



A segmentation-based method improving the performance of N4 bias field correction on T2weighted MR imaging data of the prostate

Aikaterini Dovrou^{a,*}, Katerina Nikiforaki^a, Dimitris Zaridis^{b,c,d}, Georgios C. Manikis^a,
Eugenia Mylona^{b,c}, Nikolaos Tachos^{b,c}, Manolis Tsiknakis^{a,e}, Dimitrios I. Fotiadis^{b,c},
Kostas Marias^{a,e}

^a Computational BioMedicine Laboratory (CBML), Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Heraklion, Greece

^b Unit of Medical Technology and Intelligent Information Systems, Department of Materials Science and Engineering, University of Ioannina, Ioannina, Greece

^c Biomedical Research Institute, Foundation for Research and Technology – Hellas (FORTH), Ioannina, Greece

^d Biomedical Engineering Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

^e Department of Electrical & Computer Engineering, Hellenic Mediterranean University, Heraklion, Greece

ARTICLE INFO

Keywords:

Prostate imaging
N4 bias field correction
Periprostatic fat segmentation
Full width at half maximum

ABSTRACT

Magnetic Resonance (MR) images suffer from spatial inhomogeneity, known as bias field corruption. The N4ITK filter is a state-of-the-art method used for correcting the bias field to optimize MR-based quantification. In this study, a novel approach is presented to quantitatively evaluate the performance of N4 bias field correction for pelvic prostate imaging. An exploratory analysis, regarding the different values of convergence threshold, shrink factor, fitting level, number of iterations and use of mask, is performed to quantify the performance of N4 filter in pelvic MR images. The performance of a total of 240 different N4 configurations is examined using the Full Width at Half Maximum (FWHM) of the segmented periprostatic fat distribution as evaluation metric. Phantom T2weighted images were used to assess the performance of N4 for a uniform test tissue mimicking material, excluding factors such as patient related susceptibility and anatomy heterogeneity. Moreover, 89 and 204 T2weighted patient images from two public datasets acquired by scanners with a combined surface and endorectal coil at 1.5 T and a surface coil at 3 T, respectively, were utilized and corrected with a variable set of N4 parameters. Furthermore, two external public datasets were used to validate the performance of the N4 filter in T2weighted patient images acquired by various scanning conditions with different magnetic field strengths and coils. The results show that the set of N4 parameters, converging to optimal representations of fat in the image, were: convergence threshold 0.001, shrink factor 2, fitting level 6, number of iterations 100 and the use of default mask for prostate images acquired by a combined surface and endorectal coil at both 1.5 T and 3 T. The corresponding optimal N4 configuration for MR prostate images acquired by a surface coil at 1.5 T or 3 T was: convergence threshold 0.001, shrink factor 2, fitting level 5, number of iterations 25 and the use of default mask. Hence, periprostatic fat segmentation can be used to define the optimal settings for achieving T2weighted prostate images free from bias field corruption to provide robust input for further analysis.

1. Introduction

Prostate cancer (PCa) is the most frequently diagnosed cancer in men, affecting 1.4 million men per year worldwide [1]. Imaging plays an important role for non-invasive patient classification in order to avoid overtreatment of indolent cancers as well as undertreatment of more aggressive cases requiring prompt intervention [2]. The role of

radiomics studies as a cost-effective, supportive tool has emerged through a large number of studies where high throughput extraction of quantitative features lead to machine learning models able to enhance confidence on medical decision [3]. However, a series of preparatory steps are required to minimize measurement errors such as variability under different conditions, imaging protocols, vendors or between measurements on the same subjects aiming at improved reproducibility

* Corresponding author at: Computational BioMedicine Laboratory (CBML), Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), 100 Nikolaou Plastira Str., Vassilika Vouton, Heraklion, Crete GR 70013, Greece.

E-mail address: dovrou@ics.forth.gr (A. Dovrou).

<https://doi.org/10.1016/j.mri.2023.03.012>

Received 19 October 2022; Received in revised form 15 March 2023; Accepted 21 March 2023

Available online 31 March 2023

0730-725X/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of quantitative image post-processing. Inhomogeneity correction, noise filtering and intensity normalization are the most frequently used pre-processing steps in this direction and can be deployed either as discrete actions or in the frame of a multi-step workflow for image preparation.

