Reproducibility of radiomic features using network analysis and its application in Wasserstein *k*-means clustering

Jung Hun Oh,[a,*,†] **Aditya P. Apte**,[a,†] **Evangelia Katsoulakis**,[b] **Nadeem Riaz**,[c] **Vaios Hatzoglou**,[d] **Yao Yu**,[c] **Usman Mahmood**,[a] **Harini Veeraraghavan**,[a] **Maryam Pouryahya**,[a] **Aditi Iyer**,[a] **Amita Shukla-Dave**,[a] **Allen Tannenbaum**,[e,f] **Nancy Y. Lee**,[c] **and Joseph O. Deasy**[a]

[a]Memorial Sloan Kettering Cancer Center, Department of Medical Physics,
New York, United States
[b]Veterans Affairs, James A Haley, Department of Radiation Oncology,
Tampa, Florida, United States
[c]Memorial Sloan Kettering Cancer Center, Department of Radiation Oncology,
New York, United States
[d]Memorial Sloan Kettering Cancer Center, Department of Radiology, New York, United States
[e]Stony Brook University, Department of Computer Science, Stony Brook,
New York, United States
[f]Stony Brook University, Department of Applied Mathematics and Statistics,
Stony Brook, New York, United States

Abstract

Purpose: The goal of this study is to develop innovative methods for identifying radiomic features that are reproducible over varying image acquisition settings.

Approach: We propose a regularized partial correlation network to identify reliable and reproducible radiomic features. This approach was tested on two radiomic feature sets generated using two different reconstruction methods on computed tomography (CT) scans from a cohort of 47 lung cancer patients. The largest common network component between the two networks was tested on phantom data consisting of five cancer samples. To further investigate whether radiomic features found can identify phenotypes, we propose a *k*-means clustering algorithm coupled with the optimal mass transport theory. This approach following the regularized partial correlation network analysis was tested on CT scans from 77 head and neck squamous cell carcinoma (HNSCC) patients in the Cancer Imaging Archive (TCIA) and validated using an independent dataset.

Results: A set of common radiomic features was found in relatively large network components between the resultant two partial correlation networks resulting from a cohort of lung cancer patients. The reliability and reproducibility of those radiomic features were further validated on phantom data using the Wasserstein distance. Further analysis using the network-based Wasserstein *k*-means algorithm on the TCIA HNSCC data showed that the resulting clusters separate tumor subsites as well as HPV status, and this was validated on an independent dataset.

Conclusion: We showed that a network-based analysis enables identifying reproducible radiomic features and use of the selected set of features can enhance clustering results.

Keywords: radiomics; reproducibility; Wasserstein distance; network-based Wasserstein *k*-means clustering.

Paper 20214SSR received Aug. 13, 2020; accepted for publication Apr. 2, 2021; published online Apr. 30, 2021.

[*]Address all correspondence to Jung Hun Oh, ohj@mskcc.org

[†]These authors contributed equally to this work.

1 Introduction

Radiomics enables an in-depth measurement of tumor phenotypes by quantifying imaging signals from radiologic images that can reflect key information of signatures associated with patient outcomes.[1,2] Recently, the close connection of radiomics with machine learning has accelerated the development of new imaging features and radiomic outcomes modeling, showing the potential of using radiomics to build predictive models of individual cancer outcomes.[3,4] Despite such great progress in radiomics in recent years, the development of computational techniques to identify repeatable and reproducible radiomic features remains challenging and relatively unadvanced.[5] This has led to the lack of success of many radiomic models in subsequent external validation on independent data, impairing the reliability of the models.[6,7] One of the reasons for this is likely due to the susceptibility of radiomic features to image reconstruction and acquisition parameters.[8,9] Since radiomic features are computed via multiple tasks including imaging acquisition, segmentation, and feature extraction, the selection of parameters present in each step may affect the stability of features computed.[10] As such, prior to model building, the development of radiomic features with high repeatability and high reproducibility as well as the development of tools that can identify such features is more likely to be urgently needed in the field of radiomics.

In this study, for the identification of reliable and reproducible radiomic features, we propose a graph (network)-based computational method, consisting of a partial correlation network analysis and the graphical lasso (linear absolute shrinkage and selection operator). We demonstrate its potential to identify reproducible radiomic features, using computed tomography (CT) images in lung cancer patients and validating the findings on phantom data. We employ the W_1-Wasserstein distance (also known as Earth Mover's distance: EMD) as a quantitative metric to assess the reproducibility of radiomic features. To investigate whether radiomic features found by the regularized partial correlation network analysis can identify phenotypes, we further propose a method, the network-based Wasserstein k-means (NWK) clustering algorithm, to identify subgroups of tumors, employing the Wasserstein distance as a cost function in the conventional k-means algorithm. The method is tested on CT images in head and neck squamous cell carcinoma (HNSCC) patients downloaded from The Cancer Imaging Archive (TCIA) and validated using an independent dataset.

The hypothesis behind the proposed methodology is that radiomic features that retain the strong correlation with other features in a network form after removing spurious or false positive connections are likely to be reliable and reproducible.

2 Methods

2.1 *Radiomics on CT Scans in Lung Cancer Patients*

This study was approved by the Internal Review Board (IRB). In total, 47 lung tumors were segmented on contrast-enhanced CT images of lung cancer patients that were scanned using a GE MEDICAL SYSTEMS scanner and reconstructed with standard and lung convolution kernels. Initially, lung tumors were segmented on scans belonging to the lung reconstruction and were copied over to scans belonging to the standard reconstruction. For each reconstruction method, a set of 132 radiomic features was extracted using the CERR radiomics toolbox.[11] Extracted features are categorized into three groups: (1) first-order statistics, (2) shape-based features, and (3) higher order texture features including gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), neighborhood gray tone difference matrix (NGTDM), and neighboring gray-level dependence matrix (NGLDM).

