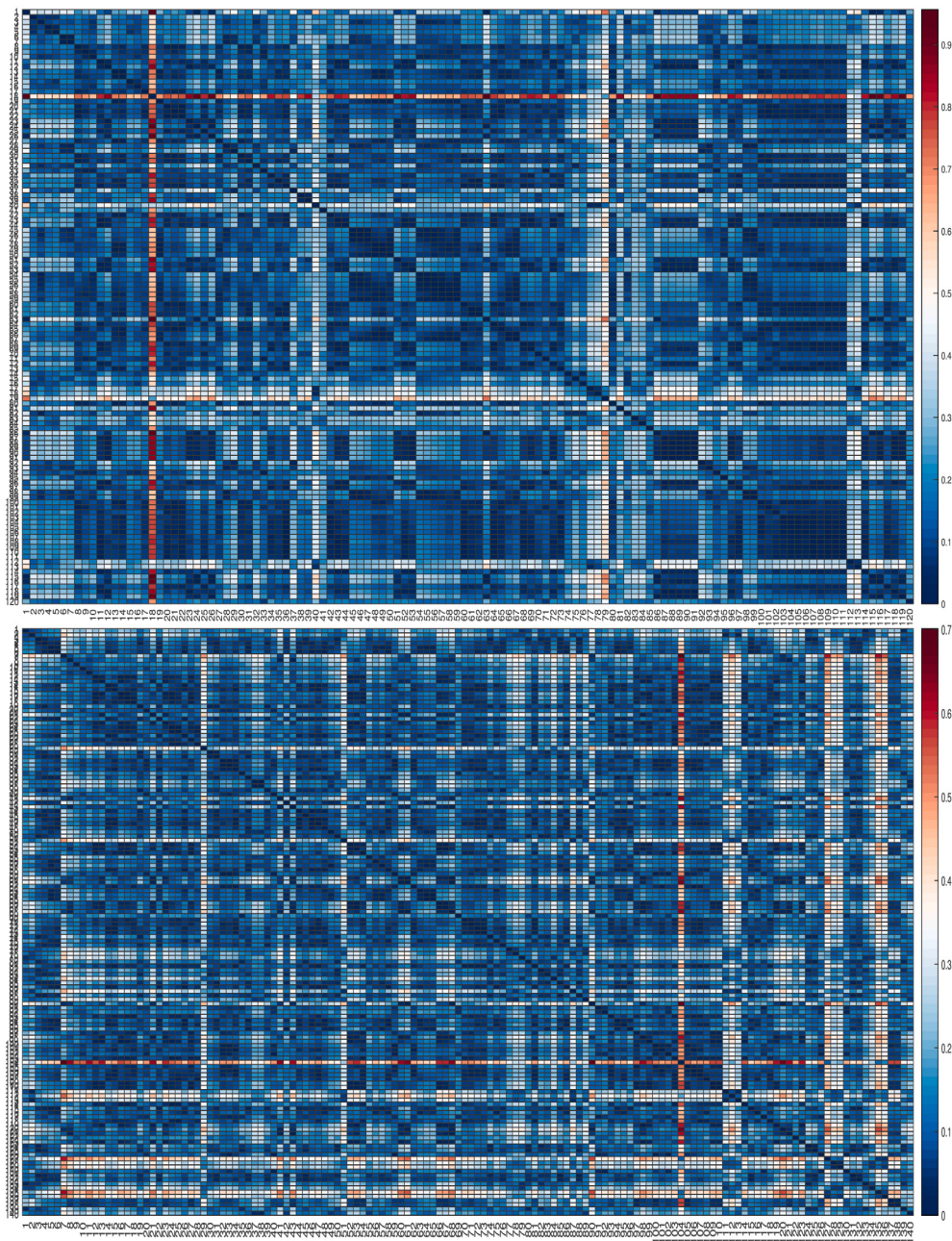


Fig. 3. Distance similarity matrices between all training pairs of samples in Class 1 “ Dr_1 ” (upper panel), and Class 2 “ Dr_2 ” (lower panel). The rows and columns show the patient id number.

