

Does image normalization and intensity resolution impact texture classification?

Marcin Kociołek ^{a,*}, Michał Strzelecki ^a, Rafał Obuchowicz ^b

^a Institute of Electronics, Lodz University of Technology, ul. Wólczańska 211/215, 90-924 Lodz, Poland

^b Department of Diagnostic Imaging, Jagiellonian University Medical College, ul Kopernika 19, 31-501, Cracow Poland



ARTICLE INFO

Article history:

Received 29 May 2019

Received in revised form 4 March 2020

Accepted 5 March 2020

Keywords:

Image procession
Texture features
Normalization

ABSTRACT

Image texture is a very important component in many types of images, including medical images. Medical images are often corrupted by noise and affected by artifacts. Some of the texture-based features that should describe the structure of the tissue under examination may also reflect, for example, the uneven sensitivity of the scanner within the tissue region. This in turn may lead to an inappropriate description of the tissue or incorrect classification. To limit these phenomena, the analyzed regions of interest are normalized. In texture analysis methods, image intensity normalization is usually followed by a reduction in the number of levels coding the intensity. The aim of this work was to analyze the impact of different image normalization methods and the number of intensity levels on texture classification, taking into account noise and artifacts related to uneven background brightness distribution. Analyses were performed on four sets of images: modified Brodatz textures, kidney images obtained by means of dynamic contrast-enhanced magnetic resonance imaging, shoulder images acquired as T2-weighted magnetic resonance images and CT heart and thorax images. The results will be of use for choosing a particular method of image normalization, based on the types of noise and distortion present in the images.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Image texture is one of the most important components of many types of images, including medical images. Visualization by means of various medical imaging modalities represents the properties of internal organs and tissues in terms of texture. The texture of the cross-sections of such structures, observed for example by tomographic imaging, provides information that can be further supported by a diagnostic process. Texture parameters reflect the physiological properties of tissues. This enables, *inter alia*, the segmentation of organs, the detection of lesions and an assessment of the degree of pathological changes. The significance of texture analysis for aided imaging diagnostics has been demonstrated for all kinds of imaging modalities, including computer tomography (CT) (Thevenot et al., 2013), Magnetic Resonance Imaging (Larroza et al., 2017), ultrasound (Chrzanowski et al., 2008) and optical imag-

ing (Kropidłowski et al., 2014). The efficiency of texture analysis is strongly affected, however, by image quality.

Medical images are often corrupted by noise and affected by artifacts. These phenomena are specific for a given imaging modality. In Magnetic Resonance Imaging (MRI) (Aja-Fernández et al., 2009; Aja-Fernández and Vegas-Sánchez-Ferrero, 2016), noise is produced by the stochastic motion of free electrons in the radio frequency receiver coils and by eddy current losses in the patient. Optical imaging is affected by optical detector noise, which usually has Poisson or Gaussian distributions (Liu et al., 2017). In CT, Poisson noise is caused by the statistical error of the low photon counts in the detectors (Boas and Fleischmann, 2012). In ultrasound images, the superposition of acoustical echoes with random phases and amplitudes produces speckle noise (Michailovich and Tannenbaum, 2006).

Numerous papers have described noise reduction techniques applied to medical images. Median filters can be used for removal of impulse noise (Chervyakov et al., 2017). Anisotropic filtering can be utilized for CT images (Yilmaz et al., 2017). Soft thresholding of wavelet transform coefficients can be applied to remove speckle noise from ultrasound images (Gupta et al., 2004). Partial differential equations have been applied to denoise MRI images (Lysaker et al., 2003). Other noise removal techniques have been

* Corresponding author.

E-mail addresses: marcin.kociolek@p.lodz.pl (M. Kociołek),
michal.strzelecki@p.lodz.pl (M. Strzelecki), [\(R. Obuchowicz\).](mailto:r.obuchowicz@gmail.com)

documented in review papers (Ouahabi, 2013; Priya et al., 2017; Somkuwar and Bhargava, 2013). There are two general issues with noise removal techniques. Such techniques can be tedious (Pieciak et al., 2017), and their application often causes changes to the image content, especially at the texture level (Aja-Fernández et al., 2009). Thus, the susceptibility of texture analysis methods to noise should be considered.

The results of texture analysis depend on the image acquisition parameters. These include image resolution, the size of the field of view (FoV), the matrix size and the background brightness heterogeneity. In addition, analyses performed on images of different patients often use different values for the acquisition parameters. This may cause the brightness and contrast of individual regions of interest (ROIs) to differ. The values can vary for ROIs within an image (Styner et al., 2005) and for ROIs in subsequent images. As a result, the values of some textural features depend not only on the texture, but also on the average brightness and/or contrast of the ROI (Materka et al., 2000). For this reason, some of the features that should describe the structure of the tissue under examination may also reflect, for example, the scanner's uneven sensitivity within the analyzed tissue region (Materka and Strzelecki, 2013). This may in turn lead to an inappropriate description of the tissue or incorrect classification. To limit these phenomena, the ROIs are normalized (C. P. Loizou et al., 2009).

The aim of this work was to analyze the impact of implementing different image normalization methods on the results of texture classification, with attention to noise and artifacts related to uneven background brightness distribution. The results may be useful for choosing a particular method of image normalization based on the types of noise and interference present in the images. In texture analysis methods, image intensity normalization is usually followed by the reduction of levels coding the intensity. For this reason, the influence of the number of intensity levels on the accuracy of texture classification was also examined. This study constitutes an extension of research described in (Strzelecki et al., 2018) and (Kociolek et al., 2018). The analyses presented here were performed for textures from the Brodatz album (Brodatz, 1966), for kidney area images obtained by means of dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) (Eikelfjord et al., 2017), shoulder T2 weighted MRI images (Obuchowicz et al., 2019) and thorax CT images. Brodatz textures were used for the simulation of various types of noise and background nonuniformity artefacts. The paper is structured as follows. Section 2 presents the concept of normalization and intensity reduction for texture analysis. Section 3 describes the data utilized in our study as well as the experimental procedure. Experimental results are discussed in Section 4. The results and a summary of the experiments are presented in Section 5.

2. Normalization and intensity level reduction for texture analysis

A typical procedure for obtaining texture-based features is shown in Fig. 1

Normalization usually constitutes the first step of the procedure. It leads to an extension of the image histogram in the ROI into the entire available intensity range. This has two effects: there is more contrast between the bright and dark structural elements creating the investigated texture, and the influence of the local mean intensity is reduced. Both of these effects usually improve the quality of the features obtained. The next step in the feature estimation procedure is the reduction of the number of intensity levels used for texture representation. Usually, the intensity ranges in ROIs are smaller than the intensity range of the whole image. For example, for 16 bit coded images there are 65536 intensity levels, so normalization to the full range will produce a sparse his-

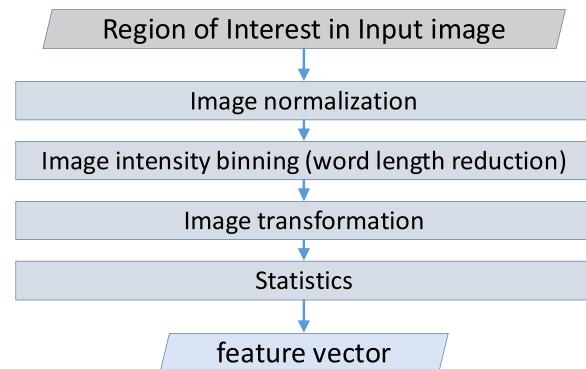


Fig. 1. Typical procedure for obtaining texture-based features.

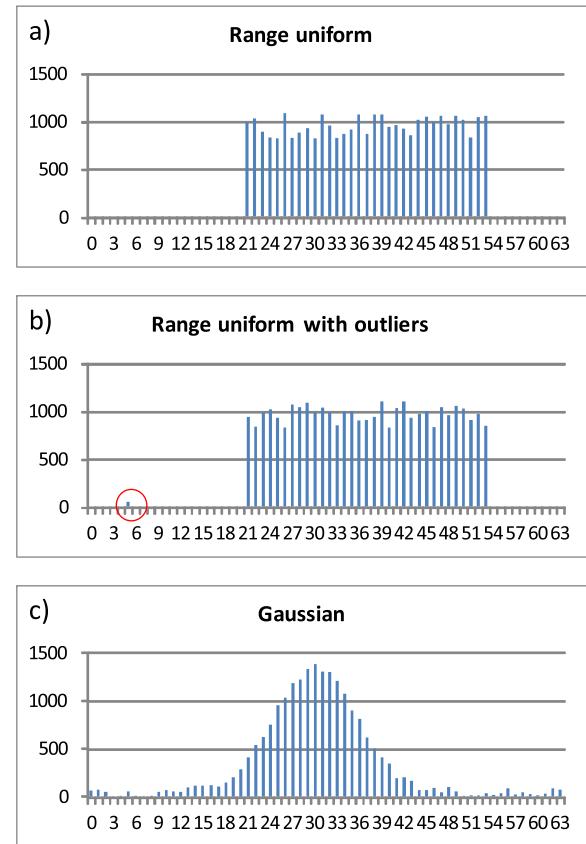


Fig. 2. Example of image histograms suitable for different normalization techniques.

togram. Moreover, in the case of features based on the gray level co-occurrence matrix (GLCM) (Haralick, 1979; Haralick et al., 1973), which is a two-dimensional matrix, the matrix for 16 bit images will have $65\,536 \times 65\,536 = 4294\,967\,296$ elements. A matrix of this size will require an unnecessary increase in the computation time needed to estimate GLCM-based features.

Normalization and reduction of intensity levels are very often performed simultaneously. It is worth noticing that for some software packages the number of intensity levels is reduced to the range $(0, \dots, 2^k - 1)$. The procedure can be described by Eq. (1)

$$I_{norm}(x, y) = \begin{cases} 2^k - 1 & \text{for } N(x, y) > 2^k - 1 \\ N(x, y) & \text{for } 0 \leq N(x, y) \leq 2^k - 1 \\ 0 & \text{for } 0 < N(x, y) \end{cases} \quad (1)$$

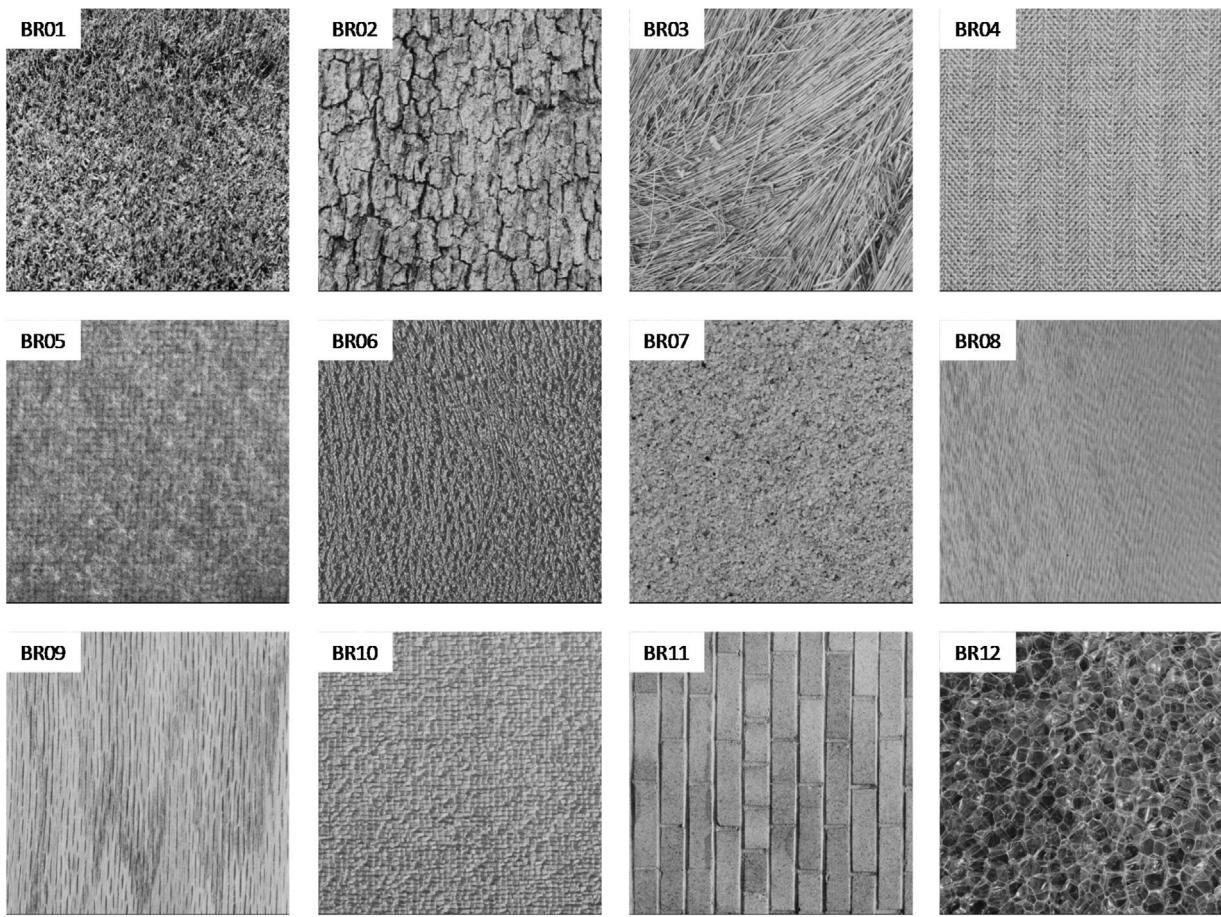


Fig. 3. Images from the Brodatz album (Brodatz, 1966) downloaded from (“SIPPI Image Database - Textures,” n.d.).