2.2 *Phantom Data*

A multimaterial 3D printer (PolyJet Objet 260 Connex 3, Stratasys, Eden Prairie, Minnesota) with Voxel Print software was used for the deposition of droplets of ultraviolet-curable photopolymer resins in a layer-by-layer manner.[12,13] In this study, two base photopolymer resins were employed including TangoPlus (material A) and VeroWhite (material B), which range in attenuation of ~65 Hounsfield unit (HU) to 125 HU at 120 kVp. The resolution of the printer is on the order of $48 \times 84 \times 30$ μm. Because the resolution is finer than that of typical CT scanners used at our institution ($0.625 \times 0.625 \times 0.625$ mm),[12] multiple resin droplets were mixed in a single voxel to reproduce the contrast differences of tumor-specific patterns observed on patient CT scans. To 3D print, the fine gradients of tumor intensity patterns, the Floyd Steinberg dithering algorithm was used.

As a first step to produce phantoms, two separate 3D prints were developed. The first 3D print was modeled after a single patient from the RIDER dataset that consisted of patients with nonsmall-cell lung carcinoma (NSCLC). It was printed to physically visualize the lung vasculature, extent of the tumor, and its structure. The second 3D print was designed to capture the morphologic appearance and different intensity patterns seen on CT scans for the single NSCLC and four pancreatic ductal adenocarcinoma tumors. Since the surrounding tissue influences the local resolution and noise properties of tumors, the lesions were immersed in a heterogeneous background that was modeled after patients with advanced stage hepatic cirrhosis. The area and number of slices of the final 3D print were dictated by the size of the largest tumor.

Prior to dithering, the HU values of each tumor and background were separately converted to double precision intensities within a range of 0 to 1. The decimal values in the new intensity range dictated the amount of photopolymer material that would be deposited within any given voxel. Then, each slice was supersampled to the resolution of the 3D printer using the Whittaker–Shannon (SINC) interpolation.

Lastly, each slice was dithered to yield a set of layered raster images using the Floyd–Steinberg dithering algorithm. A single raster layer encodes the spatial allocation of a resin material. Since the 3D printer is capable of depositing three different resin materials, three sets of raster files were generated. Within any raster layer, a value of 1 indicates the deposition of material A and a value of 0 indicates that no material is deposited. The first set of bitmaps was inverted to mix two materials within a single voxel such that a value of 0 in the first set of bitmaps (material A) has a value of 1 in the second set of bitmaps (material B). Since two materials with opposing densities can generate the desired pattern differences, the third set of bitmaps consisted of all zeros.

After 3D printing, the phantom was scanned sequentially 30 times with a typical abdominal CT protocol using the following parameters: 120 kVp, 280 mA, pitch of 0.984, reconstructed slice thickness of 5 mm, and a reconstruction interval of 5 mm. Images were reconstructed using the filtered backprojection algorithm. Two sets of 132 radiomic features were extracted using the CERR radiomics toolbox for the standard and lung reconstruction kernels.[11]

2.3 *CT Scans in Head and Neck Cancer Patients*

For further radiomic network analysis, pretreatment CT scans with IV contrast for HNSCC patients were downloaded from the TCIA.[14] The data were previously used in our other study.[15] Below we briefly introduce the preprocessing and radiomic feature extraction steps. In total, 188 available cases were imported into the Eclipse treatment planning system (Varian Medical System, Palo Alto, California) for segmentation. Before delineating lesions of interest (ROIs), samples that do not meet the inclusion criteria such as primary tumor size and image quality were excluded. For the evaluable scans that fulfill the inclusion criteria, the primary tumor on CT scans was manually delineated by a radiation oncologist and the delineation was independently confirmed by a neuroradiologist. The presence of CT artifacts was further assessed within the primary tumor. Slices with streak artifacts were not delineated and excluded from the analysis. If the proportion of slices with streak artifacts in each tumor was larger than 50% of the number of slices, the case was excluded from the study.[16] This quality test resulted in a set of 77 cases with 28 laryngeal, 11 oropharyngeal, and 38 oral cavity tumors. There were 13 HPV-positive and 64 HPV-negative tumors.[17]

Radiomic features were extracted using the CERR radiomics toolbox on resampled scans at the resolution of $0.6 \times 0.6 \times 3.5$ mm.[11] Due to the effect of subsampling slices, which is caused by dental artifacts, two-dimensional (2D) radiomic features were extracted: 104 radiomic features including first-order statistics and higher order texture features were computed from artifact-free slices.

Further feature stability and volume-independent tests were performed. For each scan, 100 independent datasets were generated, each of which consisted of 75% of the artifact-free slices and for each dataset 104 radiomic features were recomputed. Features with a median coefficient of variation >0.1 across all samples were removed. Features with high correlations with tumor volume (Spearman's correlation coefficient > 0.4) were also removed. These two tests resulted in 67 radiomic features that were used for subsequent analyses.

The radiomic analysis results on the TCIA data were validated using an independent dataset with 83 HNSCC patients treated at our institution. The validation cohort consisted of 1 laryngeal tumor, 31 oropharyngeal, and 51 oral cavity tumors, all with pretreatment CT scans with IV contrast. Among 32 patients with laryngeal and oropharyngeal tumors, 27 had HPV-positive and 5 had HPV-negative tumors. However, HPV status for 51 patients with oral cavity tumors was not available since HPV status on oral cavity tumors is not routinely obtained due to the low or rare prevalence of HPV positivity. We utilized the identical radiomic analysis pipeline (including segmentation, preprocessing, feature extraction, and network analysis) used in the analysis of the TCIA data.

2.4 *Regularized Partial Correlation Network*

For a network representation of radiomic features, we adopted a Gaussian graphical model of partial correlation coefficients in which each connection (link) in a network is represented as a partial correlation coefficient between two radiomic features (nodes) after conditioning on all other available features.[18] This network may be intractably complex including many spurious connections, particularly for data with numerous features. To remove potential false positives and make the network representation more interpretable for a meaningful understanding of the data, a lasso-type regularization (graphical lasso) was employed in which a tuning parameter λ controls the sparsity of the network by shrinking partial correlation coefficients.[19,20] More specifically, higher λ values make the network sparser whereas lower values make the network denser possibly with false-positive connections. To optimize the λ value that controls the tradeoff between keeping spurious connections and removing true connections in a network, we employed a method that optimizes the fit of the network to the data by minimizing the extended Bayesian information criterion (EBIC).[21] Lastly, weak connections whose absolute partial correlation coefficients are <0.2 were further removed since those are likely to be false positives.

