International Journal of Medical Imaging
2023; 11(2): 34-41
http://www.sciencepublishinggroup.com/j/ijmi
doi: 10.11648/j.ijmi.20231102.13
ISSN: 2330-8303 (Print); ISSN: 2330-832X (Online)

The Influence of Post-Acquisition Image Processing on a Radiomic Signature Constructed from Planar Images

Tamarisk Du Plessis[1], Gopika Ramkilawon[2], Christophe Van de Wiele[3], Mike M. Sathekge[1]

[1]Department of Nuclear Medicine, University of Pretoria, Pretoria, South Africa

[2]Department of Statistics, University of Pretoria, Pretoria, South Africa

[3]Department of Diagnostic Sciences, University of Ghent, Ghent, Belgium

Email address:

tamarisk.duplessis@gmail.com (Tamarisk Du Plessis)

To cite this article:

Tamarisk Du Plessis, Gopika Ramkilawon, Christophe Van de Wiele, Mike M. Sathekge. The Influence of Post-Acquisition Image Processing on a Radiomic Signature Constructed from Planar Images. *International Journal of Medical Imaging*. Vol. 11, No. 2, 2023, pp. 34-41. doi: 10.11648/j.ijmi.20231102.13

Received: April 17, 2023; **Accepted**: May 9, 2023; **Published**: May 18, 2023

Abstract: For radiomics to be accepted as a definite tool in medicine, the outputs must be robust, repeatable and reliable. Image processing alters the quality of the input data which might have an impact on the values of the extracted features and ultimately the signatures developed. This study evaluated the magnitude of the influence of various interpolation and post-acquisition processing methods on the radiomic feature values extracted from planar images and radiomic signatures. Three different interpolation methods were applied to a chest x-ray dataset before 2-dimesional (2D) radiomic features were extracted using Pyradiomics. The influence of image size, cropping and re-segmentation were also evaluated by changing the respective variable before applying bilinear interpolation and extracting 2D features. ANOVA and post-hoc Bonferroni corrections were used to assess the differences in the radiomic feature values. Of the 93 first order- and texture- features extracted, 42 texture features (56.8%) proved to be significantly influenced ($p \leq 0.05$) by the interpolation method. Only 2 first order features (10.5%) were significantly influenced ($p \leq 0.05$) by the image size and 62 texture features (83.8%) by the other pre-processing methods evaluated. Pearson's Correlation Analysis was then applied to develop a separate radiomics signature from each of the six image processing datasets under consideration. Five identical signatures were developed, with only the uncropped dataset that resulted in a unique signature. This study showed that the interpolation algorithms and other processing applied to planar images do have a noticeable influence on most radiomic feature values extracted. But regardless of the differences seen in the feature values, the radiomic signatures were reproducible for most datasets using different image processing methods.

Keywords: Radiomics, Image Processing, Interpolation, Chest X-rays, Radiomic Signatures

1. Introduction

Radiomics is a field of study that has the potential to extract large amounts of quantitative features from medical images [1]. It makes use of statistically based imaging analysis algorithms to identify image features which can be used to quantify disease characteristics [2]. Radiomic feature extraction has the hypothesis that the correct combination of these algorithms, together with the clinical data, can express meaningful tissue properties useful in the management of a disease [3]. Numerous articles have been published on the use of radiomics in tumours [4-7], and a few on the application in non-neoplastic diseases such as pulmonary tuberculosis and Covid-19 [8-10]. In recent years this data mining tool advanced to the point where it incorporates machine learning and deep learning approaches to build state of the art radiomic signatures and models [11]. However, radiomic studies can only be used as a definite tool in medicine once the outcomes are reliable, repeatable, robust and validated.

The complex multi-step process of radiomics includes; Image acquisition, image post-acquisition processing, image segmentation, feature extraction, dimensionality reduction, association analysis, model construction and database

development [2].

The foundation of this multi-step process however remains the input image, either two- (2D) or three dimensional (3D). But with the fast development of quantitative imaging methods the focus on this foundational step has been lost. It was shown that the quality of the input data has a considerable impact on the value of the extracted features [3]. And that variables such as a variety of acquisition, reconstruction and post processing parameters influence the image texture and noise and consequently the value of the extracted features [3]. The robustness of radiomic features therefor depend significantly on the image post-processing applied [12]. The Imaging Biomarker Standardization Initiative (IBSI) was published in 2019 with the aim to standardize image biomarker nomenclature and definitions, to suggest tools for verifying radiomics software implementations and to standardize reporting guidelines [13]. The IBSI gives useful suggestions on the radiomics workflow with detailed technical instructions regarding the image post-acquisition processing workflow required [13]. It recommends that the following steps should be followed before feature extraction algorithms are applied; dataset classification, data conversion, post-acquisition processing, segmentation, interpolation and re-segmentation [13].