Concerning MR images, robust annotation, segmentation, texture analysis or classification assume an image of high diagnostic value, free from artifacts. However, MR images are very frequently affected by a low frequency variation in the acquired signal known as bias field corruption. This non-uniformity is the result of a number of contributing factors, such as poor radiofrequency (RF) coil design, gradient eddy currents, local variations in flip angle and inhomogeneous excitations caused by interactions between radio-frequency waves and electromagnetic properties of the tissues. Scanners with higher main magnetic field strength are more severely affected, as well as acquisitions performed with surface coils.

Bias field corruption in MR images appears randomly at different locations among patients and at different acquisitions, even subsequent acquisitions of the same individuals. Moreover, each MR pulse sequence is affected to a different degree from this corruption and thus in the frame of a multicentric study with non-identical vendor-specific pulse sequence designs, images are sensitive to a variable degree. Although B1 mapping can be performed based on the reciprocity principle [4] during scanning, it costs valuable scan time and, therefore, retrospective approaches are commonly used to correct the bias field. Bias correction methods are broadly categorized into prospective and retrospective methods [5]. The former eliminates the bias field caused by the hardware devices by calibrating and improving the acquisition process. The retrospective approaches reduce the bias field arising from the properties of the object in the scanner and are more general methods in their concept.

The N4 bias field correction method [6] is a popular retrospective histogram-based method and has been very widely used for addressing bias field corruption as a preprocessing step in recent bibliography in classification [7], segmentation [8–12] and radiomic studies [13,14]. The N3 filter [15], which is the predecessor of the N4, has been reported to be the most frequently used for comparative studies [16] and outperforms other methods for addressing bias field corruption [17]. While the N4 generally performs well, an undesired effect (stronger than necessary correction) of a spot-like correction can result from a non-optimal N4 configuration, i.e. when converging to local either than global minima by setting a dense spline mesh combined with a high number of iterations [18]. Such pitfalls can be mitigated by appropriately selecting the parameters to be tailored to the needs of the specific problem, considering the main magnetic field strength, the number of coil segments and the general characteristics of the scanned object.

Since it is not easy to establish the ground truth, most published works focus on simulated data where the bias field is introduced by a known function and is corrected [17,19]. Phantoms have also been used to provide an indication of the minimum expected variability in measurements induced by scanning conditions. However, they lack the ability to simulate anatomy related susceptibility among subjects and tissue heterogeneity. Many studies assess the reproducibility of radiomic features to evaluate the performance of the bias field correction algorithm, as well as of other preprocessing techniques [20–23]. Additionally, Martin et al. [21] applied the N4ITK algorithm on breast phantoms and evaluated the values of specific N4 parameters. They identified the number of 50 iterations, fitting level 5 and the use of a full mask as optimal configuration for the reduction of the intensity inhomogeneities by assessing the segmentation results and measuring the coefficient of variation in the mean intensity of specific regions. The impact of the mask's shape and the spline distance on the performance of N3 was assessed on brain MR images acquired on 3 T scanners in a paper published prior to the N4 introduction by Boyes et al. [24]. In this work, the masks that enclose more precisely the brain tissue and smaller values of spline distances resulted in better performance of the N3 filter by measuring the variation of the normalized white matter intensity and

the variation of the normalized image difference. The effectiveness of using smaller spline smoothing distances in N3 filter in 3 T brain MR images was also confirmed by Zheng et al. [25]. Some published works on patient data focus on brain tissue and use the segmentation performance as a metric to evaluate the performance of the bias field correction algorithm [25–28].