2.5 *Wasserstein Distance*

The optimal mass transport (OMT) is an active research area with an ever-increasing growth in its application in numerous fields, including medical imaging analysis, statistical physics, machine learning, and genomics.[22–25] Here, we briefly describe the basic concepts underlying OMT. Let $P(\mathbb{R}^n)$ denote the space of probability measures on $\mathbb{R}^n$ with finite second moments. Then, using the Kantorovich relaxed formulation[26] of OMT, the Wasserstein distance (EMD) between $\mu, \nu \in P(\mathbb{R}^n)$ is defined as follows:

$$W_1(\mu,\nu) = \inf_{\pi\in\Pi(\mu,\nu)} \int_{\mathbb{R}^n\times\mathbb{R}^n} \|x-y\| \mathrm{d}\pi(x,y), \tag{1}$$

where $\Pi(\mu,\nu)$ denotes the set of all joint probability measures π on $\mathbb{R}^n \times \mathbb{R}^n$ whose marginals are μ and ν. A computationally efficient reformulation of the Wasserstein distance can be defined such that the flux vector $\mathbf{m} \in \mathbb{R}^n$ is optimized in the following manner:

$$W_1(\mu,\nu) = \inf_{\mathbf{m}} \left\{ \int_{\mathbb{R}^n} \|\mathbf{m}(x)\| \mathrm{d}x | \mu - \nu - \nabla \cdot \mathbf{m} = 0 \right\}, \tag{2}$$

where $\|\cdot\|$ is the Euclidean norm.

An alternative graph-theoretic formulation of Eq. (2) to compute the Wasserstein distance on a network modeled as a weighted graph is defined as follows:

$$W_1(\mu, \nu) = \min_u \left\{ \sum_{i=1}^{m} \|u_i\| | \mu - \nu - Du = 0 \right\}, \tag{3}$$

where m is the number of edges in a network, u_i are fluxes on the edges, and D is the incidence matrix with rows and columns indexed by the nodes and edges in the network such that every entry (i, k) is set to 1 if the node i is assigned to be the head of the edge k and is set to -1 if it is the tail of k. Using Eq. (3), we can compute the Wasserstein distance between two samples on a connected component of partial correlation network consisting of radiomic features. We used the CVX toolbox in the R language to optimize the OMT problem on a network.[27]

2.6 *Network-Based Wasserstein k-Means Algorithm*

The k-means algorithm is one of the most commonly used clustering algorithms that partition a given set of samples into c clusters, by minimizing the within-cluster sum of squares:

$$\operatorname{argmin} \sum_{i=1}^{c} \sum_{x_j \in C_i} \|x_j - \mu_i\|^2, \tag{4}$$

where μ_i is the mean of samples in cluster C_i.[28] Here, we propose a clustering method in which the cost function in the conventional k-means algorithm is replaced with the Wasserstein distance metric as shown in the following equation:

$$\operatorname{argmin} \sum_{i=1}^{c} \sum_{x_j \in C_i} W_1(x_j, \mu_i). \tag{5}$$

In this approach, during the process of k-means clustering, distances from each sequentially updated centroid in c clusters to samples on a fixed optimized radiomic network are computed and clustering of the samples is performed such that the within-cluster sum of Wasserstein distance is minimized. More specifically, the Wasserstein distance computed using Eq. (3) on a given network is used in the k-means algorithm and then centroids are updated. Based on the updated centroids, the Wasserstein distance is computed again and the centroids are newly computed. This process is iterated until the centroids do not change or the number of iterations is reached to the predefined maximum number. The k-means algorithm coupled with the Wasserstein distance intuitively enables us to cluster samples in a network form of the given data. We call this method the NWK clustering algorithm.

3 Results

3.1 *Regularized Partial Correlation Network*

Two different reconstruction kernels (lung and standard) were applied to CT scans for a cohort of 47 lung cancer patients, generating a set of 132 radiomic features for each reconstruction method. For each set of features, a radiomic network was constructed using regularized partial correlation coefficients, i.e., an integrated method of partial correlation network analysis and graphical lasso. For the network optimization, the EBIC was employed with a default parameter setting. As a result, two different radiomic networks were constructed. To remove potential false-positive (or week) relationships between nodes (features) in the networks, connections with absolute partial correlation coefficients <0.2 were further removed. This may disconnect the network, creating more network components (islands). Hereafter, we define a component as a connected set of nodes in a network.

Figure 1 shows the two largest network components for each reconstruction method with the number of links ≥9: Figs. 1(a) and 1(b) resulted from the standard reconstruction method whereas Figs. 1(c) and 1(d) resulted from the lung reconstruction method. In a comparison of Figs. 1(a) and 1(c), 10 radiomic features were common, including first-order statistics: entropy (10), shape: surface-to-volume ratio (39), GLRLM: run entropy (74), NGTDM: coarseness (82),