Data classification refers to sorting the images in the dataset to only include images of the same modality, patient orientation and photometric interpretation. Data conversion of raw image data into more meaningful data is only required for certain image modalities without definite grayscale values, e.g. PET [11]. To homogenize datasets further various post-acquisition processing steps are required; image format conversion, normalization and discretization of the grayscale intensities and image interpolation [13].

Interpolation algorithms translate image intensities from the original image grid to an interpolation grid. Pixels are spatially represented by their centres in such grids [11, 13]. Isotropic pixel spacing is required for texture features to be rotationally invariant, and to allow comparison between different datasets [13]. There are currently no clear recommendations on whether up-sampling or down-sampling should be the preferred interpolation method for radiomic studies [12]. Only a recommendation that a calculated decision should be made regarding this, as up-sampling to the smallest pixel dimension can introduce artificial information, while down-sampling to the largest pixel dimension can result in information loss [13].

Various interpolation algorithms are commonly used for volumetric image pre-processing in medicine, e.g. nearest neighbour, trilinear, tricubic convolution and tricubic spline interpolation [13]. The 2D equivalents of these popular interpolation methods are 2D-nearest neighbour-, bilinear- and bicubic interpolation, but no literature could be found on the optimal choice for interpolation of planar medical images. Nearest neighbour interpolation adapts the intensity of the nearest neighbouring pixel without regard to the intensities of other neighbouring pixels [14]. This is the simplest interpolation method, but may result in blocky images. Bilinear interpolation takes the intensity of 4 neighbouring pixels into account and applies two linear interpolations to obtain a new pixel intensity [14]. Bicubic interpolation results in the smoothest interpolation as it uses 16 pixels and applies a third order polynomial function to interpolate the new pixel intensity [14].

In the recommended workflow discussed above only 3D images were mentioned, which leaves the question as to whether the same labour intensive steps are required for planar images. Also no literature could be found that quantitatively shows the influence that each of these post-acquisition processing steps have on the feature values extracted or on the radiomic signatures being developed.

This study was performed with the aim to evaluate the scale of influence that different interpolation methods and other common post-acquisition processing applied to planar images will have on the extracted radiomic features. It also evaluated which radiomic features are most sensitive to these post-acquisition processing methods and how these feature values influence radiomics signatures. A chest x-ray (CXR) dataset was selected for this study since these planar images are still clinically used in many countries for diagnosis and disease management [15].

2. Methods

2.1. Dataset

This is a retrospective study consisting of 103 posteroanterior (PA) CXR of patients diagnosed with active pulmonary tuberculosis. Ethical clearance was granted by the researchers' tertiary institution to use these images for this study.

2.2. Image Pre-Processing

The original dataset consisted of DICOM images of various sizes and width-to-height ratios. To remove as many variables as possible that might have an underlying influence on the interpolation, all images were cropped to a square (equal width-to-heigh ratio). Only one dimension of the images was cropped to preserve the maximum dimensions of the CXR. Unfortunately automatic batch cropping was unsuccessful as most CXR were acquired at asymmetrical patient positions and automatic cropping of these images removed lung pixels. An expensive added layer of manual cropping of all images was therefore performed using commercial image processing software, Total Image Converter (by CoolUtils.com file converters) version 8.2.0.237. To eliminate the dual processing required in DICOM images, where the corresponding header group element must be changed with the image data, images were converted to another format. The same software was used to convert all images to PNG format and to correct the unconventional photometric interpretation of Monochrome 1. During the conversion from DICOM to PNG-format, the bit depth was changed from 14 or 12 bits to 8 bits to obtain a uniform dataset and the scalar DICOM type was converted to

conventional RGB type for PNG formatted images. Image size, pixel size and image resolution were preserved. Since the radiomics library used can only apply the feature algorithms to scalar images, the images were finally converted to L-mode with the ITU-R 601-2 luma transform.