However, to the best of our knowledge, the bias field correction on the pelvic area has not been widely investigated. The main challenges of this area, compared to the brain, are the increased anatomic heterogeneity among patients, the larger variability of MR sequences and the larger field of view. Furthermore, the impact of the parameters' values of the N4ITK filter on the bias field correction has been overlooked, as the majority of the studies use the default values of the filter. The present study aims to quantitatively and automatically evaluate the performance of the N4ITK filter for bias field correction in prostate images. Furthermore, the effect of the different values of each parameter on the performance of the N4 filter is investigated to identify specific optimal values for each parameter and thus improve the bias field correction in prostate images. The innovation of the current study is the rigorous exploration analysis of the performance of the N4ITK filter using various configurations. A novel automatic evaluation metric based on the periprostatic fat distribution is proposed to quantitatively assess the performance and identify optimal configurations of the N4ITK filter for the bias field correction of MR prostate images in different scanning conditions.

To this end, the FWHM of the periprostatic fat distribution was computed to assess the impact of the N4 filter on MR prostate images from two different magnetic field strengths (1.5 Tesla (1.5 T) and 3 Tesla (3 T)) and different receive coil configurations (surface coil and combination of surface and endorectal coil). As a preparatory step, the analysis on a homogeneous corn oil phantom was performed to define the minimum expected FWHM value and the minimum requirements on the parameters' values of the N4 filter. The fat tissue has consistent magnetic properties across individuals and also resides in the vicinity of the prostatic tissue. Small values of FWHM indicate narrow fat distribution and thus better bias field correction. Each parameter of the N4 filter was handled individually, while keeping the rest of the parameters unchanged, to have an overview of the effect of each parameter.

2. Methods

2.1. N4 filter

The N4 filter is available in python by the open-source image analysis toolkit `simpleITK` as an improvement of the N3 bias correction method. The N4 filter provides an improved B-spline fitting routine allowing for the use of multiple resolution levels and an optimized iterative process updating constantly the residual bias field. The N4 filter is a histogram-based method for intensity inhomogeneity correction in images and is available as a python module, called `N4ITK` [6]. The algorithm of the N4 bias filter is an iterative process between deconvolving the image histogram by a Gaussian, estimating the “corrected” intensities and spatially smoothing the resulted bias field estimation using the B-spline model. This process is repeated until the coefficient of variation (CV) in the ratio between subsequent bias field estimations drops below the convergence threshold or the maximum number of iterations is reached. Thus, the method seeks to estimate the smooth, slowly spatially varying, multiplicative field that sharpens the peaks of the intensity histogram by using an improved B-spline fitting routine.

The N4 bias field correction filter requires the selection of the values of its parameters, such as the convergence threshold, the shrink factor, the fitting level, the number of iterations and the use of mask. The convergence threshold is the stopping criterion for the iterative bias field estimation procedure. The shrink factor defines how much the original image will be downsampled before estimating the inhomogeneity field, leading to lower image complexity. The number of fitting levels defines

how many levels will be used to determine the resolution of the B-spline grid. At each fitting level, the previous mesh grid resolution is doubled and thus the spline distance gets smaller. The number of iterations defines the maximum number of iterations at each level of resolution. Regarding the use of a mask, if a binary mask is provided, the algorithm uses the voxels of the image that correspond to the voxels of the mask to estimate the bias field. If a mask is not provided, the algorithm generates a mask or uses the non-zero voxels of the image. More specifically, some implementations of the N4 filter use the simple Otsu thresholding in order to generate a mask. In the ITK implementation, the input image is log transformed and the values of the original image that are <1 are zeroed out in the log domain. Thus, these values are not used by the algorithm for the bias field estimation as only the non-zero voxels of the image are used. The computational complexity of the algorithm and the execution time are decreased by using larger values of convergence threshold and shrink factor and lower values of fitting level and number of iterations. For instance, the runtime of the N4 filter with configuration “mask False, threshold 0.001, shrink factor 2, fitting level 4, iterations 50” is approximately 25 s using a computer with an Intel(R) Core (TM) i7–9700 8-core 3.00GHz CPU and 16GB RAM, when applied to one MR prostate image with size of 16 MB.