where:

$$N(x, y) = \text{round_to_int} \left(\frac{I(x, y) - \min_{\text{norm}}}{\max_{\text{norm}} - \min_{\text{norm}}} (2^k - 1) \right)$$

\min_{norm} – minimum normalized value,

\max_{norm} – maximum normalized value,

k – no. of image bits per pixel after normalization.

The image normalization techniques investigated in this study were as follows:

min-max – in this type of normalization, the \min_{norm} and the \max_{norm} from Eq. (1) are the minimum and maximum intensities taken directly from the histogram. Such normalization is useful if the histogram of the investigated texture is range uniform (Fig. 2a).

1%–99% – for such normalization $\min_{\text{norm}} = \arg(\text{cumulative}_{\text{histogram}} = 1\%)$ and $\max_{\text{norm}} = \arg(\text{cumulative}_{\text{histogram}} = 99\%)$.

This type of normalization should be useful, for example, for range-uniform intensity distributions with limited numbers of outliers (Fig. 2b). Such outliers are very often caused by dead pixels in imaging sensors.

$\pm 3\sigma$ – the range of intensities for this normalization can be defined as $\min_{\text{norm}} = \mu - 3\sigma$ and $\max_{\text{norm}} = \mu + 3\sigma$, where μ is the mean intensity and σ is the standard deviation of the image intensities in the ROI. Such normalization is useful when the intensity histogram of the texture is in close to a Gaussian distribution.

3. Material and methods

The analyses described in this study were performed on four sets of images.

3.1. Test set 1 – Brodatz images

The first set consisted of modified images from the Brodatz album (Brodatz, 1966).

Twelve textures from the Brodatz album were used to build the test set (Fig. 3). Originally, 256 Gy level images with dimensions 512×512 pixels were downloaded from (SIPPI, 2020).

Fig. 4 shows the intensity histogram for an example Brodatz image without any added transformations or distortions.

The original images were converted into 32 bit signed format and the following image distortion combinations were applied to each image:

- no distortion;
- Gaussian noise with standard deviations of 6.4, 12.8 and 25.6, constituting respectively 2.5%, 5% and 10% of the full intensity range of the original image;
- Rician noise with an s parameter of 6.4, 12.8 and 25.6, constituting 2.5%, 5% and 10% of the full intensity range of the original image;
- low frequency background intensity non-uniformity in the form of an added intensity gradient from the top left to the bottom right of the image;
- a combination of Gaussian noise and background intensity non-uniformity,
- a combination of Rician noise and background intensity non-uniformity.

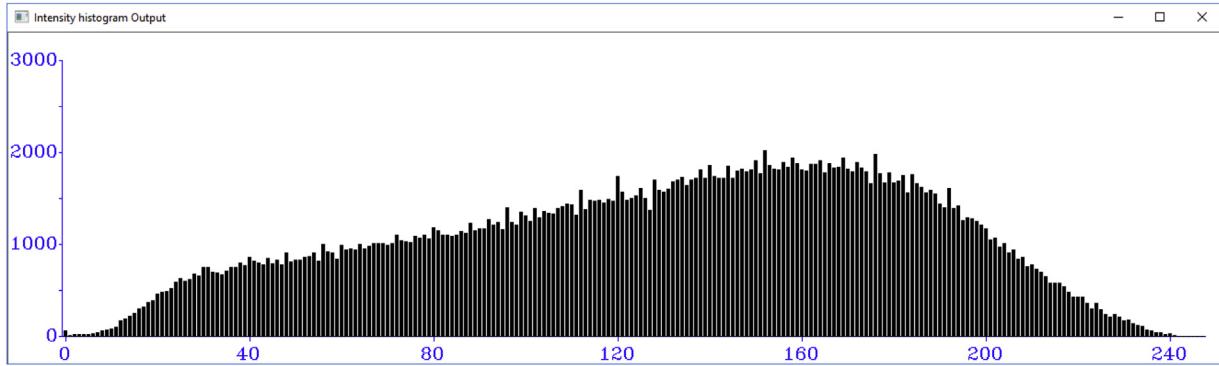


Fig. 4. Intensity histogram for the example image (BR01).

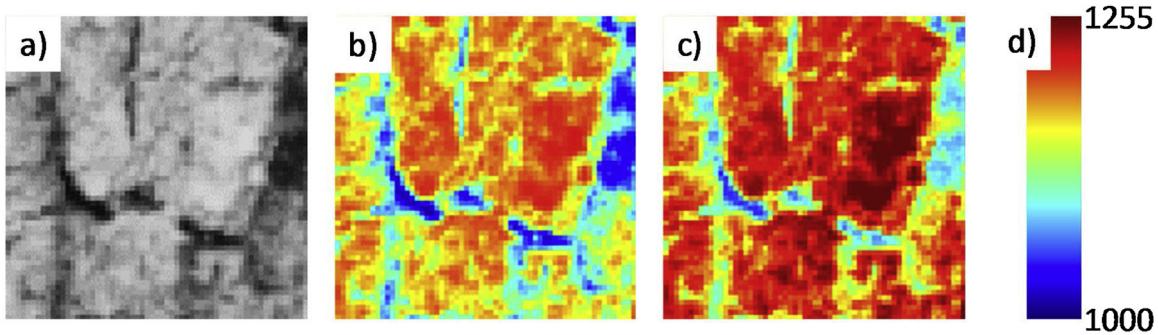


Fig. 5. Small piece of the example test image(Br02): a) original, b) offset of 1000 added, c) image with background intensity non-uniformity and offset added d) the color scale for pseudo color representation.

Additive Gaussian noise was obtained by the application of Eq. (2),

$$I_{GN}(x, y) = I(x, y) + \text{rand}(\mu = 0, \sigma = 1) \cdot k \quad (2)$$

where:

$I_{GN}(x, y)$ – intensity of the image with Gaussian noise,

$I(x, y)$ – intensity of the original image,

$\text{rand}(\mu = 0, \sigma = 1)$ – a randomly generated value with a Gaussian distribution of mean $\mu = 0$ and a standard deviation $\sigma = 1$,

k – the noise level coefficient.

Rician noise was generated in the way described in (Ridgway, 2020) by means of Eq. (3)

$$I_{RN}(x, y) = \sqrt{(s \cdot \text{rand}(\mu = 0, \sigma = 1) + I(x, y))^2 + (s \cdot \text{rand}(\mu = 0, \sigma = 1))^2} \quad (3)$$

Where:

$I_{RN}(x, y)$ – intensity of image with Rician noise,

$I(x, y)$ – intensity of the original image,

$\text{rand}(\mu = 0, \sigma = 1)$ – a randomly generated value with a Gaussian distribution of mean $\mu = 0$ and standard deviation $\sigma = 1$;

s – the noise level coefficient.

Background intensity non-uniformity was created by the addition of an intensity gradient to the image, according to Eq. (4).

$$I_{BNU}(x, y) = I(x, y) + 0.5 \cdot x \cdot y \quad (4)$$

After the application of distortion, an intensity offset of 1000 was added to the image in order to remove negative intensity values. The resulting images were converted and stored in 16 bit Tagged Image File format. Fig. 5 shows a small piece of an example test image (taken from the upper left corner of texture Br02 from Fig. 3) accompanied by the image with an offset of 1000 added and the same image combined with offset and intensity non-uniformity (the latter two represented in pseudocolors for better visualization of the introduced artifacts).

Fig. 6 shows other distortion combinations applied to the image from Fig. 5a.

In total, 13 combinations of distortion were applied to each of the 12 base images. For each image, a set of 64 uniformly distributed non-overlapping ROIs was defined. Each ROI is a square of 61×61 pixels. Of the 64 test ROIs, 13 were used as a training set, and remaining 51 were used as the test set (Fig. 7).

Fig. 8 shows histograms for images without distortion for an ROI located close to the center of the image (ROI #28 marked on Fig. 7).

As can easily be seen, each image has a different histogram and none of the histograms follows closely the histograms from Fig. 2. Fig. 9 shows histograms of an example image (mainly Br02) and selected ROIs after the application of 5% Rician noise and background non-uniformity.

Clearly, noise modifies the shape of the histogram and low frequency background non-uniformity offsets intensities inside the ROI.

3.2. Test set 2 – kidney images (DCE MRI)

The second set analyzed in this work consisted of 130 dynamic contrast-enhanced magnetic resonance images (DCE-MRI). This was a subset of the images described in (Eikefjord et al., 2017) and used in (Kociołek et al., 2019). The original set consists of 20 sequences taken for 10 healthy nonsmoking volunteers acquired by means of a 1.5 T Siemens Magnetom Avanto scanner (Erlangen, Germany) using a standard phased-array coil. Two acquisitions for each volunteer were taken at intervals of 7 days. Each time sequence includes 74 volumetric images. Each volumetric image was acquired by means of a 3D spoiled gradient-recalled (SPGR) pulse-sequence with the following parameters: flip angle: $FA = 20^\circ$; echo time $TE = 0.8\text{ ms}$; repetition time $TR = 2.36$; parallel imaging factor $PIF = 3$; Field of view $425\text{mm} \times 425\text{mm} \times 90\text{mm}$; Voxel size $2.135\text{mm} \times 2.135\text{mm} \times 3\text{mm}$. This results in a 3D vol-