2.3. Image Interpolation

All images in the original DICOM dataset had dimensions equal to or larger than 1024x1024 pixels after being cropped. These large images ensured that no image had to be extrapolated. All cropped images were therefore down sampled to 1024x1024 pixels using 3 basic interpolation algorithms commonly used in image processing; bilinear interpolation, bicubic interpolation and 2D nearest neighbour interpolation. This was done using the cv2.resize() function (cv2.INTER_LINEAR, cv2.INTER_CUBIC and cv2.INTER_NEAREST) in Python version 3.7.6.

2.4. Segmentation and Feature Extraction

A fully automatic in-house U-net based segmentation model was used to segment the lung region-of-interest (ROI) [16]. The segmentation model resizes images to 256x256 pixels (using bilinear interpolation) before segmenting the lungs as a 256x256 pixel mask. All masks therefore had to be extrapolated to 1024x1024 pixels to have similar dimensions to its corresponding image. Extrapolation of the masks were done using nearest neighbour interpolation. The extrapolation method will have no influence on the segmentation, or the results of this study, as a mask is simply a binary matrix. Each image therefore had 3 versions (one version for each interpolation method) associated to a single mask.

The Pyradiomics library has 103 2D features available for extraction (10 2D shape features, 19 first order- and 74 texture features). Pyradiomics (version 3.0) in Python (version 3.7.6) was used to extract a total of 93 2D features (first order- and texture features only) from each image version. No shape-based features were evaluated as these are calculated using the shape of the ROI defined by the mask. Since the same mask is used for all 3 image versions this will naturally result in identical features. Pyradiomics have five different texture feature groups namely; GLCM = Gray Level Co-occurrence Matrix, GLDM = Gray Level Dependence Matrix, GLRLM = Gray Level Run Length Matrix, GLSZM = Gray Level Size Zone Matrix and NGTDM = Neighbouring Gray Tone Difference Matrix [1]. All variables in the Pyradiomics library were left at default and no filters were applied.

2.5. Evaluating the Influence of Image Size, Cropping and Re-segmentation

The influence of 3 other common post-acquisition image processing techniques was also evaluated. To do this the bilinearly interpolated dataset that was discussed in the above methodology (section 2.3), was used as baseline. Mask segmentation and feature extraction was then repeated three times while only 1 post-acquisition processing variable was changed at a time to obtain a second, third and fourth feature set for comparison.

Firstly, the influence of image dimension was studied by interpolating the baseline dataset to 256x256 pixels instead of 1024x1024 pixels with bilinear interpolation. This dataset was again segmented and radiomic features extracted to obtain the second set of features called *Size*. Secondly, the influence of image cropping before applying the segmentation model and radiomic algorithms was evaluated. This dataset was left uncropped before all other image pre-processing steps were performed. The masks were segmented and radiomic features extracted to obtain the third set of features called *Uncropped*. Lastly, the influence of re-segmenting the images after image interpolation was evaluated. The fourth set of features called *Re-segmented* was therefore obtained by using masks that were re-segmented after image interpolation.

2.6. Statistical Analysis

Statistical analysis for both image processing sections of this study was performed using SPSS 28.0. ANOVA and post-hoc Bonferroni corrections were used to assess the differences between features obtained for the different variables using the different image post-processing methodologies under study.

2.7. Dimensionality Reduction

A Shapiro-Wilk test was used to ensure no normality assumption violations in the features. This dataset does not fully adhere to the assumption of normality, but since this sample size is sufficiently large (larger than 100), Pearson Correlations can still be applied as it is robust against such violations [17]. Pearson Correlation Analysis was therefore applied to obtain a radiomics signature for each of the six above-mentioned datasets. Feature pairs with absolute correlations greater than 0.8 were removed.

3. Results

3.1. Section 1: Influence of the Interpolation Algorithm

When comparing bilinear, bicubic and nearest neighbour interpolation, significant results (F-test) between these three methodologies were found for 42 of the 93 features extracted. The features with significant p-values ($p \leq 0.05$) are summarized in table 1. To evaluate which interpolation method contributed to the significantly lower or higher feature values, individual group comparisons were performed by comparing the mean values of each feature to the feature group mean. The results are also summarized in table 1 as either < or > to indicate a significantly lower mean (<) or a significantly higher mean (>) respectively.