2.2. Phantom

The corn oil phantom was used to assess the impact of the different configurations of the N4 filter on the bias field correction in homogeneous fat mimicking material. The corn oil phantom data was acquired on a SonataVision Siemens scanner using a 1.5 T magnet and 3 different Time to Echo (TE) values, which are 60, 80 and 120 ms, resulting in 3 scanning settings. More precisely, a commercially available 12-channel head matrix coil (Siemens Medical Solutions, Erlangen, Germany) was used for imaging the oil phantom. The lower and upper parts of the head matrix coil were symmetrical and housed six partially overlapping RF receiving elements each. This coil had an inner vertical diameter of 26 cm and an inner horizontal diameter of 25 cm. The oil sample was placed in a 16×14cm container (quasi-rectangular bottom part) with 18 cm height, excluding the bottleneck part. The sample contained edible corn oil from the same brand that has T2 relaxometry constants approximating adipose tissue, as shown in earlier experiments of our team [29,30]. The bottleneck part faced the feet direction to ensure maximum coil loading. Prior to imaging, the cylinder was topped up with material from an identical sample in order to restrict susceptibility effects from air. Imaging slices were centered in the middle part of the container, for the same reason. The container was surrounded by a thin layer of foam material and was further immobilized by sand cushions that covered the empty space between the coil and the phantom.

The N4ITK filter was applied to the 3 produced phantom T2weighted (T2W) series of different TEs frequently applied in clinical practice. More precisely, the examined values were: i) convergence threshold 0.01, 0.001 and 0.0001; ii) fitting level 3, 4, 5 and 6; and iii) number of iterations 5, 10, 25, 50 and 100. All possible combinations of these parameters values were applied, resulting in 60 different settings of the N4 filter. The initial B-spline mesh element size is 1x1x1 with a spline distance of 200 mm at the first level, resulting in a 32x32x32 mesh element with a spline distance of 6.25 mm at the last sixth level. The effect of the shrink factor parameter could not be examined as the corn oil phantom consists of only 2 slices. The intensity histogram of the central region of each phantom and the corresponding FWHM were calculated. The central region was manually positioned by using a cubic box that covers the largest area of the phantom and excludes the area with the low intensity values that exists close to the phantom's wall.

For each corn oil phantom, the filter's setting that resulted in the minimum FWHM was identified and considered as optimal setting. The minimum FWHM corresponds to the narrowest distribution, demonstrating more homogeneous distribution of the intensity values. This minimum FWHM value was then compared with the FWHM of the

original unfiltered image to quantify the improvement in the intensity values after the N4 filter.

2.3. Datasets

Two public datasets of T2W prostate images were used for our analysis. The first dataset is the PROSTATE-DIAGNOSIS which consists of 89 prostate cancer T2W MR images acquired on a 1.5 T Philips Achieva scanner using combined surface and endorectal coil [31,32] (the data can be downloaded from [33]).

The second dataset used was the PROSTATEx which consists of 204 prostate T2W MR images that were acquired on two different types of Siemens 3 T MR scanners, the MAGNETOM Trio and Skyra (the data can be downloaded from [34]). The T2W images were acquired using a turbo spin echo sequence without an endorectal coil and had a resolution around 0.5 mm in plane and a slice thickness 3.6 mm [32,35,36].

Two external public datasets were used to validate the findings. The Prostate-MRI dataset consists of 26 prostate T2W MR images acquired on a 3 T Philips Achieva scanner using combined phased-array surface and endorectal coil [32,37] (the data can be downloaded from [38]). This dataset was used to validate the performance of the derived optimal configuration in prostate MR images acquired by a combined surface and endorectal coil.

The second dataset used for validation was the PI-CAI dataset [39] (the data can be downloaded from [40]). This dataset consists of 1500 prostate T2W MR images acquired on Siemens Healthineers or Philips Medical Systems-based MR scanners at 1.5 T or 3 T using surface coils. More precisely, the Aera, Avanto and Espree scanners of Siemens Healthineers and the Achieva and Intera scanners of Philips were used at 1.5 T magnetic field strength. The Skyra, TrioTrim and Prisma scanners of Siemens Healthineers and the Ingenia scanner of Philips were used at 3 T magnetic field strength. This dataset was used to validate the performance of the derived optimal configuration of the N4 filter in prostate MR images acquired by a surface coil. Two subsets of T2W images were selected based on the two different magnetic field strengths, i.e., 1.5 T and 3 T, used for scanning in PI-CAI dataset, as the magnetic field strength affects the degree of the bias field corruption in images. In each subset, images of the same magnetic field strength were randomly selected in order to include heterogeneity and avoid biases. The whole dataset cannot be used due to the high computational complexity and the time-consuming process of the N4 filter. All the images of the PI-CAI dataset that were scanned with 1.5 T magnetic field strength were included in the first validation set, as there were only 59 such T2W prostate images. A subset of 80 random images that were scanned with 3 T magnetic field strength was selected for the second validation set. Hence, the effect of the magnetic field strength on the performance of the N4 filter was assessed using two sets of images acquired with the same receive coil configuration, i.e., surface coil, but different magnetic field strength, i.e., 1.5 T and 3 T, to account for the severity of the bias field corruption induced by the magnetic field strength.