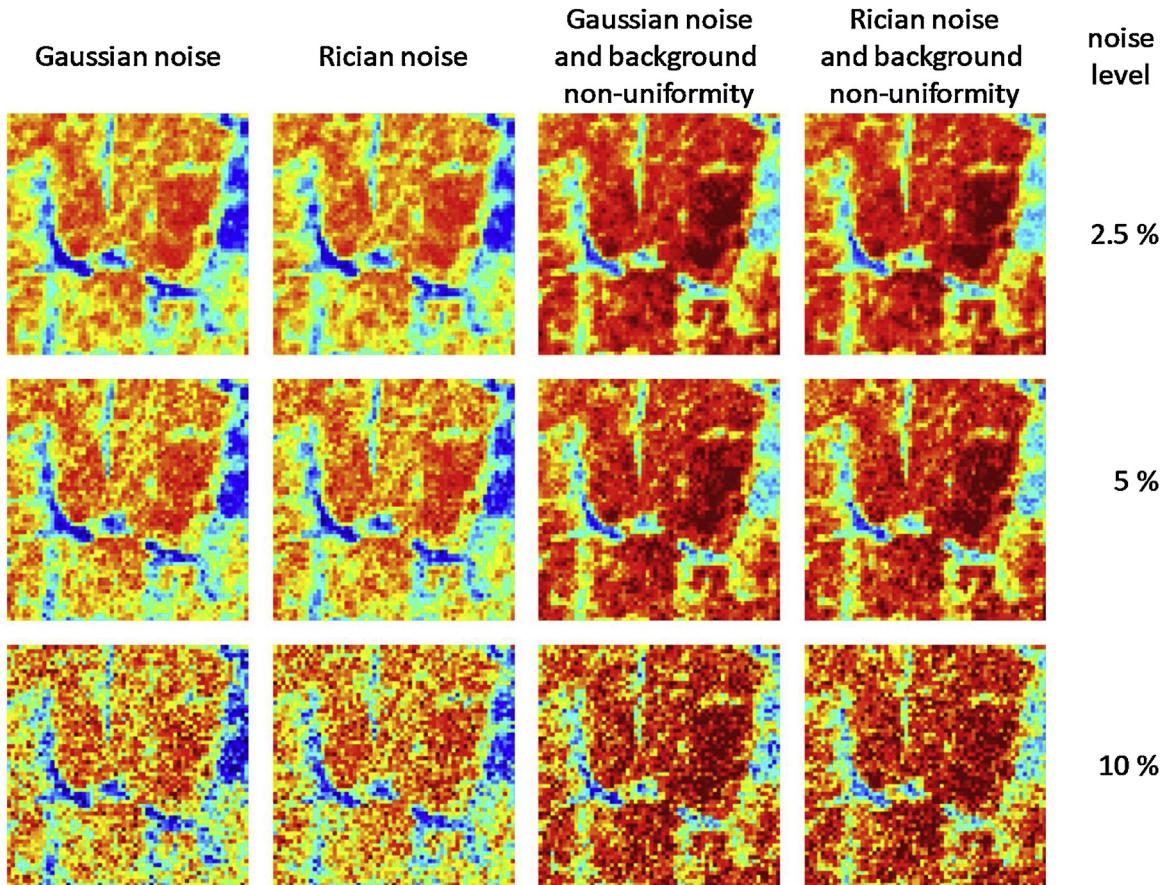


Fig. 6. Distortion combinations for the image from Fig. 5a. The color scale is the same as in Fig. 5d.

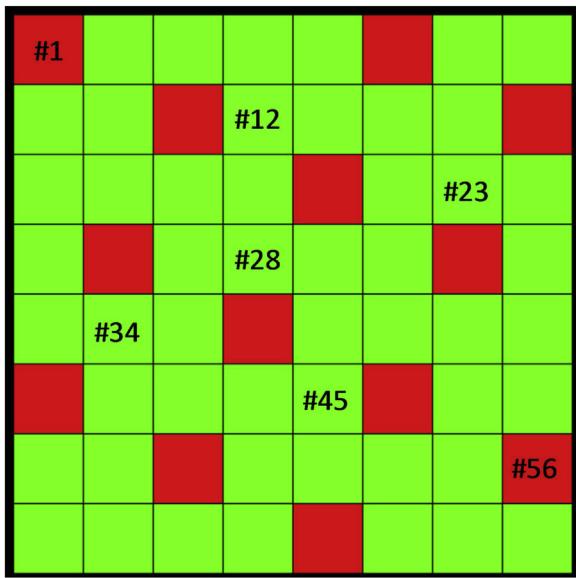


Fig. 7. ROIs for Brodatz images. The red ones were used as a training set while the green ones were used as a test set (For interpretation of the references to colour in the Figure, the reader is referred to the web version of this article).

ume of $192\text{pixels} \times 192\text{pixels} \times 30\text{pixels}$, which can be treated as 30 slice images with dimensions of $192\text{pixels} \times 192\text{pixels}$. The data were obtained from the Haukeland University Hospital Bergen, Norway. All processed image data were anonymized before being handed to us. All volunteers gave their written informed consent for participation in the examinations. The study protocol was approved

by the Institutional Review Board at the Haukeland University Hospital Bergen, Norway.

Each time sequence captures the filtration action of the kidneys. A bolus injected into the bloodstream of the patient is filtered out to the urinal tract. As it passes through the kidney structures, it causes changes in the kidney image. Although the kidney region did not constitute homogeneous region in (Kociołek et al., 2019) we found that the values for multiple textural features are related to the state of the kidney during the filtration process. In this work, we are therefore using textural features which apply to the whole kidney region. Each image was accompanied by a corresponding ROI mask. These masks were prepared semi-manually. We noticed that for a given slice the shape of the kidney remained unchanged during respiratory movements. Thus, the shape of the kidney was manually outlined for the first timeframe. For consecutive frames, only the position of the mask was adjusted manually in order to match the position of the kidney. For each time sequence, slices containing a clear view of the medulla and pelvis were selected. Usually there were two such slices in each time sequence. This gives 2960 images ($2\text{slices} \times 74\text{time frames} \times 2\text{sequences} \times 10\text{patients}$). From the available images, we handpicked 130 belonging to one of three classes:

- no contrast agent in the kidney (43 images);
- contrast agent in the right kidney cortex (42 images);
- contrast agent in the pelvis (45 images);

A half of the images from each class were selected randomly as the training set. The other half was used as a test set.

Fig. 10 shows an example of DCE-MRI with the superimposed contour of the right kidney ROI for two volunteers.

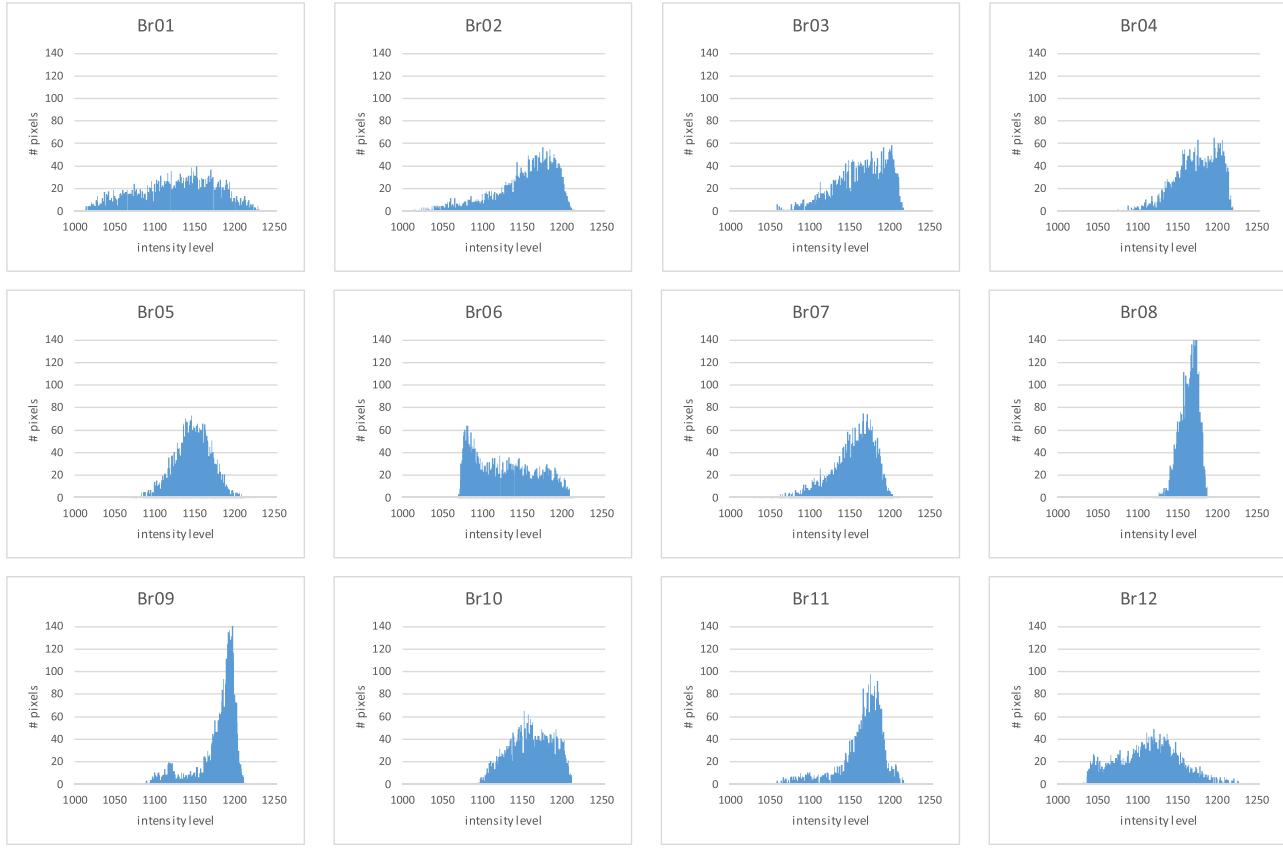


Fig. 8. Histograms for a 61×61 pixels ROI(#28) located close to the center of the image. Histograms were plotted for images without noise or distortions.

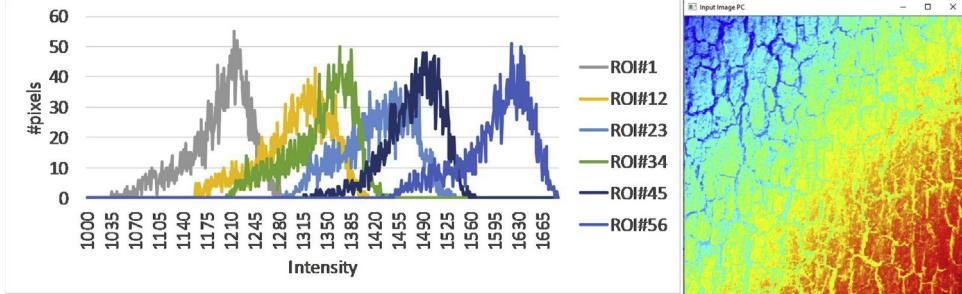


Fig. 9. Example histograms for selected ROIs in image BR02 distorted by 5% Rician noise and background non-uniformity.

[Fig. 11](#) shows example histograms obtained for the right kidney region in the three considered classes of DCE-MRI images

3.3. Test set 3 – shoulder images (MRI T2 weighted)

The third set analyzed in this work was composed of T2 weighted MRI images acquired with different acquisition times. The study protocol was designed according to the guidelines of the Declaration of Helsinki and the Good Clinical Practice Declaration Statement. Special care was taken regarding personal data safety, with all images anonymized before processing. Written acceptance to conduct the study was obtained from the Ethics Committee of the Jagiellonian University (no.1072.6120.15.2017 dated: 20.06.2017). This was a subset of images described in ([Obuchowicz et al., 2019](#)). The original set consists of 12 T2-weighted coronal volumetric images of the shoulder taken from patients between the ages of 23 and 56 years old. Volumetric images were acquired using a 1.5 T Siemens Essenza (Erlangen, Germany) equipped with 12 dedicated table coils and 8 channel shoulder coils. Regular GRAPPA1

and three shortened sequences, GRAPPA2, GRAPPA3 and GRAPPA4, were performed using parallel imaging. Each volumetric image for the GRAPPA1 was acquired by means of a T2 weighted sequence with the following parameters: flip angle $FA = 150^\circ$; echo time $TE = 101\text{ ms}$; repetition time $TR = 3.54\text{ ms}$; phase oversampling of 100; slice spacing 3.9 mm ; field of view $200\text{ mm} \times 200\text{ mm} \times 60\text{ mm}$; voxel size $0.625\text{ mm} \times 0.625\text{ mm} \times 3\text{ mm}$. GRAPPA2 has acquisition time reduced by 2, GRAPPA3 has the acquisition time reduced by 4, and GRAPPA4 has the acquisition time reduced by 8. Each GRAPPA sequence results in a 3D volume of $320\text{ pixels} \times 320\text{ pixels} \times 20\text{ pixels}$, which can be treated as 20 slice images with dimensions of $320\text{ pixels} \times 320\text{ pixels}$. As can be seen from [Fig. 12](#), reducing the acquisition time deteriorates image quality, introducing distortions and noise.