Table 1. *Summary of features that displayed significant differences ($p \leq 0.05$) when comparing the different interpolation methods with their respective individual group comparison results.*

Feature no	Feature name	Significant (F-test)	Bilinear	Bicubic	Nearest Neighbour
23	glcm_Contrast	0.0001	<	>	>
24	glcm_Correlation	0.0001	>		
25	glcm_DifferenceAverage	0.0001	<		>
26	glcm_DifferenceEntropy	0.0001	<		>
27	glcm_DifferenceVariance	0.0001	<	>	>
28	glcm_Id	0.0001	>		<
29	glcm_Idm	0.0001	>		<
30	glcm_Idmn	0.0001	>		
31	glcm_Idn	0.0001	>		
32	glcm_Imc1	0.0001	<		>
33	glcm_Imc2	0.0001	>		
34	glcm_InverseVariance	0.0001	<	>	>
37	glcm_JointEntropy	0.0320	<		>
38	glcm_MCC	0.0001	>		
43	gldm_DependenceEntropy	0.0290	<		>
44	gldm_DependenceNonUniformity	0.0050	>	<	<
45	gldm_DependenceNonUniformityNormalized	0.0080	>		<
46	gldm_DependenceVariance	0.0001	<	>	>
50	gldm_LargeDependenceEmphasis	0.0001	>		<
51	gldm_LargeDependenceHighGrayLevelEmphasis	0.0130	>	<	<
54	gldm_SmallDependenceEmphasis	0.0001	<		>
55	gldm_SmallDependenceHighGrayLevelEmphasis	0.0001	<		>
56	gldm_SmallDependenceLowGrayLevelEmphasis	0.0010	<		
57	glrlm_GrayLevelNonUniformity	0.0001	<	>	>
65	glrlm_RunEntropy	0.0001	>		
66	glrlm_RunLengthNonUniformity	0.0001	<		>
67	glrlm_RunLengthNonUniformityNormalized	0.0001	<		
68	glrlm_RunPercentage	0.0001	<	>	>
70	glrlm_ShortRunEmphasis	0.0001	<	>	>
71	glrlm_ShortRunHighGrayLevelEmphasis	0.0020	<	>	>
73	glszm_GrayLevelNonUniformity	0.0001	<		>
78	glszm_LargeAreaHighGrayLevelEmphasis	0.0290	>		<
81	glszm_SizeZoneNonUniformity	0.0001	<	>	>
82	glszm_SizeZoneNonUniformityNormalized	0.0001	<	>	
83	glszm_SmallAreaEmphasis	0.0001	<	>	
84	glszm_SmallAreaHighGrayLevelEmphasis	0.0500	<		
86	glszm_ZoneEntropy	0.0001	>		
87	glszm_ZonePercentage	0.0001	<		>
89	ngtdm_Busyness	0.0001	<		>
91	ngtdm_Complexity	0.0001	<	>	>
92	ngtdm_Contrast	0.0001	<	>	>
93	ngtdm_Strength	0.0001	>		<

The significant differences can be appreciated by observing figure 1 where the normalized means of the 93 features extracted using the different interpolation methods were plotted. Before plotting the graphs, the means were normalized with min-max normalization to compensate for the scale variety of the radiomic features.

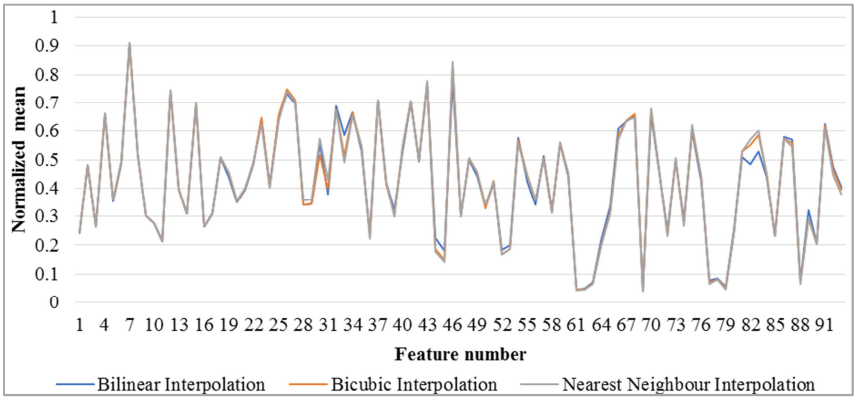

Figure 1. *Plot of the normalized means of the 93 features extracted using three different interpolation methods (Bilinear, Bicubic and Nearest Neighbour interpolation) to indicate the significant differences obtained in some of the radiomic features.*

3.2. Section 2: Influence of Image Size, Image Cropping and Re-segmentation

When comparing the baseline, Size, Uncropped and Re-segmented datasets, significant differences (F-test) were found for 63 of the 93 features extracted. The features with significant p-values ($p \leq 0.05$) are summarized in table 2. Individual group comparisons were performed to evaluate which pre-processing methods contributed to the significantly lower or higher feature values by comparing the means of each feature to the feature group mean. The results are also indicated in table 2 as either < or > to indicate a significantly lower mean (<) or a significantly higher mean (>) respectively.