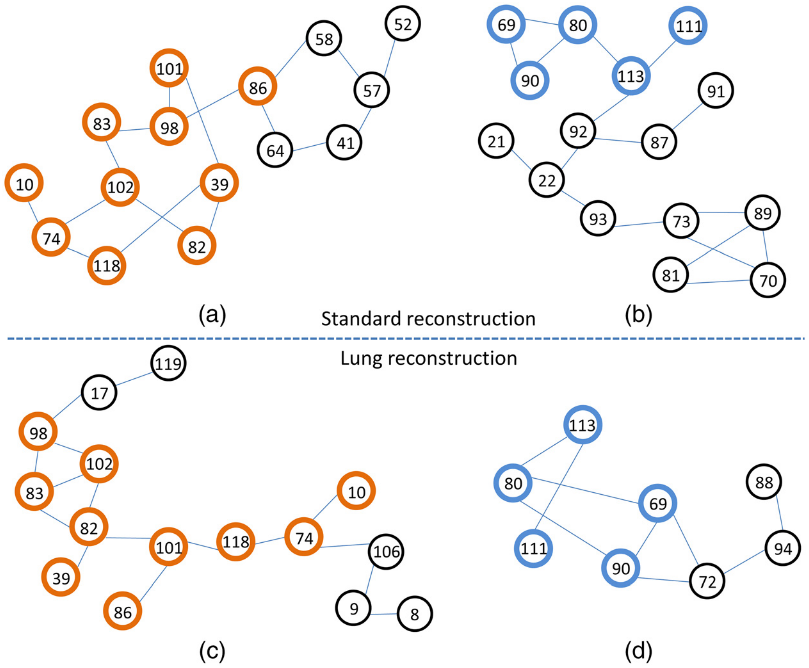

Fig. 1 The two largest network components from (a) and (b) the standard reconstruction kernel and from (c) and (d) the lung reconstruction kernel. The thick circles with the same color indicate the common radiomic features between the two reconstruction methods. The numbers in the circles indicate the order of 132 features in our data.

contrast (83), strength (86), NGLDM: dependence count nonuniformity normalized (98), dependence count entropy (101), dependence count energy (102), and GLSZM: size zone entropy (118). In Figs. 1(b) and 1(d), five radiomic features were common including GLRLM: high gray-level run emphasis (69), short run high gray-level emphasis (80), NGLDM: high gray-level count emphasis (90), GLSZM: high gray-level zone emphasis (111), and small area high gray-level emphasis (113). In the similarity test of network components using the hypergeometric distribution, statistically significant p-values were obtained with 2.3×10^{-8} between Figs. 1(a) and 1(c) and 4.2×10^{-4} between Figs. 1(b) and 1(d), showing the strong reproducibility of these radiomic features between different reconstruction methods. The direct similarity of network topologies was not compared, but it is noted that most of the radiomic features preserve the connections with other features as shown in Figs. 1(a) versus 1(c) and Figs. 1(b) versus 1(d).

3.2 *Validation on Phantom Data*

The reproducibility of those radiomic features identified using lung cancer data was further tested on phantom data. To do this, the two network components [Figs. 1(a) and 1(c)] each of which is the largest network component in each reconstruction method were combined, preserving the connections, which led to a connected network (Fig. 2) consisting of 20 radiomic features.

As previously described, five phantoms were generated in this study and for each phantom, two sets of radiomic features were extracted using the same two reconstruction kernels (standard and lung). Figure 3(a) shows a representative phantom slice used in this study. The same parameter setting in the CERR radiomics toolbox was used as in lung cancer analysis, generating 132 radiomic features for each set. Using Eq. (3), the Wasserstein distance was computed on the merged network (Fig. 2) between the two sets of 20 radiomic features for each phantom sample [Fig. 3(b)]. The average Wasserstein distance on the five phantoms was 0.21 (denoted as a reference for comparison with a simulation test shown in the following section) using the following equation:

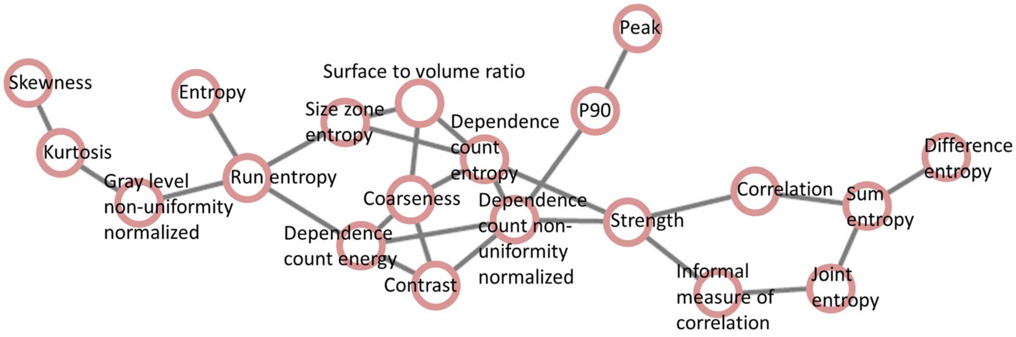

Fig. 2 A combined network constructed by merging two network components, each of which is the largest network component in each reconstruction method, i.e., Figs. 1(a) and 1(c).

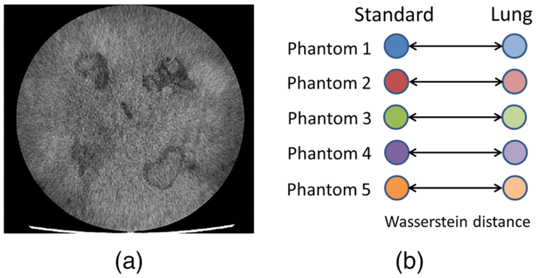

Fig. 3 (a) An example of phantom slice used in this study. (b) For each phantom, the Wasserstein distance was computed between a set of features of standard reconstruction and a set of features of lung reconstruction on a network constructed using lung cancer data.

$$\text{Average Wasserstein distance} = \frac{1}{5}\sum_{i=1}^{5} W_1(\text{Lung}_i, \text{Standard}_i). \tag{6}$$

It is likely that if the 20 radiomic features and their relationships (links) shown in Fig. 2 are reliable and stable, the Wasserstein distance computed using random radiomic features on the same network would be larger than 0.21 (reference). With this hypothesis, we randomly selected 20 features from the available 132 radiomic features and randomly assigned the 20 features to the 20 nodes in the network (Fig. 2) and then computed an average Wasserstein distance for five phantoms using Eq. (6). The whole process was repeated 1000 times, yielding 1000 average Wasserstein distance values. An overall average of the 1000 Wasserstein distance values was 0.32 and for only 49 times out of the 1000 simulation tests, the average Wasserstein distance was <0.21 (Fig. 4), implying the stability of the 20 radiomic features and their relationships in the network.