We handpicked 71 images of the shoulder. On each image, three circular ROIs with diameters of 21 pixels were marked ([Fig. 12](#)). These contained the humerus epiphysis (bone), muscle and fat regions, respectively. The training set consisted of 36 images. The remaining 35 images were used as test set.

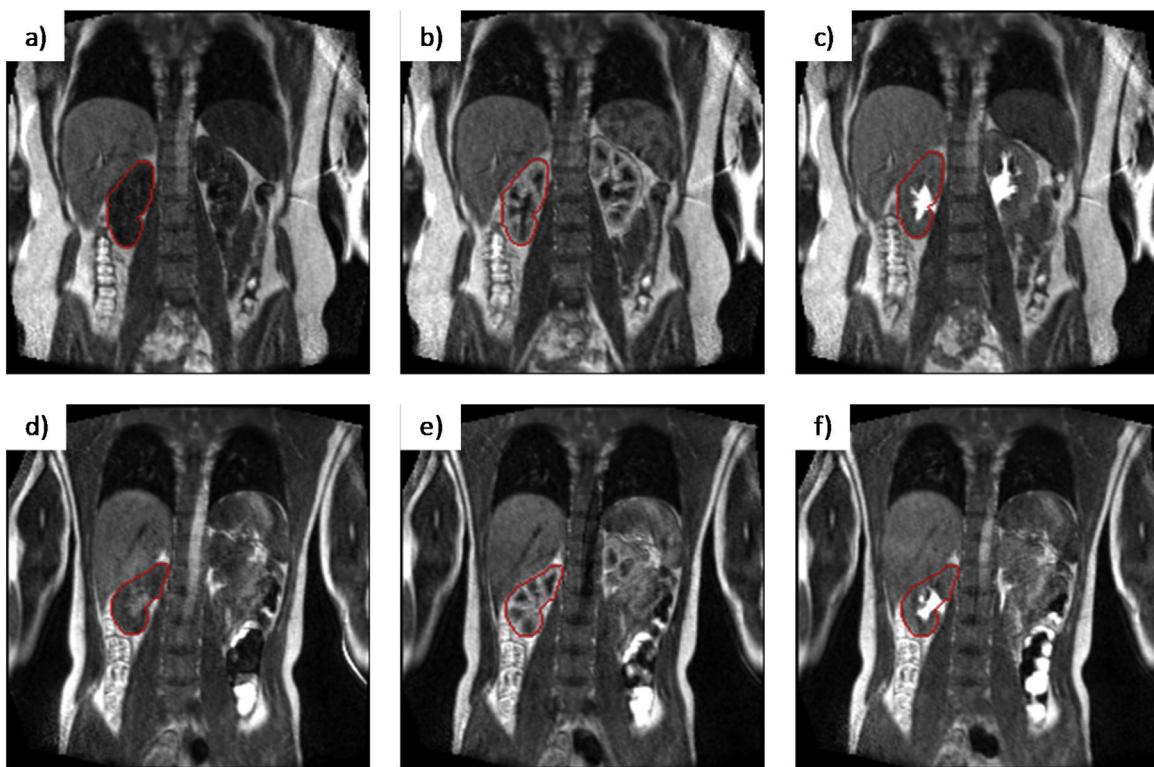


Fig. 10. Examples of dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) with superimposed contour of the right kidney ROI for two volunteers a, d) no contrast agent in the kidney b, e) contrast agent in the cortex c, f) contrast agent flushed to the pelvis.

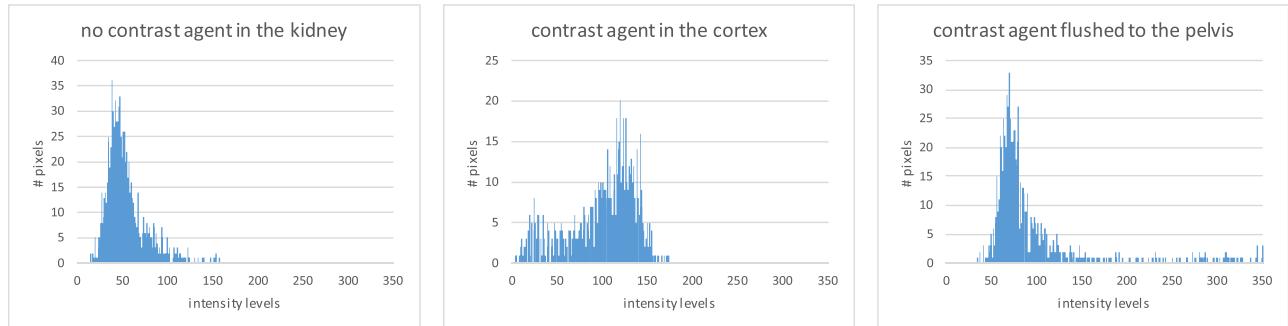


Fig. 11. Example of histograms obtained for a right kidney region in DCE-MRI images.

3.4. Test set 4 – thorax images (CT)

The last set analyzed in this work consisted of heart CT cross-sections obtained from the same raw data for different reconstruction filters. Images were acquired using a SIEMENS STOMATOM CT scanner (Erlangen, Germany). The parameters used were as follows: X-ray tube current: $TC = 1082\mu A$; X-ray tube voltage: $TV = 120kV$; exposure time $TE = 285\text{ ms}$; generator power $P = 65\text{ W}$; slice thickness 0.6 mm ; Field of view $205mm \times 205m$; pixel size $0.4mm \times 0.4mm$; image size 512×512 . Three different reconstruction kernels were implemented, with different degrees of image sharpness and different purposes (depending on the type of organ for which they provided the best visualization). These were: HeartView smooth (B26f), Medium smooth (B30f), and HeartView medium (B36f). Sharper filters generate higher noise in the reconstructed images. To partly remove this noise, the SAFIRE iterative reconstruction-based filter was used at level 3 (Grant and Raupach, 2012). It was demonstrated in (Völgyes et al., 2017) that the application of different reconstruction ker-

nels (resulting in different noise levels) significantly affects the resulting image quality, including the possibility of low contrast tissue detection. The influence of CT heart image quality (expressed in terms of applied reconstruction kernels) on the possibility of building accurate 3D/4D model of the heart has been investigated in (Bielecka and Piórkowski, 2015). Our aim was to determine whether image normalization can influence tissue classification in such data, assuming different noise levels are generated by the kernels.

Particular care was taken to ensure the safety of personal data. All processed image data were anonymized before being handed to us. Written informed consent for the processing and publication of anonymized clinical images was obtained from the management of the Department of Diagnostic Imaging at Jagiellonian University Medical College. The usual requirement for informed consent from patients was waived in view of the retrospective nature of the research.

Five slices were obtained for two patients. On each image, three sets of circular ROIs with diameters of 35 pixels were marked

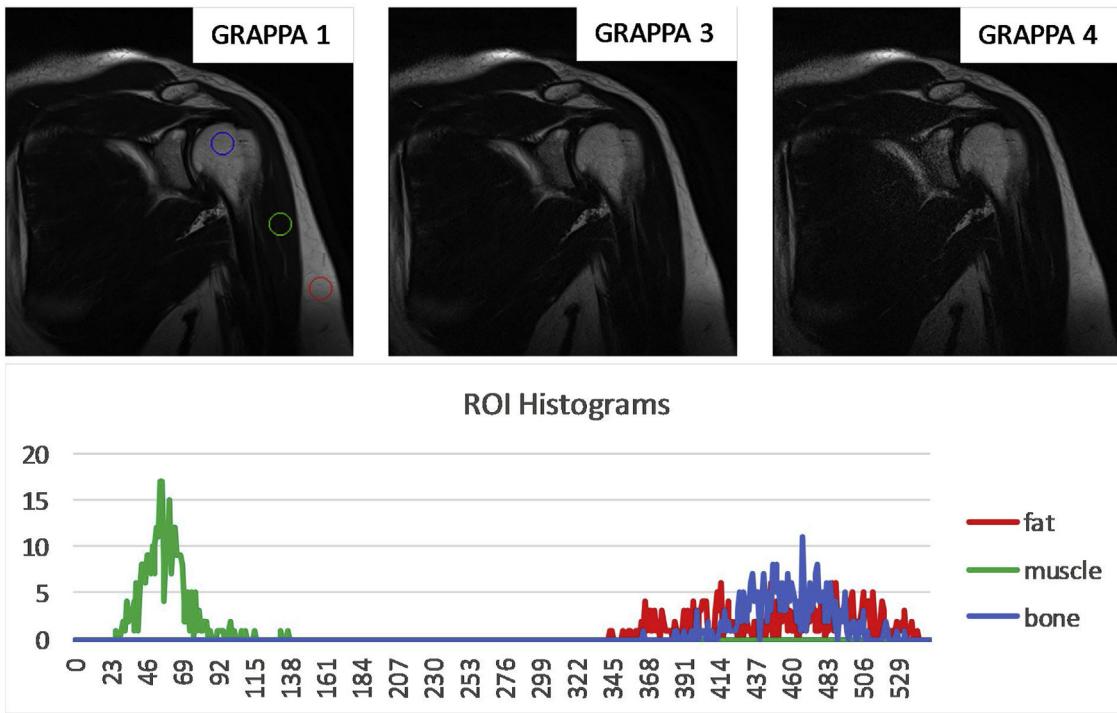


Fig. 12. T2 weighted images of the shoulder acquired with three different sequences GRAPPA 1, GRAPPA3 and GRAPPA4 and histograms of example ROIs with three ROIs marked: red circle - fat, green circle - muscle, blue circle - bone (For interpretation of the references to colour in the Figure, the reader is referred to the web version of this article)..

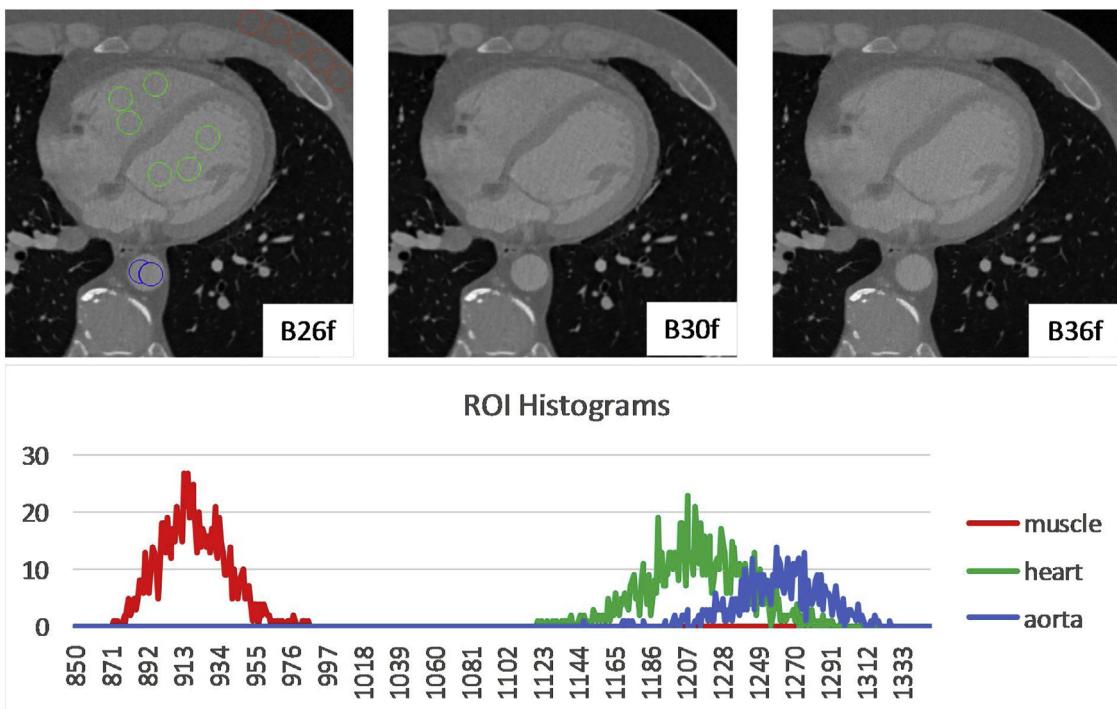


Fig. 13. CT images of a thorax obtained using three reconstruction filter kernels B26f, B30f and B36f and histograms of example ROIs. Three ROI groups are marked on the B26f image: red circles - muscle, green circles - heart, blue circles - aorta (For interpretation of the references to colour in the Figure, the reader is referred to the web version of this article).