Table 2. *Summary of features that displayed significant differences ($p \leq 0.05$) when comparing the various pre-processing methods with their respective individual group comparison results.*

Feature no	Feature name	Significant (F-test)	Baseline	Size	Uncropped	Re-segmented
3	firstorder_Energy	0.0001		<		
16	firstorder_TotalEnergy	0.0001		<		
19	glcm_Autocorrelation	0.0250		<		
23	glcm_Contrast	0.0001		>		
24	glcm_Correlation	0.0001		>		
25	glcm_DifferenceAverage	0.0001		>		
26	glcm_DifferenceEntropy	0.0001		>		
27	glcm_DifferenceVariance	0.0001		>		
28	glcm_Id	0.0001		<		
29	glcm_Idm	0.0001		<		
30	glcm_Idmn	0.0001		<		
31	glcm_Idn	0.0001		<		
32	glcm_Imc1	0.0001		>		
33	glcm_Imc2	0.0001		<		
34	glcm_InverseVariance	0.0001		>		
35	glcm_JointAverage	0.0390		<		
36	glcm_JointEnergy	0.0001		<		
37	glcm_JointEntropy	0.0001		>		
38	glcm_MCC	0.0001		<		
39	glcm_MaximumProbability	0.0001		<		
40	glcm_SumAverage	0.0390		<		
41	glcm_SumEntropy	0.0050		>		
43	gldm_DependenceEntropy	0.0001		>		
44	gldm_DependenceNonUniformity	0.0001		<	<	
45	gldm_DependenceNonUniformityNormalized	0.0001		<		
46	gldm_DependenceVariance	0.0001		>		
47	gldm_GrayLevelNonUniformity	0.0001		<		
50	gldm_LargeDependenceEmphasis	0.0001		<		
51	gldm_LargeDependenceHighGrayLevelEmphasis	0.0001		<		
54	gldm_SmallDependenceEmphasis	0.0001		>		
55	gldm_SmallDependenceHighGrayLevelEmphasis	0.0001		>		
56	gldm_SmallDependenceLowGrayLevelEmphasis	0.0001		>		
57	glrlm_GrayLevelNonUniformity	0.0001		<		
60	glrlm_HighGrayLevelRunEmphasis	0.0310		<		
61	glrlm_LongRunEmphasis	0.0001		<	<	
62	glrlm_LongRunHighGrayLevelEmphasis	0.0001		<		
63	glrlm_LongRunLowGrayLevelEmphasis	0.0030		<		
65	glrlm_RunEntropy	0.0001		<		
66	glrlm_RunLengthNonUniformity	0.0001		<		
67	glrlm_RunLengthNonUniformityNormalized	0.0001		>		
68	glrlm_RunPercentage	0.0001		<		
69	glrlm_RunVariance	0.0010		<		
70	glrlm_ShortRunEmphasis	0.0001		<		
71	glrlm_ShortRunHighGrayLevelEmphasis	0.0001		>		
72	glrlm_ShortRunLowGrayLevelEmphasis	0.0001		>		
73	glszm_GrayLevelNonUniformity	0.0001		<		
74	glszm_GrayLevelNonUniformityNormalized	0.0010		<		
75	glszm_GrayLevelVariance	0.0001		>		
77	glszm_LargeAreaEmphasis	0.0001		<	<	
78	glszm_LargeAreaHighGrayLevelEmphasis	0.0001		>	<	
79	glszm_LargeAreaLowGrayLevelEmphasis	0.0001		<	<	
80	glszm_LowGrayLevelZoneEmphasis	0.0410		>		
81	glszm_SizeZoneNonUniformity	0.0001		<		

Feature no	Feature name	Significant (F-test)	Baseline	Size	Uncropped	Re-segmented
82	glszm_SizeZoneNonUniformityNormalized	0.0001		<		
83	glszm_SmallAreaEmphasis	0.0001		<		
86	glszm_ZoneEntropy	0.0170		>		
87	glszm_ZonePercentage	0.0001		>		
88	glszm_ZoneVariance	0.0001		<	<	
89	ngtdm_Busyness	0.0001		<		
90	ngtdm_Coarseness	0.0001		>		
91	ngtdm_Complexity	0.0001		>		
92	ngtdm_Contrast	0.0001		>		
93	ngtdm_Strength	0.0001		>		

The individual group comparisons are graphically presented in figures 2, 3 and 4. The normalized mean values of the baseline feature set were plotted against the *Size* features (Figure 2), *Uncropped* features (Figure 3) and the *Re-segmented* features (Figure 4) respectively.