GLCMs are computed within the lung fields, and similarity between all training pairs is assessed in each label class through the Wasserstein-1 distance. Fig. 3 illustrates the similarity distance matrix of class 1 and class 2. As shown in Fig. 3, some samples are closer in the sense of Wasserstein distance, compared to others, to some other samples within the same class. For instance, sample 18 from class 1 is further away from other samples in that class compared to the remaining samples. This observation has led us to the second step of our classification pipeline, which is to rank the similarity between samples within the same class through Spearman’s rank-order correlation. Then based on the ordered averaged correlation coefficient, we decide about the “good /bad” samples in a given class. Fig. 4 shows the ranked correlation coefficients for all training samples in C1 and C2. We focus on the n_1 samples from the tail of the ordered correlation vectors and n_2 samples from its head. We optimized the accuracy of the trained SVM model to select the best values of n_1 and n_2 , denoted n_1^* and n_2^* , using Bayesian

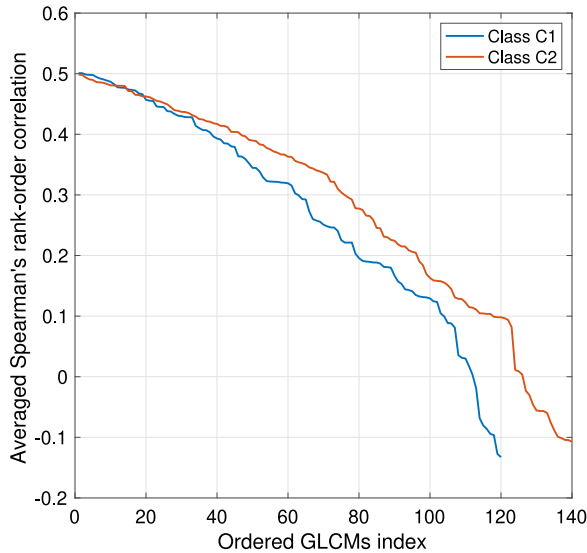
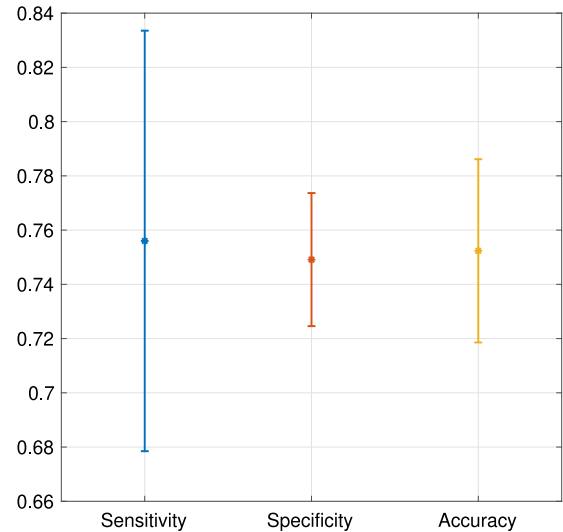
optimization and by performing 5-folds CV. Given the optimal values of reference samples in each class, we build an optimized classifier that accounts for the additional spatial texture features and will test its accuracy using the “no seen” samples. The averaged performance of the proposed classifier is shown in Fig. 5. The results of the best classifier, which achieves an area under the curve (AUC) value of 0.82225, are depicted in Fig. 6. The confusion matrix depicted in Fig. 6 shows that 78.125% of the total number of test samples were correctly classified.

The obtained results are very promising considering the size of the training set, which is not large, and also given that the base classifier that is used (a third order polynomial SVM) does not inherently rely of many hyper-parameters, which may reduce the over-fitting problem that usually exists in prediction models that rely on deeper structures. Moreover, to better appreciate the added value of the newly introduced spatial GLCM features, we compared our proposed supervised classifier

Table 2

Classification results for test (“never seen”) data. The **mean** and **std** represent the average and standard deviation of the classification statistics. **OpWspafeat**: SVM classifier with optimized spatial features based on Wasserstein metric, **OpL2spafeat**: SVM classifier with optimized spatial features based on Euclidean L2 metric, **Nospafeat**: SVM classifier without spatial features (only statistical features), **NOpWspafeat**: SVM classifier with non-optimized Wasserstein-based spatial features.