Fig. 13a. one containing the muscle region, the second containing the heart region and the third containing the aorta region. The ROIs for the muscle and heart did not overlap, while the ROIs for the aorta partially overlapped. In total, there were 45 ROIs in the mus-

cle region, 56 ROIs in the heart region and 20 ROIs in the aorta region. Five images, three slices from the first patient and two from the second, were used as a training set. The remaining five images were used as a test set.

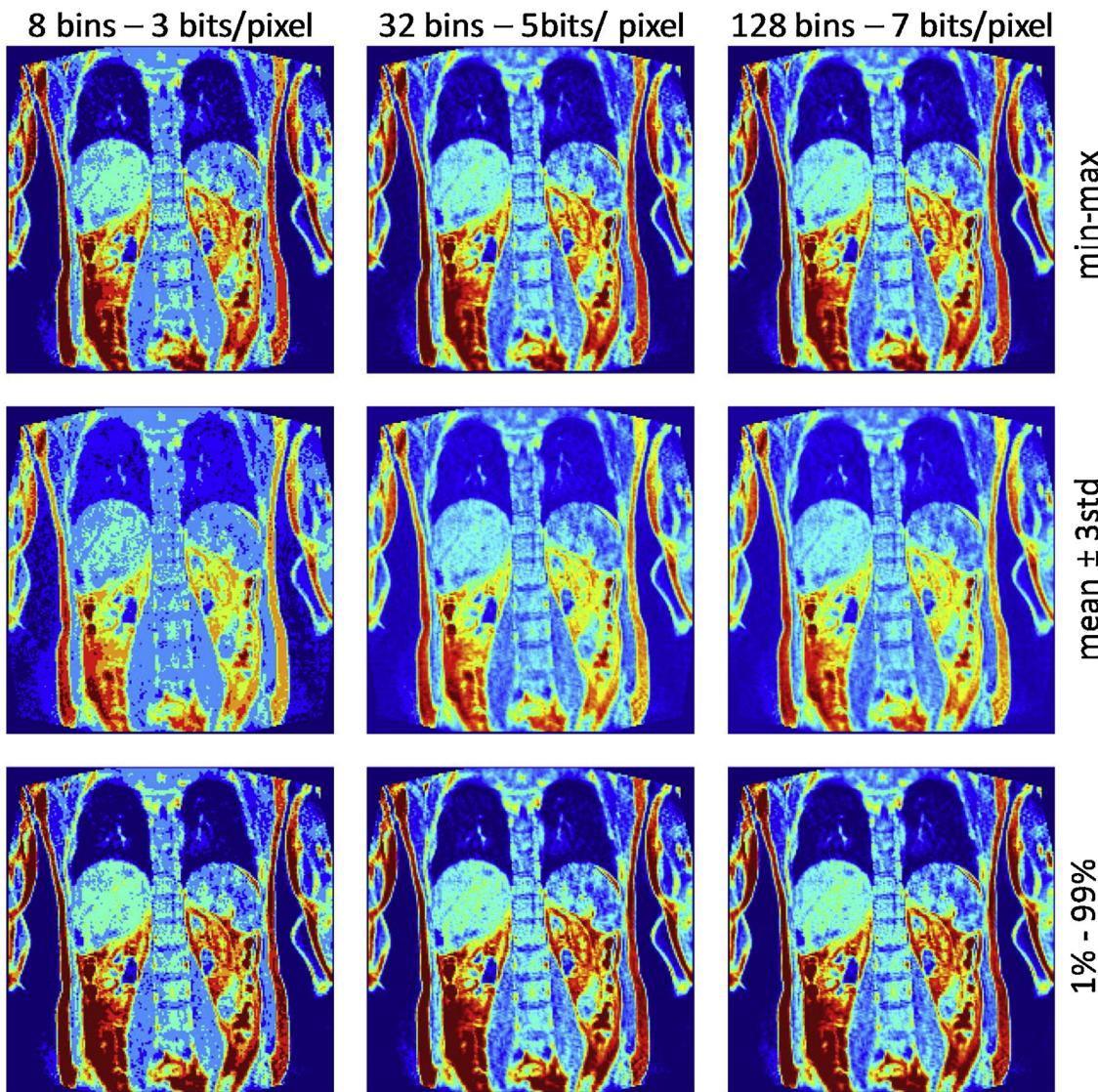


Fig. 14. Example image after different combinations of the normalization scheme and with various intensity level numbers. The images were normalized according to the intensity distribution in ROI containing the right kidney area.

3.5. Test methodology

Three normalization schemes were tested, min-max, 1 %–99 % and $\pm 3\sigma$, followed by a reduction in the intensity level number (intensity binning/quantization) to 8, 16, 32, 64, 128 and 256 levels (corresponding to 3, 4, 5, 6, 7 and 8 bits coding intensity). Fig. 14 shows an example image from the MR kidney test set after the application of normalization methods with 8, 32, and 128 intensity levels. Normalization was performed based on the intensity distribution for the right kidney region.

Fig. 15 shows histograms of the right kidney region for images from Fig. 14.

For each image, ROI and combination of normalization scheme and number of intensity bins, the following set of textural features was calculated:

- 24 based on the Gabor Transform – Gabor transforms were created for a combination of 4 directions (0° , 45° , 90° , 135°) and 6 sizes of Gaussian envelope (4, 6, 8, 12, 16, 24);
- 220 based on the Grey Level Co-occurrence Matrix (GLCM) – 11 features for GLCM were built for a combination of 4 directions (0° , 45° , 90° , 135°) and 5 distances/offsets (1, 2, 3, 4, 5);

- 24 based on the Grey-Level Run-Length Matrix (GRLM) – 6 features for GLRLMs were built for 4 directions (0° , 45° , 90° , 135°);
- based on the Gradient Map;
- based on the Histogram of Oriented Gradients (HOG) – 3 angular bins and 4 orientations;
- based on 2D Harr Wavelet Transform (3 filters and 4 scales).

Feature calculations (including normalization and quantization) were obtained using QMaZda software (Szczypinski et al., 2017). A detailed description of the features can be found in (Szczypinski, 2018).

For each type of distortion, a series of classification experiments was performed involving features calculated for each normalization scheme – quantization level combination. The classification experiments were based on linear discriminant analysis built into the MzReport tool from the QMaZda package. The classification experiment involved two phases:

- supervised training performed on the training set,
- a testing phase performed on the testing set.



Fig. 15. Histograms of kidney region for example image created for different combinations of normalization method and number of intensity bins.

During the training phase, a series of binary classifiers was trained separately for each pair of classes. The following algorithm was used in the training phase for each classifier:

- 1) All feature vectors associated with the training set were loaded into the MZReport tool.
- 2) All feature values were standardized according to their mean and standard deviation.
- 3) The best combination of three features was found for each pair of classes according to their discriminative power using LDA.

The following linear classification training settings were utilized:

- normalization: standardize,
- threshold: Bayes,
- direction: LDA,
- time: 2400s,
- dimensionality: 3.

As a result of the training phase, a set of $(K^2-K)/2$ classifiers for K classes was found, each utilizing the best ranked combination of three features.

4. Results

Classifiers trained during the training phase were used in the testing phase of the experiment. In the testing phase, feature vectors calculated for the test set were fed to the set of classifiers. This resulted in a confusion matrix, shown in Fig. 16.

The sample under consideration was assigned to a given class if all $(K-1)$ classifiers for the class agreed. Otherwise, no decision was reported. This situation is also treated as a classification error ("no decision" column in Fig. 16).

Fig. 17 presents the results of classification experiments for Brodatz images distorted by means of Rician noise and background non-uniformity (detailed results for all considered distortions can be found in the Supplementary Information). The classification error is reported for each combination of normalization technique and intensity level number separately for three noise levels (2.5%, 5% and 10%) and six feature groups.

Class	Classification result												
	Br01	Br02	Br03	Br04	Br05	Br06	Br07	Br08	Br09	Br10	Br11	Br12	No decision
Br01	51												
Br02		44									4	3	
Br03			47									4	
Br04				51									
Br05					51								
Br06	2					49							
Br07							50					1	
Br08								51					
Br09									51				
Br10										51			
Br11											47	4	
Br12											50	1	

Fig. 16. Example confusion matrix obtained in the testing phase of the classification experiment for Brodatz images with added Gaussian noise with standard deviation = 6.4 (2.5 % of original intensity range). Empty cells correspond to 0 values. The analysis was performed for GLCM features calculated for $\pm 3\sigma$ normalization and 256 intensity levels.

In the case of Gabor-based features $\pm 3\sigma$ normalization gives the best classification results. Changing the number of intensity levels has a limited influence on the classification results, although there is a slight increase in error for a small number of coding intensity levels. This set of features was relatively immune to moderate levels (2.5 % and 5%) of noise and to background non-uniformity.

For GLCM-based features, there was no clear trend in terms of classification error versus normalization method and number of intensity levels. However, the lowest classification error was observed for GLCM-based features. The maximum observed error was below 10% regardless of the number of intensity levels and normalization scheme. For low and moderate noise levels (2.5% and 5%, Rican and Gaussian) the maximum observed error was below 6%. Similarly, the differences in classification error for different combinations of normalization method and number of intensity levels was low, with a maximum equal to 6%.

For gradient features, better classification results were obtained mostly for large numbers of intensity levels (128 and 256). But there was no major improvement in comparison to low numbers of intensity levels. With moderate noise levels (2.5% and 5% Rican and Gaussian) without background non-uniformity, 1%-99% normalization performs better than other two methods. For moderate noise levels with background non-uniformity, $\pm 3\sigma$ normalization gives the lowest classification error. With higher noise levels, 1%-99% normalization performs similarly to $\pm 3\sigma$ normalization and both methods perform better than min-max normalization.

The best classification results for GLRM-based features in all combinations of the tested distortion were obtained for the lowest numbers of intensity levels. All normalization schemes performed similarly, although $\pm 3\sigma$ normalization gave the lowest classification error in 8 out 14 cases.

For HOG-based features with low noise level (2.5%) and a combination of moderate noise level (5%) and background non-uniformity, the best classification results were obtained in the case of 1%-99% normalization. Slightly worse results were obtained for $\pm 3\sigma$ normalization. In the case of moderate noise levels without background non-uniformity and with a high noise level (10%), $\pm 3\sigma$ normalization performed slightly better than 1%-99% normalization. In most cases, min-max normalization gave the worst classification. The number of intensity levels had a limited influence on the classification results, but in most cases higher numbers of intensity levels gave slightly better classification.

For wavelet-based features, with a low noise level (2.5%) min-max normalization usually gave the best classification results. For moderate and large noise levels, the best results were usually obtained using $\pm 3\sigma$ normalization. Changing the number of intensity levels had a limited influence on the classification outcome, but in some cases the use of 8 intensity levels introduced a slightly higher classification error.

Fig. 18 shows the results of classification experiments performed on DCI - MRI of kidneys.