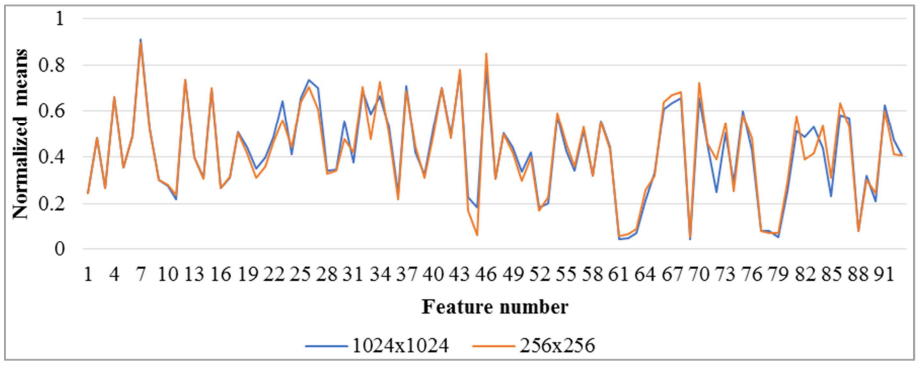

Figure 2. *Plot of the normalized means of the 93 features extracted using different image sizes (1024x1024 pixels and 256x256 pixels) to indicate the significant differences obtained in some of the radiomic features.*

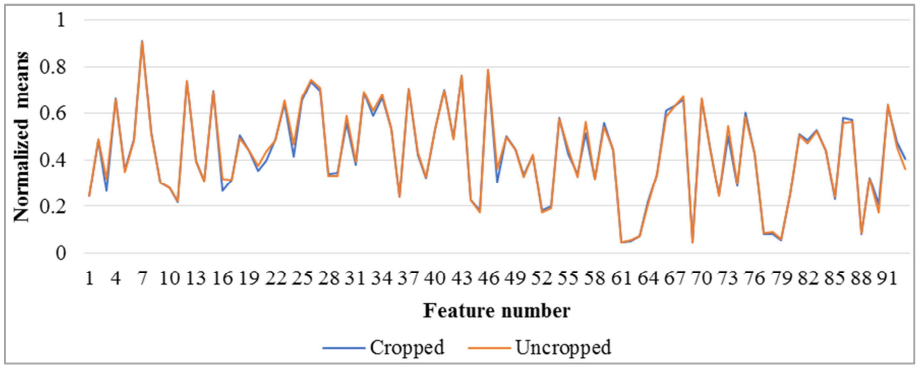

Figure 3. *Plot of the normalized means of the 93 features extracted using cropped and uncropped images to indicate the significant differences obtained in some of the radiomic features.*

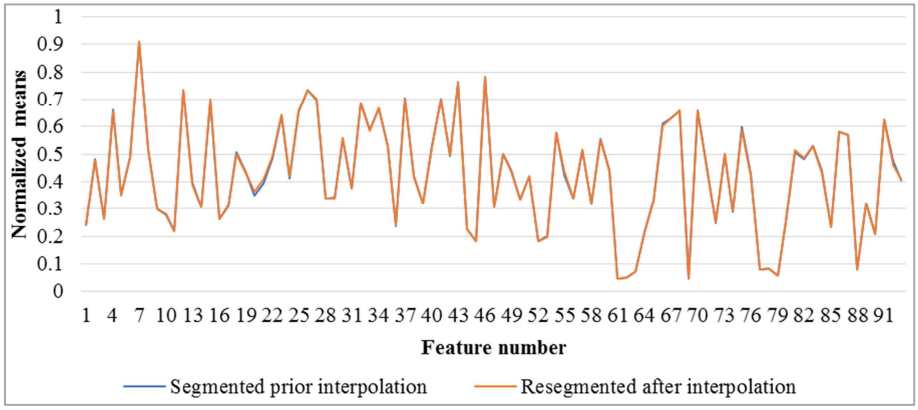

Figure 4. *Plot of the normalized means of the 93 features extracted using images segmented prior to interpolation and re-segmented images to indicate that no significant differences are seen in the extracted radiomic features.*

3.3. Section 3: Influence on Radiomic Signatures

Six datasets were under consideration in this study, three evaluating the influence of the interpolation algorithm (*Bilinear*, *Bicubic* and *Nearest Neighbour*) and three evaluating the influence of other image post processing methods (*Size*, *Uncropped* and *Re-segmented*). Table 3 summarizes the signatures obtained for each of these datasets.