3.3 *Network-Based Wasserstein k-Means Clustering*

Using the TCIA HNSCC data, consisting of 77 samples and 67 radiomic features for each sample, a regularized correlation network was constructed. After applying the EBIC and cutting the links with partial correlation coefficients <0.2, a final network was built. The three largest network components with the number of links ≥9 were chosen for further analysis, which consisted of 26 radiomic features (Fig. 5).

Based on the silhouette criterion, the optimal number of clusters was 2. Using the NWK algorithm with $k = 2$, clustering was performed. During the k-means clustering process, the Wasserstein distance was computed for each network component and the three Wasserstein distance values were averaged. For the purpose of visualization, samples along with their clustering membership were represented on a low-dimensional space mapped from the 26 radiomic

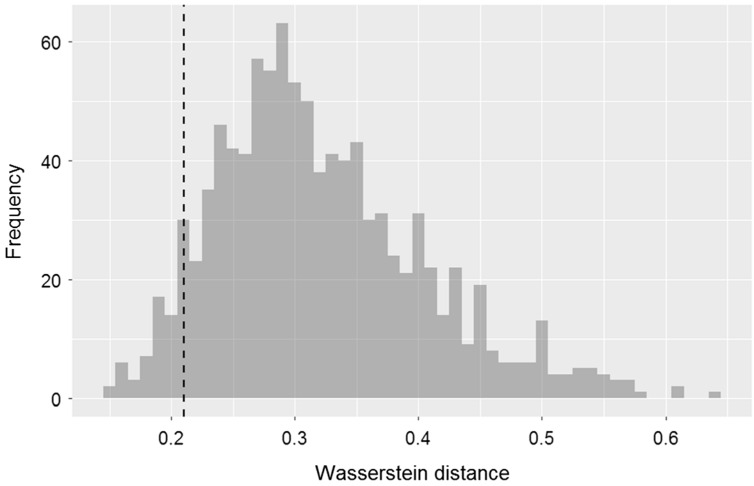

Fig. 4 On the network shown in Fig. 2, a random simulation test was performed, by randomly selecting 20 features from the available 132 radiomic features and randomly assigning the 20 features to the 20 nodes in the network. This histogram shows the results of Wasserstein distance after 1000 iterations between a set of features of standard reconstruction and a set of features of lung reconstruction. The dotted vertical line indicates 0.21 that is an average Wasserstein distance of the five phantoms computed on the original network shown in Fig. 2.

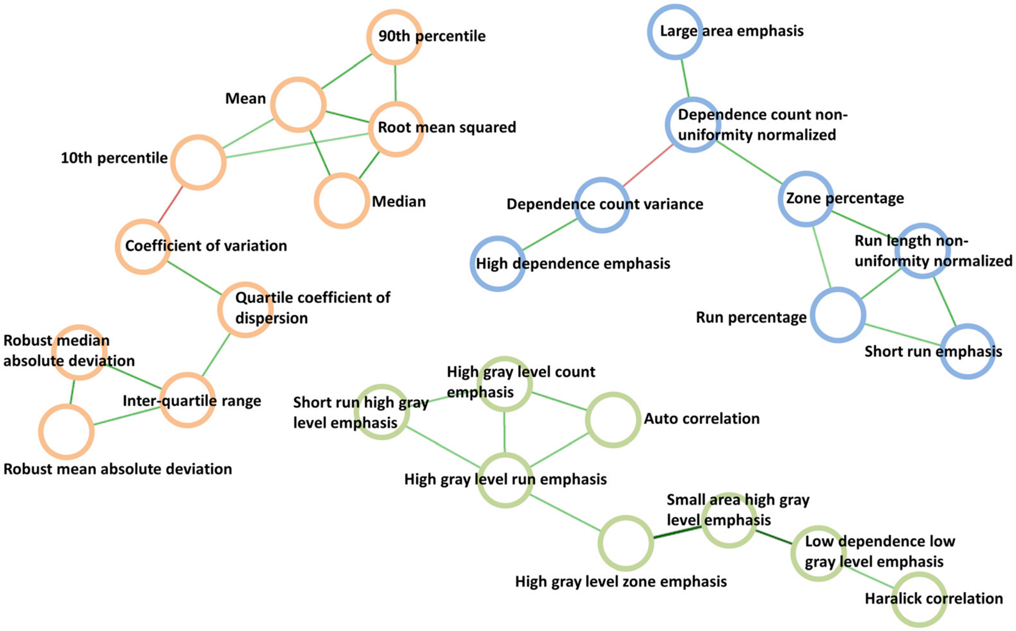
90th percentile
Mean
Root mean squared
10th percentile
Median
Coefficient of variation
Quartile coefficient of dispersion
Robust median absolute deviation
Inter-quartile range
Robust mean absolute deviation
Large area emphasis
Dependence count non-uniformity normalized
Dependence count variance
Zone percentage
High dependence emphasis
Run length non-uniformity normalized
Run percentage
Short run emphasis
High gray level count emphasis
Short run high gray level emphasis
Auto correlation
High gray level run emphasis
Small area high gray level emphasis
Low dependence low gray level emphasis
High gray level zone emphasis
Haralick correlation

Fig. 5 The three largest network components of partial correlation network that resulted from the TCIA head and neck cancer data.

features, using principal component analysis (PCA). Figure 6 shows the clustering results visualized on the first two principal components.