2.4. Assessment of N4 settings in MR images

The effect of the N4 filter's parameters on the enhancement of the MR image is assessed in the heterogeneous MR prostate images. The analysis was performed independently on each dataset. The same N4 parameters that were applied to the phantoms were also used to the patients' images with the addition of the shrink factor parameter and the use of a mask. The shrink factor defines how much the original image will be down-sampled, reducing the size of the image along the three dimensions. The mask was derived automatically by the proposed deep learning model [41] that extracts a cubic box around the estimated position of the whole prostate gland. The deep learning model requires slice-wised min-max normalization of the images and resizing to 256 × 256. Hence, 240 different configurations of the N4 filter were applied to the patients' images by using all combinations of the following parameters: i) with

and without mask; ii) convergence threshold 0.01, 0.001 and 0.0001; iii) shrink factor 2 and 3; iv) fitting level 3, 4, 5 and 6; and v) number of iterations 5, 10, 25, 50 and 100.

The workflow of the proposed pipeline is presented in Fig. 1. The original and the filtered images were cropped to include only the periprostatic region by removing the heterogeneous prostate gland and the area distant from the prostate. The prostate gland is located at the center of the image; thus, each image was divided in 3 equal parts vertically and the middle part was selected for subsequent analysis to automatically restrict the working area in the prostate region based on human anatomy. The mask of the prostate was applied to the image for the heterogeneous prostate removal. The K-means clustering algorithm was implemented on the image intensity values of the masked image to identify the periprostatic fat distribution. The number of clusters (i.e., K) was set equal to 2 to discriminate the low intensity values from the high intensity values that correspond to the fat distribution, since after masking the fat and muscle tissue dominated in the image working space. The fat tissue corresponds to high intensity values in the T2W MR images. The FWHM was calculated on the frequency distribution of the cluster with the high intensity values for the original and the filtered images of each patient. The voxels with intensity values belonging to the cluster of the high intensity values were marked with red color and were considered as an approximation of the fat tissue. The minimum FWHM achieved among the filtered images with different settings corresponds to the optimal filter's setting for each patient and was compared with the FWHM value of the original unfiltered image. The fat distribution should become narrower after applying the N4 bias field correction as the tissue intensities become more homogeneous. The narrowest fat distribution, and subsequently the minimum FWHM, indicates the reduction of the inhomogeneities that account for fat distribution broadening. Furthermore, the relative difference between the FWHM of a specific setting and the minimum FWHM obtained from the optimal setting for each patient was also computed to identify the set of parameter values which are optimal for all the dataset.

3. Results

3.1. Phantom results

In Table 1, the results for the corn oil phantom with the 3 different TE values relevant for T2W imaging are presented. The optimal settings of the N4 filter are similar for all scanning settings of the phantom. The fitting level was 6 in the configurations that resulted in the minimum FWHM in all examined cases of the phantom. The convergence threshold is 0.001 in the optimal configurations of the phantoms with TE equal to 60 and 80, while it is 0.01 for the phantom with TE 120. The number of iterations that are used in these derived optimal configurations vary among the three phantoms. More specifically, the minimum FWHM achieved for the phantom with TE 120 is independent of the number of iterations, as it is achieved using either 5 or 10 or 25 or 50 or 100 iterations. However, the values of the FWHM achieved by configurations using a threshold 0.01 or 0.001, a fitting level 5 or 6 and independently

Table 1

Selection of the optimal filter settings for the corn oil phantoms based on the difference of FWHM before and after N4 filtering. For each corn oil phantom, the optimal settings, the FWHM of the distribution of the original image and the minimum FWHM of the distribution of the filtered image with the optimal filter configuration are depicted. In the last 2 columns, the absolute and the relative difference between the FWHM of the unfiltered image and the FWHM of the selected filtered images are presented. The (–) sign in the relative difference denotes the decrease in the FWHM value after applying the N4 bias filter.