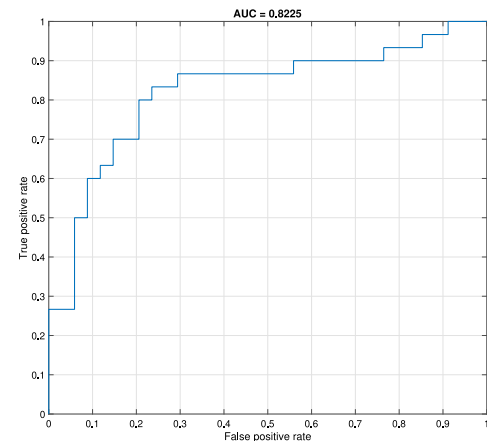
Method	OpWspafeat mean (std)	OpL2spafeat mean (std)	Nospafeat mean (std)	NOpWspafeat mean (std)
Sensitivity	0.7523 (0.0775)	0.6240 (0.0829)	0.6603 (0.0614)	0.6627 (0.0879)
Specificity	0.7491 (0.0245)	0.7256 (0.0168)	0.7329 (0.0220)	0.7106 (0.0820)
Accuracy	0.7560 (0.0338)	0.6780 (0.0412)	0.6989 (0.0304)	0.6881 (0.0604)

**Fig. 4.** Averaged Spearman's rank-order correlation for training Class 1 and Class 2.**Fig. 5.** Classification statistics among 100 training/testing repetitions, the mean value is highlighted with stars and standard deviation with bar plot.

that accounts for the newly defined texture features with other SVM-based classifiers that account for different types of texture features. Prediction performance for all the classifiers is provided in Table 2. The proposed pipeline shown in Fig. 1 is denoted **OpWspafeat**. The algorithm **OpL2spafeat** follows the same optimized pipeline except that the metric used to assess the closeness of GLCMs is the Euclidean L_2 distance instead of the Wasserstein distance. The **NOpWspafeat** algorithm consists of the SVM classifier that includes the statistical features and the non-optimized spatial features that consist of the Wasserstein distance between a given GLCM sample and all remaining samples. Last, the **Nospafeat** is the SVM classifier that uses only the statistical texture features. We observe from the table that the proposed classification pipeline outperforms other variants of SVM classifiers. This suggests the usefulness of both the optimization step in the training phase to select the reference GLCM samples and the Wasserstein metric to assess the closeness of GLCM samples. We note here that the superiority of the Wasserstein metric compared to L2 distance emerges from its ability to incorporate both spatial and intensity information when comparing images such as texture ones. Moreover, we took advantage of the robustness property that the Wasserstein metric offers compared to other distances or divergences.

The MATLAB code of the proposed classification pipeline is made publicly available.³ It can be used to find the best reference samples in other data cohorts and which can be added in building different prediction models.

It is worth noting the limitations and the potential future extensions of this study. Firstly, the use of the idea of defining reference samples



	Predicted Class 1	Predicted Class 2	total
True Class 1	25	5	30
True Class 2	9	25	34
Total	34	30	64

Fig. 6. Best Classification performance for the testing set. Upper panel: A receiver operating characteristic (ROC) curve for the proposed classifier with best area under the curve (AUC) value. Lower panel: Confusion matrix for the maximum achieved classification accuracy.

along with their fitness metric and its inclusion when analyzing high-dimensional data and building prediction models is worth generalizing in other applications beyond image classification. For instance, in outcome prediction and treatment planning, where in the latter, it is

³ <https://github.com/Supervised-image-classification-algorithm-using-representative-texture-features>

crucial to measure and know the dose difference between incoming patients and optimized reference patients having different survival outcomes. Secondly, we only made use of the GLCM texture matrices to prove the concept of the additional value of the reference samples and the newly defined optimized spatial texture features. Other texture matrices, such as the Run Length Matrix (RLM), size zone matrix (SZM), and/or neighborhood gray-tone difference matrix (NGTDM), can also be included in the pipeline and their effect on the classification accuracy investigated. Finally, it is evident from this study that capturing the fitness of samples to optimized reference samples using the Wasserstein distance improves image classification accuracy. However, it would be of great potential for clinical application to have a deeper analysis of the optimized reference images by expert radiologists and investigate the possibility of correlating them with other clinical outcomes.