A significant difference in the tumor subsites was found between the two clusters with extended Fisher's exact test $p = 0.0027$. One cluster (blue) in Fig. 6 had 1 oropharyngeal, 24 oral cavity, and 10 laryngeal tumors whereas the other cluster (red) had 10 oropharyngeal, 14 oral cavity, and 18 laryngeal tumors, and this cluster was significantly enriched for HPV-positive tumors with extended Fisher's exact test $p = 0.030$. These results were similar to what we previously reported in another study.[15] It should be noted that we used a network as a basic representation of radiomic features in which subnetworks identified by the proposed method

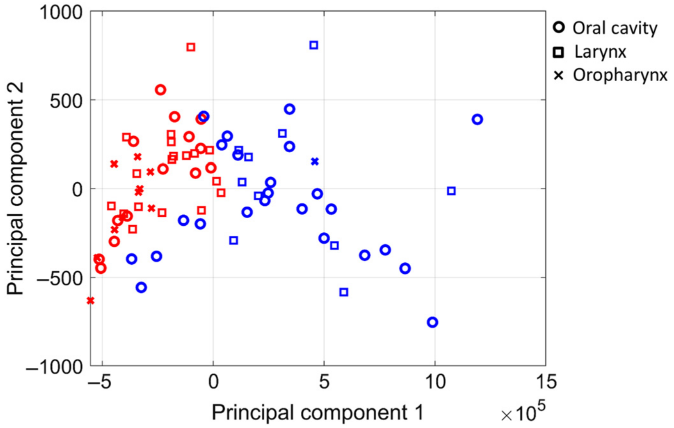

Fig. 6 The NWK algorithm was performed on the TCIA head and neck cancer data. For visualization purpose, PCA was performed and the final clustering results were projected to the first two principal components. The blue and red colors indicate the two different clusters.

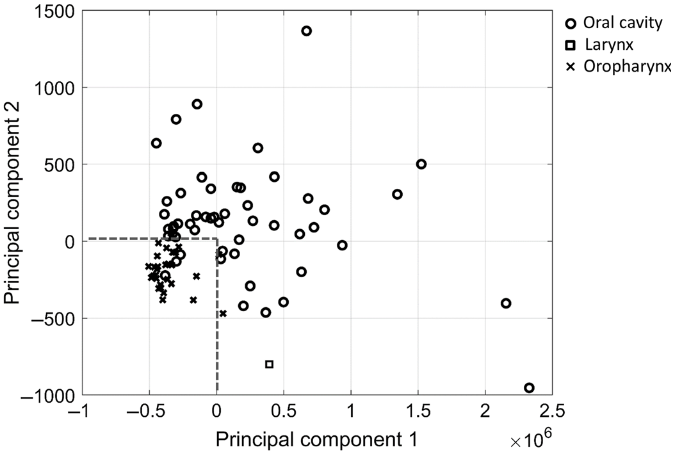

Fig. 7 For validation, the 26 radiomic features were extracted from CT scans of 83 head and neck cancer patients treated at our institution. PCA was then carried out on the data. This scatter plot shows the projection results on the first two principal components.

may indicate distinct physiological functions (which is out of scope in this study), and the clusters of samples that resulted from the NWK algorithm on the network may reflect phenotypes.

For validation, the same 26 radiomic features were extracted from CT scans of 83 HNSCC patients treated at our institution and PCA was performed on the 26 features. Figure 7 shows the projection results on the first two principal components. Similarly, oropharyngeal tumors were clustered and separated from oral cavity tumors. Using a classifier with the dotted line boundary (both principal components have 0), the accuracy of classifying oropharyngeal tumors was 87.1% (27/31) and the accuracy of classifying oral cavity tumors was 94.1% (48/51). When all 67 radiomic features were used in PCA, the use of the same boundary on the first two principal components achieved an accuracy of 67.7% (21/31) and 90.2% (46/51) in classifying oropharyngeal and oral cavity tumors, respectively.

4 Discussion

Radiomics has shown great promise as a powerful tool in quantitatively characterizing tumor phenotypes and in improving predictive power of clinical outcomes modeling, particularly in conjunction with machine learning techniques. While over the last few years a number of new radiomic features have been developed and applied to outcomes modeling, the investigation into the reproducibility of such features still remains immature despite its importance for identifying reproducible and stable radiomic features, and thereby building reliable predictive models.

The use of unstable features in predictive modeling can lead to failure in validating the models on independent data. Virginia et al.[6] reported a significant association between survival in NSCLC patients and primary mass entropy on CT scans using a training set, but this finding was not validated on a test set. Dissaux et al. investigated the capability of radiomic features extracted ^{18}F-FDG PET/CT to predict local recurrence in early-stage NSCLC patients treated with stereotactic radiotherapy. For the two multivariate models that achieved significant predictive power on a training set, a model using two PET radiomic features with information correlation 2 (IC2) from GLCM and texture strength from NGTDM remained statistically significant on a test set, whereas the other model using IC2 from PET and flatness from CT failed to reach statistical significance.[7]

As observed in our study, a partial correlation network analysis in connection with the graphical lasso showed the capability to identify reproducible radiomic features. The hypothesis underpinning this idea is that radiomic features that retain strong correlation with other features in a network formed after removing spurious or false-positive connections are likely to be reliable and reproducible. More specifically, the process to identify stable radiomic features consists of three steps, taking into account the connectivity among features in a network representation: (1) the correlation between two features is assessed, conditioning on all other available features; (2) spurious and potential false-positive connections are removed from the network using graphical lasso; and (3) weak connections are further removed. As a result, radiomic features identified through this filtering process are likely to be stable and reproducible across reconstruction methods.

Very few studies have been conducted for imaging analysis in a network form of radiomic features. Recently, a study was performed in a network analysis framework, using imaging features extracted from magnetic resonance images in intracranial ependymoma, and showed that subnetworks clearly separated tumor and healthy tissues, and even reflected tissue heterogeneity inside the tumor.[29] However, to our knowledge, the present study is the first to exploit network analysis for identifying reproducible radiomic features.

We demonstrated the potential of network-based approaches to identify reproducible radiomic features on data obtained using two different reconstruction kernels on lung cancer CT scans. In a comparison of the largest network components between the two reconstruction methods, statistically significant p-values in terms of the common number of radiomic features were obtained with 2.3×10^{-8} between Figs. 1(a) and 1(c) and 4.2×10^{-4} between Figs. 1(b) and 1(d). Interestingly, it was found that most of the radiomic features preserved the relationships (links in the network) with other features between different reconstruction settings.