The kidney images were obtained by means of the DCI-MRI technique, so Rican noise may be present. For Gabor transform-based features, the lowest classification error was observed for min-max normalization and 32 or 64 intensity bins. For GLCM features, the best results (classification error equal to 0) were obtained for all normalization methods with 8–64 intensity levels and 1%-99% normalization with 128 intensity levels as well as min-max normalization with 256 intensity levels. For gradient-based features, the lowest classification error was with 1%-99% normalization and 32–256 intensity levels. For GLRM-derived features, the best classification results were obtained for $\pm 3\sigma$ normalization with 8 intensity levels. A little higher but comparable error level was obtained for $\pm 3\sigma$ normalization with 64 and 256 intensity levels. For HOG-based features, the lowest classification error was obtained for min-max normalization with 8 intensity levels. For wavelet-based features, the best classification was obtained for min-max normalization with 16 intensity levels.

Fig. 19 shows classification error in the analysis of shoulder MRI images acquired for GRAPPA 1, GRAPPA 3 and GRAPPA 4 sequences. Only GLCM, GLRM and HOG-based features are presented here, since for other feature families the minimal classification error was well over 40%.

For GLCM-based features, min-max normalization with 256 intensity levels gave the lowest classification error for GRAPPA 1 sequence. Similar but slightly higher classification error can be obtained for $\pm 3\sigma$ normalization with the same number of intensity levels. In the case of GRAPPA3 sequence the best classification results were achieved for 1%-99% normalization with 8 intensity levels. However, the normalization $\pm 3\sigma$ with 256 intensity levels gave similar result. In the cases of GRAPPA4 sequence the best classification results were achieved for min-max normalization with 32 intensity levels.

In the cases of GRAPPA1 and GRAPPA3 sequences, for GLRM-based features the best classification results were achieved for



Fig. 17. Classification error results of classification experiments for Brodatz images distorted by means of Rician noise and background non-uniformity.

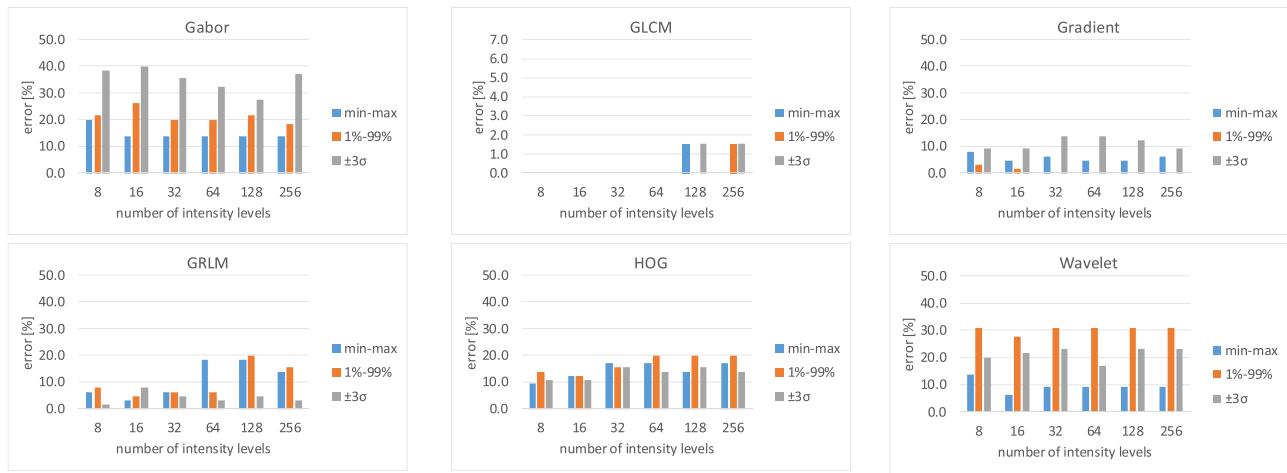


Fig. 18. Classification error for kidney Images presented separately for different feature groups.

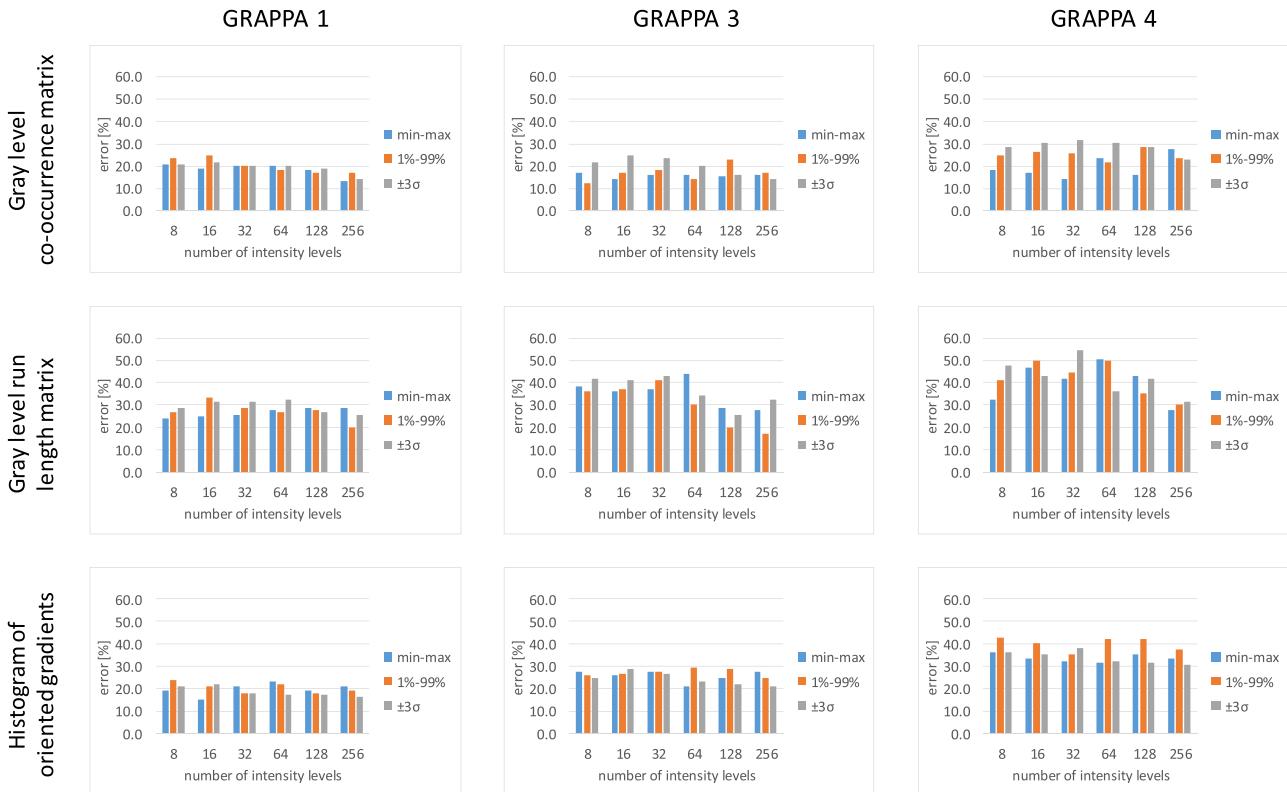


Fig. 19. Classification error plotted for GRAPPA 1, GRAPPA 3 and GRAPPA 4 MRI sequences with shoulder images.

1 %–99 % normalization with 256 intensity levels. In the case of GRAPPA4 the best classification was obtained for min-max normalization with 256 intensity levels.

For histogram of oriented gradients, min-max normalization with 16 intensity levels gave the lowest classification error for GRAPPA 1 sequence. Similar but slightly higher classification error can be obtained for $\pm 3\sigma$ normalization with 256 intensity levels. In the cases of GRAPPA 3 and GRAPPA 4 sequences the best classification results were achieved for $\pm 3\sigma$ normalization with 256 intensity levels.

Fig. 20 shows classification error in the analysis of thorax CT images obtained by means of three reconstruction filters, B26f, B30f and B36f. As in the case of the MRI shoulder results, only GLCM, GRLM and HOG-based features are presented here because in other

feature families the minimal classification error was well over 30 %.

For GLCM-based features, 1 %–99 % normalization with 256 intensity levels gave the lowest classification error for B26f reconstruction filter. In the case of B30f filter the lowest error was obtained for min-max normalization with 256 intensity levels. However, 1 %–99 % normalizations with the same number of intensity levels gave similar result. For B36f reconstruction filter min-max normalization with 128 intensity levels allow for the lowest classification error. Slightly higher classification error can be obtained for 1 %–99 % normalization with 256 intensity levels.

In the case of B26f reconstruction filters, the best classification result for GRLM-based features was obtained with 1 %–99 % normalization and 8 intensity levels. However, the 1

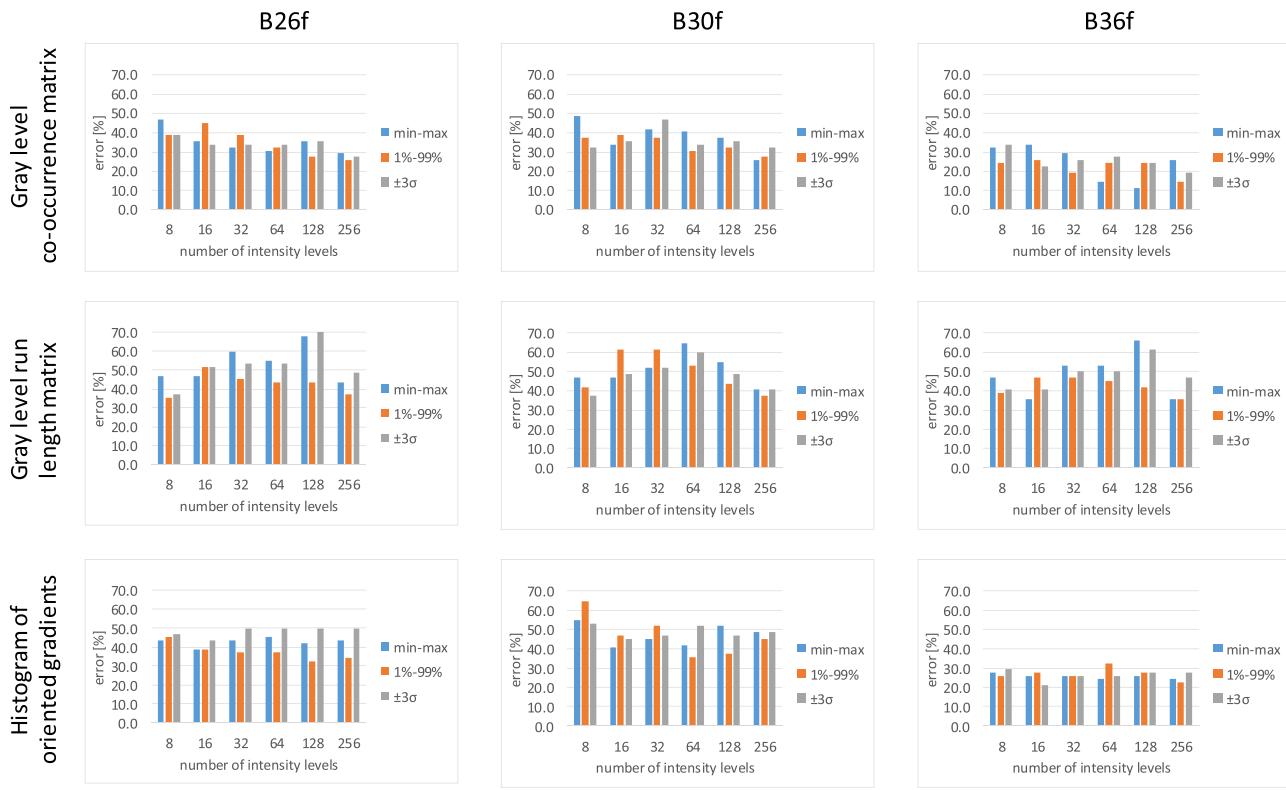


Fig. 20. classification error plotted for analysis of thorax CT images obtained by means of three reconstruction filters, B26f, B30f and B36f.