4. Conclusion

In this study, we proposed a supervised classification pipeline for medical images that relies mainly on multidimensional GLCM-based texture features together with Bayesian optimization and the Wasserstein-1 metric. The proposed method uses an optimal subset of samples to represent the training set and to define a new set of geometric texture features based on the given optimal subset. The obtained results indicated the importance of the optimized reference samples' selection step and the newly defined spatial features in the classification workflow. The present work provides a way forward to more extensive studies with more data in order to build efficient prediction methods for analyzing medical images for various clinical scenarios. Moreover, the proposed pipeline for identifying representative texture matrices can be generalized to detect other types of reference samples from a given data cohort. The selected reference samples can be used for other purposes besides medical image classification, e.g., treatment planning and outcome prediction.

CRedit authorship contribution statement

Zehor Belkhatir: Designed and performed research, Developed the classification pipeline and its numerical implementation, Analyzed data and results, Writing of paper. **Raúl San José Estépar:** Pre-processed and segmented CT images, Analyzed data and results, Writing of paper. **Allen R. Tannenbaum:** Designed research, Analyzed data and results, Writing of paper.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Prof. Marc Niethammer for the discussions about lung segmentation, Prof. Joseph Deasy for the fruitful discussions about the proposed classification algorithm, and Dr. Saad Nadeem for discussions about texture analysis and the data used in this study.

This study was supported by the School of Engineering and Sustainable Development at the University of De Montfort, United Kingdom. Allen Tannenbaum was supported by AFOSR, United States of America grants (FA9550-17-1-0435, FA9550-20-1-0029), NIH, United States of America grant (R01-AG048769), MSK Cancer Center Support, United States of America Grant/Core Grant (P30 CA008748), and a grant from Breast Cancer Research Foundation, United States of America (BCRF-17-193).