We further validated the reproducibility of those radiomic features on phantom data, employing the OMT metric, W_1-Wasserstein distance (EMD), as a quantitative metric. To further investigate whether radiomic features found by the regularized partial correlation network analysis can identify phenotypes, we proposed an NWK clustering method. This approach was applied to radiomic features from the TCIA HNSCC data after the regularized partial correlation analysis, resulting in two subgroups of tumors that significantly separated tumor sites and HPV status. On an independent dataset, a silhouette test resulted in a six-cluster solution. Due to its large number compared to the two clusters in the TCIA HNSCC data, we only assessed the capability of 26 radiomic features found in the TCIA HNSCC data to separate tumor sites on a low-dimensional space after the PCA projection. Its accuracy of classification was much better than that using all radiomic features, implying the reliability of the subset of features. Due to the lack of information on HPV status, the separation of HPV-positive tumors from HPV-negative tumors was not assessed.

In this study, radiomic features were extracted from 3D volume for lung cancer and phantom data whereas for the TCIA HNSCC CT scans, radiomic features were extracted from 2D slices due to CT scans with poor quality caused by dental artifacts. A recent study showed that there was no significant difference in the predictive power of positive and negative axillary lymph node status between 2D and 3D analysis, but further research is needed.[30]

In summary, we have adopted and developed network-based methods including regularized partial correlation analysis, the Wasserstein distance on a network, and the *k*-means algorithm coupled with the Wasserstein distance, which have the potential to identify reproducible radiomic features and their relationships in a network form as well as perform clustering of samples. Further, the proposed approach has an advantage compared to traditional statistical approaches that assess individual feature distributions in case that the number of feature distributions is limited; in this study, there are only two reconstruction datasets in the lung cancer cohort.

A major limitation of this study lies in the small sample size, in particular, in the size of phantom data. We plan to produce more phantoms and analyze additional reconstruction methods for further reproducibility test of radiomic features and its results will be compared with those in large real patients' data. It should be noted that our approach is fully unsupervised learning without overfitting in the analysis. We also plan to develop a supervised method using the concept of Wasserstein distance, which enables us to build an individualized predictive model of outcomes.

5 Conclusion

In this study, we showed the potential of using a network-based approach to identify radiomic features that are reproducible over different image reconstruction methods. This was tested on CT scans in cancer patients and phantoms using the Wasserstein distance metric designed to compute the dissimilarity of samples on a network. We further proposed an NWK algorithm to cluster samples using the Wasserstein distance metric as a cost function. The clustering results on the TCIA data were validated using independent data. Applying the OMT coupled with the network analysis to radiomics could provide a powerful tool to identify reproducible radiomic features as well as develop reliable prediction models.

Disclosures

Dr. Nancy Lee is a consultant for Merck, Merck Serono. The rest of the authors have no conflicts of interest to disclose.

Acknowledgments

This research was funded in part through the National Institutes of Health/National Cancer Institute Cancer Center Support Grant P30 CA008748, R21 CA234752; Air Force Office of Scientific Research Grant FA9550-20-1-0029; National Institutes of Health Grant R01-AG048769; and a grant from the Breast Cancer Research Foundation, BCRF-17-193.

References

1. R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology* **278**, 563–577 (2016).
2. H. J. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.* **5**, 4006 (2014).
3. P. Giraud et al., "Radiomics and machine learning for radiotherapy in head and neck cancers," *Front. Oncol.* **9**, 174 (2019).
4. M. R. Folkert et al., "Predictive modeling of outcomes following definitive chemoradiotherapy for oropharyngeal cancer based on FDGPET image characteristics," *Phys. Med. Biol.* **62**(13), 5327–5343 (2017).
5. A. Traverso et al., "Repeatability and reproducibility of radiomic features: a systematic review," *Int. J. Radiat. Oncol. Biol. Phys.* **102**(4), 1143–1158 (2018).
6. B. M. Virginia et al., "Prognostic value of histogram analysis in advanced non-small cell lung cancer: a radiomic study," *Oncotarget* **9**(2), 1906–1914 (2018).

7. G. Dissaux et al., "Pretreatment 18 F-FDG PET/CT radiomics predict local recurrence in patients treated with stereotactic body radiotherapy for early-stage non-small cell lung cancer: a multicentric study," *J. Nucl. Med.* **61**(6), 814–820 (2020).
8. B. Zhao et al., "Reproducibility of radiomics for deciphering tumor phenotype with imaging," *Sci. Rep.* **6**, 23428 (2016).
9. S. Rizzo et al., "Radiomics: the facts and the challenges of image analysis," *Eur. Radiol. Exp.* **2**, 36 (2018).
10. J. E. Park et al., "Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives," *Korean J. Radiol.* **20**(7), 1124–1137 (2019).
11. A. P. Apte et al., "Technical note: extension of CERR for computational radiomics: a comprehensive MATLAB platform for reproducible radiomics research," *Med. Phys.* **45**(8), 3713–3720 (2018).
12. N. A. Obuchowski et al., "Statistical issues in the comparison of quantitative imaging biomarker algorithms using pulmonary nodule volume as an example," *Stat Methods Med. Res.* **24**(1), 107–140 (2015).
13. U. Mahmood et al., "Anatomically informed 3D printed CT phantoms: the first step of a pipeline to identify robust quantitative radiomic features," bioRxiv 773879 (2019).
14. K. Clark et al., "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository," *J. Digit. Imaging* **26**(6), 1045–1057 (2013).
15. E. Katsoulakis et al., "Radiomic analysis identifies tumor subtypes associated with distinct molecular and microenvironmental factors in head and neck squamous cell carcinoma," Oral Oncol 110, 104877 (2020).
16. R. B. Ger et al., "Practical guidelines for handling head and neck computed tomography artifacts for quantitative image analysis," *Comput. Med. Imaging Graph.* **69**, 134–139 (2018).
17. T. J. Nulton et al., "Analysis of The Cancer Genome Atlas sequencing data reveals novel properties of the human papillomavirus 16 genome in head and neck squamous cell carcinoma," *Oncotarget* **8**(11), 17684–17699 (2017).
18. S. Epskamp and E. I. Fried, "A tutorial on regularized partial correlation networks," *Psychol. Methods* **23**(4), 617–634 (2018).
19. J. H. Oh and J. O. Deasy, "Inference of radio-responsive gene regulatory networks using the graphical lasso algorithm," *BMC Bioinf.* **15**(Suppl. 7), S5 (2014).
20. J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics* **9**(3), 432–441 (2008).
21. J. Chen and Z. Chen, "Extended Bayesian information criterion for model selection with large model spaces," *Biometrika* **95**, 759–771 (2008).
22. C. Villani, *Optimal Transport: Old and New*, Springer Science & Business Media, New York (2008).
23. Y. Chen et al., "Pediatric sarcoma data forms a unique cluster measured via the Earth Mover's Distance," *Sci. Rep.* **7**, 7035 (2017).
24. J. H. Oh et al., "A novel kernel Wasserstein distance on Gaussian measures: an application of identifying dental artifacts in head and neck computed tomography," *Comput. Biol. Med.* **120**, 103731 (2020).
25. M. Pouryahya et al., "aWCluster: A novel integrative network-based clustering of multiomics for subtype analysis of cancer data," *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2019).
26. L. Kantorovich, "On the transfer of masses," *Dokl. Akad. Nauk SSSR* **37**, 227–229 (1942).
27. A. Fu, B. Narasimhan, and S. Boyd, "CVXR: an R package for disciplined convex optimization," *J. Statistical Software* **94**(14), 1–34 (2020).
28. S. Yu et al., "Optimized data fusion for kernel k-means clustering," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(5), 1031–1039 (2012).
29. M. Dominietto et al., "Role of complex networks for integrating medical images and radiomic features of intracranial ependymoma patients in response to proton radiotherapy," *Front. Med.* **6**, 333 (2020).
30. D. Arefan et al., "Machine learning prediction of axillary lymph node metastasis in breast cancer: 2D versus 3D radiomic features," *Med. Phys.* **47**(12), 6334–6342 (2020).