%–99 % normalizations with the 256 intensity levels gave similar result. With the B30f filter, the best classification was obtained using $\pm 3\sigma$ normalization and 8 intensity levels. Similar but slightly higher classification error can be obtained for 1 %–99 % normalization with 256 intensity levels. In the case of B36f reconstruction filters, the best classification results were achieved with min-max normalization and 16 intensity levels as well as min-max and 1 %–99 % normalizations with 256 intensity levels.

For HOG-based features and images obtained using the B26f reconstruction filter, the best classification results were obtained with 1 %–99 % normalization and 128 intensity levels. In the case of B30f filter the lowest error was achieved for 1 %–99 % normalization with 64 intensity levels. However, the 1 %–99 % normalizations with 128 intensity levels gave similar result. For images obtained using the B36f reconstruction filter, the best classification results were achieved with $\pm 3\sigma$ normalization and 16 intensity levels.

5. Discussion and summary

It should be emphasized that changing the normalization technique or the number of intensity levels may lead to a completely new texture feature set. This means that a given feature formula calculated for a different normalization technique and/or number of intensities will result in different numerical values. Moreover, such feature values may be uncorrelated. Fig. 21 shows three combinations of normalization/number of intensity levels for the sum of squares feature, calculated based on GLCM, created for direction = 0° and offset = 1. All three combinations of normalization/number of intensity level had different feature values for given image/ROI combinations. The Pearson correlation coefficient between the min-max normalization with 256 intensity levels and the min-max normalization with 128 intensity levels is equal to 1.0. Between the min-max normalization with 256 intensity lev-

els and $\pm 3\sigma$ normalization with 256 intensity levels the Pearson correlation is equal to 0.55.

Table 1 shows the percentage of features with a Pearson correlation coefficient greater than 0.9 for different combinations of normalization and numbers of intensity levels, with min-max normalization and 256 intensity levels as the reference.

Some features (Gabor, WDT and HOG) were resistant to gray level reduction, while others (GRLM and GLCM) were not. It is also worth noting that $\pm 3\sigma$ normalization had the greatest effect on the texture properties, so few features correlate with those estimated for the reference texture. In any case, it should be remembered that for a given feature all the analyzed ROIs need to be normalized in the same way, with a constant number of intensity levels.

Lack of normalization can render features useless for classification problems. A very good example is provided by GLCM-based features. An intermediate step for these features is the calculation of a gray level co-occurrence matrix, which is a two-dimensional (or second order) histogram. For the 16bit images utilized in this study, without reducing the intensity levels, such a histogram will be a matrix of 65535×65535 elements. This means over 4 billion elements. Assuming 32-bit integers as matrix elements, such a structure will occupy 16GB of memory. On the other hand, it will be a sparse matrix since there are $61 \times 61 = 3721$ pixels in each of the considered ROI for Brodatz textures. A simple reduction of the number of intensity levels without normalization will lead to a situation in which the pixel values will be represented by an extremely limited number of intensities. In the case of modified Brodatz texture-based images, the intensity values for each image fall in a range of between 950 and 1550. Simple binning into 256 intensity levels will result in a maximum of 3 consecutive (2, 3 and 4) intensities. Similar behavior may be observed for MRI images, where the intensities in the kidney region fall into a range of between 0 and 700 and the bone, muscle and fat

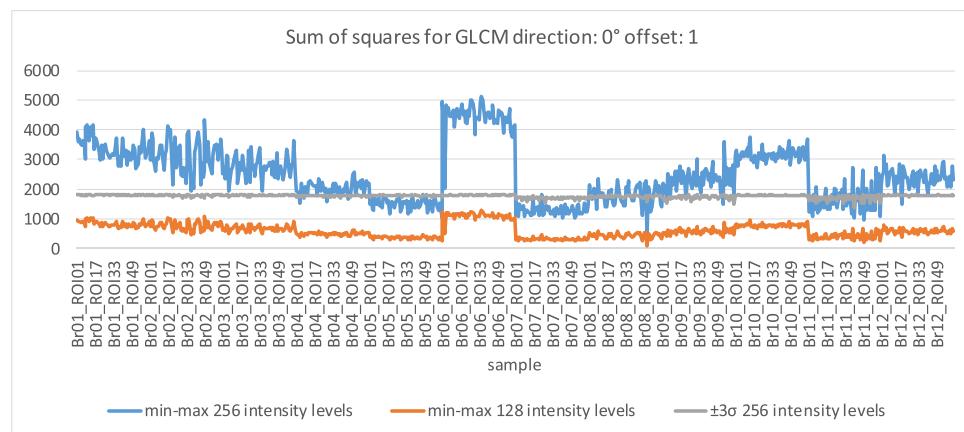


Fig. 21. Example feature values plotted for ROI/Image combinations in Brodatz images. Min-max normalization with 256 intensity levels (reference), min-max normalization with 128 intensity levels (Pearson correlation coefficient = 1.00), $\pm 3\sigma$ normalization with 256 intensity levels (Pearson correlation coefficient = 0.55).

Table 1

Percentage of features with a Pearson correlation coefficient ≥ 0.9 for different combinations of normalization technique and numbers of intensity levels with min-max normalization and 256 intensity levels. This table was created based on results from Brodatz images.

Normalization	min max				1 %-99 %				$\pm 3\sigma$									
	# gray levels	128	64	32	16	8	256	128	64	32	16	8	256	128	64	32	16	8
Gabor	100	100	100	100	100	96	96	96	96	96	96	88	71	75	75	75	71	67
GLCM	70	65	65	66	61	99	91	60	58	58	58	40	15	11	11	11	11	11
gradient	80	80	80	80	60	20	40	40	40	40	40	0	0	0	0	0	0	0
GRLM	17	0	0	0	0	100	50	0	0	0	0	96	13	0	0	0	0	0
WDT	100	100	100	100	100	92	92	92	92	92	92	8	8	8	8	8	8	8
HOG	100	100	100	100	100	100	100	100	100	100	100	75	83	75	83	67	50	

fall into a range of between 0 and 550. For tested CT images the intensity range in considered ROIs fall into a range of between 850 and 1350. The application of a normalization technique will result in better representation of the texture in the new intensity range.

There are several papers concerning the utilization of different normalization methods in texture analysis. In (Dzierżak et al., 2019) different normalization methods were applied for analysis of Spinal CT images. It was demonstrated in this paper that image normalization significantly worsened the accuracy of the automatic diagnosis of osteoporosis based on CT images of the spine. In (Loizou et al., 2009) five different normalization techniques were tested during the analysis of brain MRI images. It was shown in this work that application of min-max normalization provides more stable feature values for brain lesion analysis. In (Collewet et al., 2004) normalization techniques were tested for analysis of cheese MRI images. It was demonstrated in this work, that application of image normalization improves soft cheese texture classification accuracy.

However, to the best knowledge of the authors this is the only publication in which different normalization techniques were tested in comprehensive way for different numbers of intensity levels on images with various distortions obtained by means of various imaging techniques.

The research, of course, has limitations. Only three classes of biomedical images, obtained using two imaging methods (CT, MRI), were used. It cannot be assumed that these images were disturbed by noise from perfect Rice and Gauss distributions, as was the case of the tests performed on the textures from the Brodatz album. Regardless of the noise, however, these images had artifacts related to reduced acquisition time (MRI) or different reconstructive nuclei (CT). Therefore, the results of the analyses obtained for these images did not always coincide with those obtained for natural textures. Nevertheless, the results allow several conclusions to be drawn that should be true for a broader class of images.

A few general conclusions can be drawn from the Brodatz experiment:

- 1) For GRLM-based features, there is a clear rule that low numbers of intensity levels should be used. Improvement of the classification results for small numbers of intensity levels is due to the construction principles of the run length matrix. For small numbers of intensity levels, GRLM is less sensitive to noise, because strips of pixels with the same intensity are not broken.
- 2) For high noise distortions, both Gaussian and Rician $\pm 3\sigma$ normalization gives lower classification error than min-max and 1 %-99 % normalization. (sentence removed) This is caused by the fact that the addition of noise modifies (stretches) the histogram in such a way that it becomes similar to a Gaussian distribution, for which $\pm 3\sigma$ normalization is preferred.
- 3) For Gabor transform-based features, the normalization technique of choice is $\pm 3\sigma$. This normalization gives the minimum classification error for all tested numbers of intensity levels, regardless of the distortion type.
- 4) For wavelet features, $\pm 3\sigma$ normalization gives either the lowest or very close to lowest discrimination error for most distortion combinations and all tested numbers of intensity levels. The only exceptions occur for low levels of Gaussian noise and non-distorted images, in which case min-max normalization gives the best results.
- 5) For HOG-based features, 1 %-99 % normalization is preferred for low noise levels as well as for the combination of moderate noise and background nonuniformity. For other distortion combinations, $\pm 3\sigma$ normalization is better.
- 6) For gradient-based features, 1 %-99 % normalization gives either the lowest or very close to the lowest discrimination error for most distortion combinations. The only two exceptions are low and moderate Rician noise combined with background non-uniformity and low Gaussian noise with background non-uniformity.

Unfortunately, these findings do not hold for the kidney experiment. This is due to the fact that for images where the contrast agent is flushed to the pelvis, the histograms differ significantly from a Gaussian distribution (Fig. 11c). Such histograms have a relatively long tail of large values for intensities from the bright but relatively small pelvis area. Both $\pm 3\sigma$ and 1%–99% normalizations can cause clipping of this tail, resulting in the pelvis blending into the kidney cortex area. This can lead to the degradation of classification results for certain features. Such a situation is especially visible for Gabor and Wavelet transform-based features.

The GLCM-based features require additional comment. It is difficult to identify any clear trend in terms of discrimination performance for different normalizations or numbers of intensity levels. On the other hand, for all tested image sets the lowest discrimination error was obtained for GLCM-based features. The maximal discrimination error was 9.3% and the minimal discrimination error 0.8% for all combinations of distortion tested in the Brodatz experiment. In the kidney experiments, the minimal and maximal errors were 0% and 1.5%, respectively. It seems that such features clearly describe the properties of Brodatz and kidney textures and that their values are sufficiently distinctive to provide good classification of the analyzed classes. In such cases, the normalization scheme does not significantly influence classification accuracy, even if not all of the features maintain their values after normalization (see Table 1). This means that the proper selection of parameters which uniquely represent texture properties is crucial for further analysis.

For the MRI shoulder and CT heart datasets, it can be noticed that $\pm 3\sigma$ normalization works less well than in the case of kidney images. This can be explained by the fact that none of the images in these datasets follow the Gaussian distribution (see Fig. 12 and Fig. 13). Therefore, $\pm 3\sigma$ normalization results in missing gray level distribution, usually due to overextension of the intensity range to fit it to the normal distribution. Thus, other normalization schemes are a better choice for such medical images, such as 1%–99% or min-max, at least as far as texture classification is considered.

Another observation related to the medical datasets is that, contrary to the Brodatz images, in the case of GLCM and GRLM-based features a reduction in classification error was observed with increasing numbers of intensity levels. In this case, the noise that affected both the MR and CT images was not additive Gaussian (as it was in the experiments with Brodatz data), so no averaging effect occurred. When an image is affected by additive noise, the reduction in the gray level number corresponds to averaging the whole image data, including noise. When the noise mean is equal to zero, as it was in the experiment performed on Brodatz images, the averaging of additive noise leads to its reduction (in terms both of the standard deviation and the maximum amplitude value). In the medical images analyzed here, the noise was usually not additive (Rician in the case of MR data, or kernel dependent in the reconstructed CT images). In such cases, reducing the number of intensity levels does not reduce image noise. Moreover, the information reduction caused by this process deteriorates classification accuracy in some cases (see e.g. GLCM features for MR GRAPPA 1 and CT data, all kernels).