References

- Apte, A.P., et al., 2018. Extension of CERR for computational radiomics: A comprehensive MATLAB platform for reproducible radiomics research. *Med. Phys.* 45, 3713–3720.
- Archetti, F., Candelieri, A., 2019. *Bayesian Optimization and Data Science*. Springer Briefs in Optimization.
- Beckmann, M., 1952. A continuous model of transportation. *Econometrica* 129, 643–660.
- Belkhatir, Z., Iyer, A., Mathews, J.C., Pouryahya, M., Nadeem, S., Deasy, J.O., Apte, A.P., Tannenbaum, A.R., 2020. Optimal mass transport for robust texture analysis. Available at: <https://www.biorxiv.org/content/10.1101/855221v1.full.pdf>.
- Berrai, D., 2019. Cross-validation. In: Ranganathan, S., Gribskov, M., Nakai, K., Schönbach, C. (Eds.), *Encyclopedia of Bioinformatics and Computational Biology*. Academic Press, Oxford, pp. 542–545. <http://dx.doi.org/10.1016/B978-0-12-809633-8.20349-X>, URL <https://www.sciencedirect.com/science/article/pii/B978012809633820349X>.
- Estépar, R.S.J., Ross, J., Harmouche, R., Onieva, J., Diaz, A., Washko, G., 2015. Chest imaging platform: An open-source library and workstation for quantitative chest imaging. In: *American Thoracic Society International Conference Abstracts American Thoracic Society*. p. A4975?A4975.
- Evans, L.C., 1999. Partial differential equations and Monge-Kantorovich mass transfer. In: *Current Developments in Mathematics*. International Press.
- Frazier, P.I., 2018. A tutorial on Bayesian optimization. Available at: <https://arxiv.org/pdf/1807.02811.pdf>.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst., Man Cybern.* 3, 610–621.
- He, X., Yang, X., Zhang, S., Zhao, J., Zhang, Y., Xing, E., Xie, P., 2020. Sample-efficient deep learning for COVID-19 diagnosis based on CT scans. Available at: <https://www.medrxiv.org/content/10.1101/2020.04.13.20063941v1.full.pdf>.
- Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. CVPR, p. 2261?2269.
- Hung, C., Song, E., Lan, Y., 2019. *Image Texture Analysis: Foundations, Models and Algorithms*. Springer International Publishing.
- Kantorovich, L.V., 2006. On the translocation of masses. *J. Math. Sci.* 133.
- Kaufman, A., Naidu, S., Ramachandran, S., Kaufman, D., Fayad, Z., Mani, V., 2020. Review of radiographic findings in COVID-19. *World J. Radiol.* 12, 142–155.
- Kolouri, S., Park, S., Thorpe, M., et al., 2017. Optimal mass transport: Signal processing and machine-learning applications (survey paper). *IEEE Signal Process. Mag.* 34, 43–59.
- Lambin, P., Leijenaar, R.T., Deist, T.M., et al., 2017. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* 14, 749–762.
- Lambin, P., Rios-Velazquez, E., Leijenaar, R., et al., 2012. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* 48, 441–446.
- Li, X., Guindani, M., Ng, C.S., Hobbs, B.P., 2019. Spatial bayesian modeling of GLCM with application to malignant lesion characterization. *J. Appl. Stat.* 46, 230–246.
- Liu, J., Yin, W., Li, W., Chow, Y.T., 2018. Multilevel optimal transport: A fast approximation of wasserstein-1 distances. Available: <https://arxiv.org/pdf/1810.00118.pdf>.
- Meng, J., Liu, S., Zhu, L., et al., 2018. Texture analysis as imaging biomarker for recurrence in advanced cervical cancer treated with CCRT. *Sci. Rep.* 8, 1–9.
- Monge, G., 1781. *Mémoire Sur La Théorie Des Déblais Et Des Remblais*. De l'Imprimerie Royale.
- Moreta-Martinez, R., Vegas-Sánchez-Ferrero, G., Andresen, L., Holsting, J.Q., Estépar, R.S.J., 2020. Multi-cavity heart segmentation in non-contrast non-ECG gated CT scans with F-CNN. *Lecture Notes in Comput. Sci.* 12502 (9), 14–23. http://dx.doi.org/10.1007/978-3-030-62469-9_2.
- Morin, O., Vallières, M., Jochims, A., et al., 2018. A deep look into the future of quantitative imaging in oncology: A statement of working principles and proposal for change. *Int. J. Radiat. Oncol.* 102, 1074–1082.
- Paszke, A., Chaurasia, A., Kim, S., Culurciello, E., 2016. ENet: A deep neural network architecture for real-time semantic segmentation. Available at: <https://arxiv.org/pdf/1606.02147.pdf>.
- Peeken, J.C., Bernhofer, M., Wiessler, B., et al., 2018. Radiomics in radiooncology—challenging the medical physicist. *Eur. J. Med. Phys.* 48, 27–36.
- Roy, A.G., Navab, N., Wachinger, C., 2018. Concurrent spatial and channel squeeze & excitation in fully convolutional networks. Available at: <https://arxiv.org/pdf/1803.02579.pdf>.
- Scalco, E., Rizzo, G., 2016. Texture analysis of medical images for radiotherapy applications. *Br. Inst. Radiol.* 90, 1–15.
- Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., de Freitas, N., 2016. Taking the human out of the loop: A review of Bayesian optimization. *Proc. IEEE* 104, 148–175.
- Shiradkar, R., Podder, T.K., Algohary, A., et al., 2016. Radiomics based targeted radiotherapy planning (rad-TRaP): A computational framework for prostate cancer treatment planning with MRI. *Radiat. Oncol.* 11, 1–14.

- Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical Bayesian optimization of machine learning algorithms. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), In: *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., pp. 2951–2959.
- Vickers, A.L., Modestino, J.W., 1982. A maximum likelihood approach to texture classification. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-4*, 61–68.
- Villani, C., 2009. *Optimal Transport: Old and New*. Springer-Verlag Berlin Heidelberg.
- Yan, Z., Zhang, J., Long, H., et al., 2019. Correlation of CT texture changes with treatment response during radiation therapy for esophageal cancer: An exploratory study. *PLoS One* 14, 1–12.