Jung Hun Oh is an assistant attending computer scientist at Memorial Sloan Kettering Cancer Center. He uses cutting-edge computational and statistical methods, informed by bioinformatics and machine learning techniques, to identify diagnostic biomarkers and imaging markers and to build models that predict radiation treatment outcomes.

Aditya P. Apte is a research faculty member at Memorial Sloan Kettering Cancer. He received his PhD in the Department of Mechanical Engineering at the University of Texas, Arlington, Texas, USA. His research interests include the development of radiation oncology informatics tools and systems to enable multi-institutional research and translation to clinic.

Evangelia Katsoulakis is a clinical researcher who currently works in the Veterans' Health Administration (VA) and the VA National Program Office in radiation oncology and precision medical oncology. She has an interest in developing precision radiation oncology through research in the intersection of multiomics, biomarkers, and prediction for radiation oncology and she hopes to introduce these in the clinical setting.

Nadeem Riaz is a radiation oncologist who specializes in treating head and neck cancers. He has extensive experience with advanced techniques, such as intensity-modulated radiotherapy, image-guided radiation therapy, and stereotactic body radiotherapy. His research looks at how the genetics of a tumor influences its response to traditional therapies and newer treatments, such as immunotherapies.

Vaios Hatzoglou is an associate attending radiologist at Memorial Sloan Kettering Cancer Center and has been involved in quantitative neuroimaging for more than a decade. His daily consultations have allowed him to recognize the urgent need to further the development and application of advanced imaging capabilities. He is also an integral member of several multidisciplinary Disease Management Teams that guide therapy decisions about complex cases. His research focuses on helping patients by better characterizing their tumors with advanced imaging.

Yao Yu is an assistant attending at Memorial Sloan Kettering Cancer Center specializing in cancers of the head and neck and central nervous system.

Usman Mahmood is a lead diagnostic medical physicist at Memorial Sloan Kettering Cancer. His research interests are in radiomics, machine learning, and 3D printing anatomically realistic phantoms for the evaluation of computed tomography systems.

Harini Veeraraghavan is an associate attending computer scientist at Memorial Sloan Kettering Cancer Center. Her research interests are focused on developing computer algorithms and software for computer-aided analysis of medical images to solve difficult problems in automated and semiautomated image segmentation, using multiple imaging modalities.

Maryam Pouryahya is a postdoctoral fellow at Memorial Sloan Kettering Cancer Center. Her research interests include the application of optimal mass transport theory to biological problems using multiomics data.

Aditi Iyer is a senior application developer at Memorial Sloan Kettering Cancer Center and a key contributor to the open-source software tool CERR (Computational Environment for Radiological Research). Her research interests include developing algorithms and software for medical image analysis, incorporating machine learning methods.

Amita Shukla-Dave is an attending physicist at Memorial Sloan Kettering Cancer Center. Her group's research focuses on developing and implementing advanced quantitative imaging biomarkers derived from MR imaging physics techniques, including MR fingerprinting (MRF), diffusion-weighted MRI (DW-MRI), dynamic contrast-enhanced MRI (DCE-MRI), and MR spectroscopy (MRS) for clinical application in cancer.

Allen Tannenbaum is distinguished professor of Computer Science and Applied Mathematics and Statistics at Stony Brook University. His work has won several awards including IEEE

Fellow, O. Hugo Schuck Award, and the George Taylor Research Award. He has given numerous plenary talks at major conferences including SIAM Control, IEEE CDC, AMS, and MTNS. He does research in medical imaging, computer vision, network theory, and systems and control.

Nancy Y. Lee is a head and neck radiation oncologist. She focuses on personalizing therapy for head and neck cancer with particular interest in nasopharyngeal, oropharyngeal, and thyroid cancer. She also focuses her research on using technologies to improve the therapeutic outcomes such as proton therapy or MR Linac.

Joseph O. Deasy is the chair of the Department of Medical Physics at Memorial Sloan Kettering Cancer Center. He and his group apply statistical modeling to the analysis of large, complex datasets in order to understand the relationship between treatment, patient, and disease characteristics. He is also a chief of the Service for Predictive Informatics established to work with the Departments of Radiation Oncology, Radiology, and Pathology to conduct big medical data analyses, and to advance new clinical decision support tools, e.g., image-based radiobiology, radiomics, radiogenomics, and genome-based predictive models.