Table 3. *Summary of the signature features retained for each of the six datasets under consideration.*

DATASET	# features removed	Signature features retained
BI-LINEAR	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
BI-CUBIC	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
NEAREST NEIGHBOUR	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
SIZE	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength
UNCROPPED	89	glcm_Idmn, gldm_GrayLevelNonUniformity, ngtdm_Busyness, ngtdm_Strength
RE-SEGMENTED	90	glcm_Idmn, ngtdm_Busyness, ngtdm_Strength

4. Discussions

In radiomics first order statistics uses basic statistical algorithms to describe the value and distribution of a group of pixels without concern for spatial relationships [3]. Second order statistics describe the textural features and are calculated by the statistical inter-relationship between the pixels in the ROI [1]. This was confirmed by the results of this study where 42 (56.8%) and 62 (83.8%) out of the 74 texture features extracted were significantly influenced by the interpolation method and other post-acquisition image processing respectively. The only 2 first order feature values influenced by any image processing were Energy (p=0.001) and Total Energy (p=0.001).

4.1. Section 1: Interpolation

Since no ground truth exists regarding the correct interpolation algorithm to apply in radiomic studies, the group mean for each feature was considered as the baseline for individual comparisons. When individual group comparisons were performed, no significant differences were seen in any first order feature values. However significant differences in 56.8% of the texture features amongst all three interpolation methods were observed. The differences were random without an obvious trend or pattern that can be identified. The mean feature values for bicubic interpolation, which is the smoothest interpolation, do however have outputs closest to the group mean with only 15 features that differ considerably from the group mean. Another observation is that bilinear- and nearest neighbour interpolation were always distributed to opposite sides of the group mean in the 42 features that differed significantly. Bicubic interpolation will most likely result in the most stable feature choice. However when selecting an interpolation algorithm all factors, such as the quality of the images, available computational power and study outcomes, should be considered. By definition bilinear interpolation is considered as the conservative choice in image interpolation as it takes the intensity of 4 neighbouring pixels into account, compared to 16 in bicubic interpolation and only 1 in nearest neighbour interpolation [13]. Without a ground truth, the above results highlight the importance of consistency in radiomics studies. Regardless of the method selected only a single interpolation algorithm must be applied across an entire study and in comparative studies.

4.2. Section 2: Size, Cropping and Re-segmentation

In this part of the study all four feature datasets (*Baseline, Size, Uncropped and Re-segmented*) were interpolated with bilinear interpolation. When the *Size* CXR dataset was down sampled to a quarter (256x256 pixels) of the baseline images' dimensions, 83.8% of the texture feature values extracted were significantly influenced. This can be accounted for by the statistical inter-relationship of the pixels in the secondary matrixes that would have changed by down sampling the images. Some 3D radiomic studies also mentioned the influence of image size on the outcome of their studies [18, 19].

Only 6 out of 74 (8.1%) texture features extracted were significantly influenced by not cropping the images into a square before applying the interpolation algorithm. This can also be accounted for by the image matrix that would have changed when the interpolation algorithm had unevenly down sampled the pixels of the uncropped images in the two dimensions. It is possible that the influence of uncropped images will increase as the width-to-height ratios of the uncropped images increase.

By re-segmenting the masks after image interpolation were applied, no significant influences were observed in any features. This might hold true for this study only as the segmentation model used also applied bilinear interpolation to the images before segmenting the masks. In essence the masks were always re-segmented without doing it intentionally. Even being study specific, these results are valuable as it proves good repeatability of the interpolation algorithm, the segmentation model and the radiomics library used.

4.3. Section 3: Radiomic Signatures

Five out of the six radiomic signatures that were constructed from each dataset were identical consisting of 3 texture features; glcm_Idmn, ngtdm_Busyness and ngtdm_Strength. Only the *Uncropped* dataset signature had an additional feature, gldm_GrayLevelNonUniformity. This reproducible result is

seen regardless of the significant differences found in the extracted feature values caused by the different interpolation methods, sizes and re-segmentations applied. Also the unique signature obtained when the images were left uncropped indicates that the labour intensive process of cropping images to squares (or equal width-to-height ratios for the entire dataset) is imperative.