Finally, the best texture classification ability, in the cases of both the Brodatz texture and biomedical images analyzed in this study, was for GLCM-based features. The number of such features is much larger in comparison to other parameter groups, as stated in subsection Test Methodology. Of course, some may be correlated (especially the same features calculated for different inter-pixel distances). However, it is still easier to find a 3-element feature subset from larger input space that will be optimal for classification than to consider a parameter space with much lower dimensionality. On the other hand, our analysis of the literature shows that the statistical parameters estimated for the co-occurrence matrix are

the most commonly used texture parameters. Perhaps this is due to the fact that those features were among the first to be identified (they were developed in the 1970s (Haralick, 1979; Haralick et al., 1973)). Nevertheless, they have demonstrated their usefulness repeatedly to describe a wide class of textures found in biomedical and natural images. Of course, other methods of texture analysis should not be rejected a priori (unless there is evidence of its limited usefulness for a given class of images). Biomedical textures are characterized by very high complexity and variability, so it is worth investigating a whole series of features in their analysis.

To summarize, for Gabor transform, gradient, HOG and wavelet-based features the number of intensity levels has a minimal influence on classification error. On other hand, this error depends on the normalization method. For GLCM and GRLM-based features, both the normalization method and the number of intensity levels influence the classification error.

In the case of MRI and CT images, the lowest classification error can usually be obtained with large numbers of intensity levels (mainly 256 in our case). The normalization method of choice for such modalities is 1%–99%, since it gives the most stable results (usually the best or close enough to the best). The other two normalization methods should be considered if 1%–99% normalization does not provide sufficiently low classification error.

In the case of images affected only by additive noise, $\pm 3\sigma$ normalization with low numbers of intensity levels may be considered for GLCM and GRLM-based features.

Authors contribution

Marcin Kociołek is the main author of the presented research and vast amount of the text. He has chosen the methodology and performs all the data analysis.

Michał Strzelecki oversees whole project. He was an initiator of the study. He has great contribution in the literature revive and discussion part of the text.

Rafał Obuchowicz provided the thorax CT and Shoulder t2 weighted MRI images. He consult all medical aspects of the paper. He was also contributed in the text describing CT and MRI data.

Declaration of Competing Interest

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Funding source

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Human subjects study statement

All human-related imaging data used in this study were received from external institutions. All processed image data was anonymized before it was handed to us. The institutions which acquired the data followed the guidelines of the Declaration of Helsinki and the Good Clinical Practice Declaration Statement. Special care was taken regarding personal data safety, with all images

anonymized before processing. Written acceptance to conduct the study was obtained from the appropriate Ethics Committees.

References

- Aja-Fernández, S., Vegas-Sánchez-Ferrero, G., 2016. *Statistical Analysis of Noise in MRI: Modeling, Filtering and Estimation*, 1st ed. Springer International Publishing, Cham.
- Aja-Fernández, S., Tristán-Vega, A., Alberola-López, C., 2009. Noise estimation in single-and multiple-coil magnetic resonance data based on statistical models. *Magn. Reson. Imaging* 27, 1397–1409, <http://dx.doi.org/10.1016/j.mri.2009.05.025>.
- Bielecka, M., Piórkowski, A., 2015. Automatized fuzzy evaluation of CT scan heart slices for creating 3D/4D heart model. *Appl. Soft Comput.* J. 30, 179–189, <http://dx.doi.org/10.1016/j.asoc.2015.01.023>.
- Boas, F.E., Fleischmann, D., 2012. *CT artifacts: causes and reduction techniques. Imaging Med.* 4, 229–240.
- Brodatz, P., 1966. *Textures: A Photographic Album for Artists and Designers*. Dover Publications, New York, NY.
- Chervyakov, N.I., Lyakhov, P.A., Orazaev, A.R., Valueva, M.V., 2017. Efficiency analysis of the image impulse noise cleaning using median filters with weighted central element. In: 2017 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), IEEE, pp. 141–146, <http://dx.doi.org/10.1109/SIBIRCON.2017.8109856>.
- Chrzanowski, L., Drozdz, J., Strzelecki, M., Krzeminska-Pakula, M., Jedrzejewski, K., Kasprzak, J., 2008. Application of neural networks for the analysis of histological and ultrasonic aortic wall appearance—an invitro tissue characterization study. *Ultrasound Med. Biol.* 34, 103–113.
- Collewet, G., Strzelecki, M., Mariette, F., 2004. Influence of MRI acquisition protocols and image intensity normalization methods on texture classification. *Magn. Reson. Imaging* 22, 81–91, <http://dx.doi.org/10.1016/j.mri.2003.09.001>.
- Dzierżak, R., Omiotek, Z., Tkacz, E., Kępa, A., 2019. The influence of the normalisation of spinal CT images on the significance of textural features in the identification of defects in the spongy tissue structure. In: Advances in Intelligent Systems and Computing. Springer Verlag, pp. 55–66, http://dx.doi.org/10.1007/978-3-030-15472-1_7.
- Eikefjord, E., Andersen, E., Hodneland, E., Hanson, E.A., Sourbron, S., Svarstad, E., Lundervold, A., Rørvik, J.T., 2017. Dynamic contrast-enhanced MRI measurement of renal function in healthy participants. *Acta radiol.* 58, 748–757, <http://dx.doi.org/10.1177/0284185116666417>.
- Grant, K., Raupach, R., 2012. SAFIRE: Sinogram Affirmed Iterative Reconstruction White Paper.
- Gupta, S., Chauhan, R.C., Sexana, S.C., 2004. Wavelet-based statistical approach for speckle reduction in medical ultrasound images. *Med. Biol. Eng. Comput.*, 189–192, <http://dx.doi.org/10.1007/BF02344630>.
- Haralick, R.M., 1979. Statistical and structural approaches to texture. *Proc. IEEE* 67, 786–804, <http://dx.doi.org/10.1109/PROC.1979.11328>.
- Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* 3, 610–621, <http://dx.doi.org/10.1109/TSMC.1973.4309314>.
- Kociołek, M., Strzelecki, M., Szymajda, S., 2018. On the influence of the image normalization scheme on texture classification accuracy. In: 2018 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), IEEE, pp. 152–157, <http://dx.doi.org/10.23919/SPA.2018.8563397>.
- Kociołek, M., Strzelecki, M., Klepaczko, A., 2019. Functional kidney analysis based on textured DCE-MRI images. In: Advances in Intelligent Systems and Computing. Springer Verlag, pp. 38–49, http://dx.doi.org/10.1007/978-3-030-23762-4_4.
- Kropidłowski, K., Kociołek, M., Strzelecki, M., Czubiński, D., 2014. *Model Based Approach for Melanoma Segmentation, Lecture Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*.
- Larroza, A., Materka, A., López-Lereu, M.P., Monmeneu, J.V., Bodí, V., Moratal, D., 2017. Differentiation between acute and chronic myocardial infarction by means of texture analysis of late gadolinium enhancement and cine cardiac magnetic resonance imaging. *Eur. J. Radiol.* 92, 78–83.
- Liu, S., Mlodzianoski, M.J., Hu, Z., Ren, Y., McElmurry, K., Suter, D.M., Huang, F., 2017. sCMOS noise-correction algorithm for microscopy images. *Nat. Methods* 14, 760–761, <http://dx.doi.org/10.1038/nmeth.4379>.
- Loizou, C.P., Pantziaris, M., Seimenis, I., Pattichis, C.S., 2009. Brain MR image normalization in texture analysis of multiple sclerosis. *Final Progr. Abstr. B. - 9th Int. Conf. Inf. Technol. Appl. Biomed. ITAB*, 1–5, <http://dx.doi.org/10.1109/ITAB.2009.5394331>.
- Lysaker, M., Lundervold, A., Tai, Xue-Cheng, 2003. Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time. *IEEE Trans. Image Process.* 12, 1579–1590, <http://dx.doi.org/10.1109/TIP.2003.819229>.
- Materka, A., Strzelecki, M., 2013. On the importance of MRI nonuniformity correction for texture analysis, in: *signal Processing: algorithms, Architectures, Arrangements. Appl. Evol. Comput.* 2013, 118–123.
- Materka, A., Strzelecki Michał and Lerski, R., Schad, L., 2000. Feature evaluation of texture test objects for magnetic resonance imaging. In: *Texture Analysis in Machine Vision*. World Scientific, pp. 197–206.
- Michailovich, O.V., Tannenbaum, A., 2006. Despeckling of medical ultrasound images. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 53, 64–78, <http://dx.doi.org/10.1109/TUFC.2006.1588392>.
- Obuchowicz, R., Piórkowski, A., Urbanik, A., Strzelecki, M., 2019. Influence of acquisition time on MR image quality estimated with nonparametric measures based on texture features. *Biomed. Res. Int.* 2019, 1–10, <http://dx.doi.org/10.1155/2019/3706581>.
- Ouhabi, A., 2013. A review of wavelet denoising in medical imaging. In: 2013 8th International Workshop on Systems, Signal Processing and Their Applications (WoSSPA), IEEE, pp. 19–26, <http://dx.doi.org/10.1109/WoSSPA.2013.6602230>.
- Pieciak, T., Aja-Fernandez, S., Vegas-Sánchez-Ferrero, G., 2017. Non-stationary rician noise estimation in parallel MRI using a single image: a variance-stabilizing approach. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2015–2029, <http://dx.doi.org/10.1109/TPAMI.2016.2625789>.
- Priya, D.K., Sam, B.B., Lavanya, S., Sajin, A.P., 2017. A survey on medical image denoising using optimisation technique and classification. In: 2017 International Conference on Information Communication and Embedded Systems (ICICES), IEEE, pp. 1–6, <http://dx.doi.org/10.1109/ICICES.2017.8070729>.
- Ridgway, G., n.d. Rice/Rician distribution - File Exchange - MATLAB Central [WWW Document]. URL <https://www.mathworks.com/matlabcentral/fileexchange/14237-rice-rician-distribution?focused=5109004&tab=example> (accessed 3.25.19).
- SIPI Image Database - Textures [WWW Document], n.d. URL <http://sipi.usc.edu/database/database.php?volume=textures> (accessed 3.21.19).
- Somkuwar, A., Bhargava, S., 2013. Noise reduction techniques in medical imaging data-a review. In: 2nd International Conference on Mechanical, Electronics and Mechatronics Engineering (ICMEME'2013), June 17–18, 2013 London (UK). London, pp. 115–119.
- Strzelecki, M., Kociołek, M., Materka, A., 2018. On the influence of image features wordlength reduction on texture classification. International Conference on Information Technologies in Biomedicine, 15–26.
- Styner, M., Van Leemput, K., Landini, L., Positano, V., Santarelli, M.F., 2005. Retrospective evaluation and correction of intensity inhomogeneities. In: *Advanced Image Processing in Magnetic Resonance Imaging*. Taylor & Francis CRC Press, pp. 145–168.
- Szczypliński, P.M., [WWW Document] URL 2018. Qmazda Manual. <http://www.eletel.p.lodz.pl/pms/Programy/qmazda.pdf>.
- Szczypliński, P.M., Klepaczko, A., Kociołek, M., 2017. QMaZda - software tools for image analysis and pattern recognition. In: Signal Processing - Algorithms, Architectures, Arrangements, and Applications Conference Proceedings, SPA, <http://dx.doi.org/10.23919/SPA.2017.8166867>.
- Thevenot, J., Hirvasniemi, J., Finnillä, M., Pulkkinen, P., Kuhn, V., Link, T., Eckstein, F., Jämsä, T., Saarakkala, S., 2013. Trabecular homogeneity index derived from plain radiograph to evaluate bone quality. *J. Bone Miner. Res.* 28, 2584–2591.
- Völgyes, D., Pedersen, M., Stray-Pedersen, A., Waaler, D., Martinsen, A.C.T., 2017. How different iterative and filtered back projection kernels affect computed tomography numbers and low contrast detectability. *J. Comput. Assist. Tomogr.* 41, 75–81, <http://dx.doi.org/10.1097/RCT.0000000000000491>.
- Yilmaz, E., Kayikcioğlu, T., Kayipmaz, S., 2017. Noise removal of CBCT images using an adaptive anisotropic diffusion filter. 2017 40th International Conference on Telecommunications and Signal Processing (TSP), 650–653, <http://dx.doi.org/10.1109/TSP.2017.8076067>.