5. Conclusions

This study showed that first order feature values are not significantly impacted by the interpolation algorithms and other image processing methods applied, but that it does significantly influence most texture feature values extracted from planar images. It also showed that regardless of the significant differences seen in the extracted feature values, caused by most post-acquisition image processing methods, the outcome of the radiomics signatures remains reproducible. The only image post-acquisition processing step that resulted in a different signature was image cropping and it must therefore strongly be considered in all planar image studies.

Larger sample sizes are required to verify this study, but it preliminarily shows that image post-processing, except cropping, does not considerably influence the outcome of radiomic signatures. Focus should rather be placed on applying the correct dimensionality algorithms that are robust against any instabilities caused by image post-processing. It is however imperative to be consistent with all image processing steps applied across an entire radiomics study. Thorough reporting of all image processing applied in radiomic studies is also crucial to increase the reproducibility and validity of this field of study.

Conflict of Interest

The authors declare that no funding, grants, or other financial support were received for the preparation of this manuscript. The authors also have no competing interests to declare that are relevant to the content of this article.

References

[1] van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer research* 2017; 77: e104-e107.

[2] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging* 2012; 30: 1234-1248.

[3] Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *European Radiology Experimental* 2018; 2: 1-8.

[4] Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *AJNR. American journal of neuroradiology* 2018; 39: 208-216.

[5] Feng B, Chen X, Chen Y, Liu K, Li K, Liu X, et al. Radiomics nomogram for preoperative differentiation of lung tuberculoma from adenocarcinoma in solitary pulmonary solid nodule. *European Journal of Radiology* 2020; 128.

[6] Feng B, Chen X, Chen Y, Lu S, Liu K, Li K, et al. Solitary solid pulmonary nodules: a CT-based deep learning nomogram helps differentiate tuberculosis granulomas from lung adenocarcinomas. *European Radiology* 2020.

[7] Hu Y, Zhao X, Zhang J, Han J, Dai M. Value of 18F-FDG PET/CT radiomic features to distinguish solitary lung adenocarcinoma from tuberculosis. *European Journal of Nuclear Medicine and Molecular Imaging* 2020.

[8] Bei W, Min L, He M, Fangfang H, Yan W, Shunying Z, et al. Computed tomography-based predictive nomogram for differentiating primary progressive pulmonary tuberculosis from community-acquired pneumonia in children. In: *BMC Medical Imaging*, 2019; 1-11.

[9] Giraudo C, Frattin G, Fichera G, Motta R, Stramare R. A practical integrated radiomics model predicting intensive care hospitalization in COVID-19. *Critical care (London, England)* 2021; 25: 145.

[10] Han Y, Chen C, Tang L, Lin M, Jaiswal A, Ding Y, et al. Using Radiomics as Prior Knowledge for Abnormality Classification and Localization in Chest X-rays. In, 2020.

[11] Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. *Medical physics* 2020; 47: e185-e202.

[12] van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights into Imaging* 2020; 11: 91.

[13] Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 2020; 295: 328-338.

[14] Acharya T, Tsai P-S. Computational foundations of image interpolation algorithms. *Ubiquity* 2007; 2007: 1-17.

[15] Chassagnon G, Vakalopoulou M, Paragios N, Revel MP. Artificial intelligence applications for thoracic imaging. *European Journal of Radiology* 2020; 123.

[16] Dixon SAP. Using Deep Learning to Segment Chest X-Rays for the Analysis of Pneumonia. In: *School of Biomedical Engineering*: University of Sydney, 2019.

[17] Havlicek LL, Peterson NL. Robustness of the Pearson Correlation against Violations of Assumptions. *Perceptual and Motor Skills* 1976; 43: 1319-1334.

[18] Altazi BA, Zhang GG, Fernandez DC, Montejo ME, Hunt D, Werner J, et al. Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms. *Journal of Applied Clinical Medical Physics* 2017; 18: 32-48.

[19] Bailly C, Bodet-Milin C, Couespel S, Necib H, Kraeber-Bodéré F, Ansquer C, et al. Revisiting the Robustness of PET-Based Textural Features in the Context of Multi-Centric Trials. In: *PLOS ONE*, 2016.