Cornoil_TE	Optimal Filter Settings	FWHM original	FWHM filtered	Absolute Difference	Relative Difference
cornoil60	thres = 0.001_fit = 6_iters = 25	41.89	3.46	38.43	–91.74%
cornoil80	thres = 0.001_fit = 6_iters = 10	37.06	3.75	33.31	–89.88%
cornoil120	thres = 0.01_fit = 6_iters = 5, 10, 25, 50, 100	27.1	3.69	23.41	–86.38%

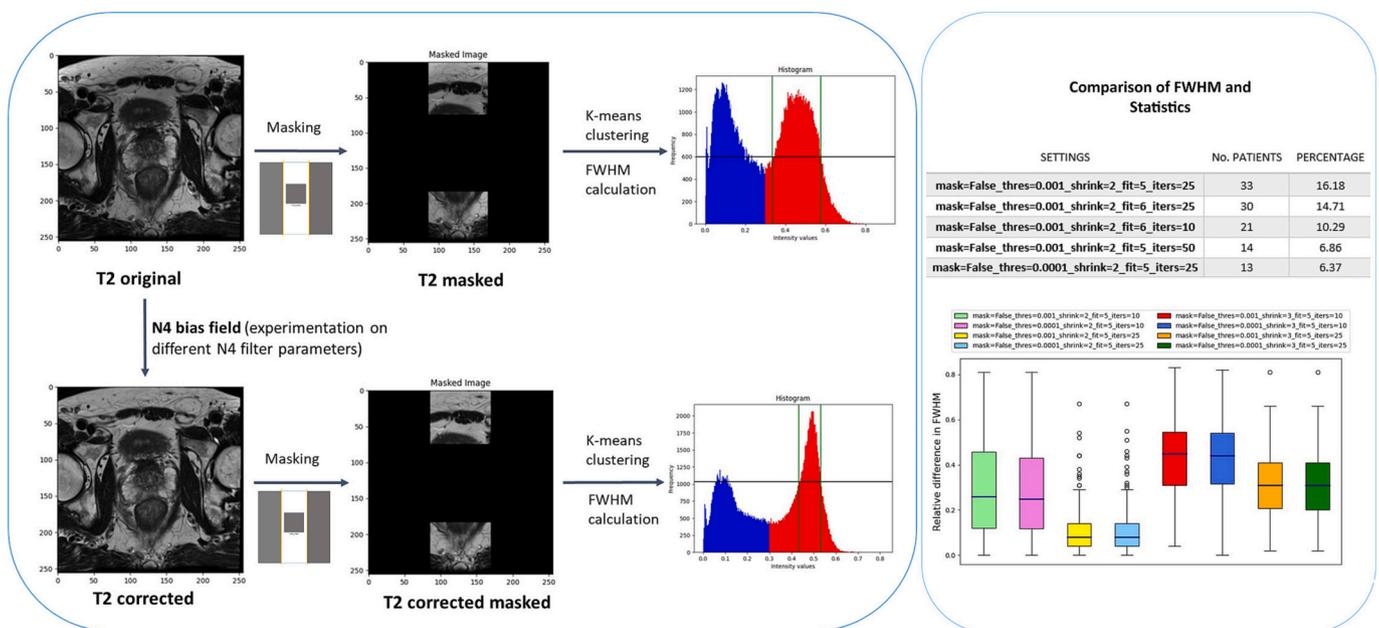


Fig. 1. Workflow of the proposed methodology for assessing the N4 parameters. The pipeline is performed to each patient and statistics are derived from the calculated FWHM values from all patients.

of the number of iterations, are the lowest and quite close to each other, as depicted in Fig. 2. The relative difference showed a decrease larger than 86% in the FWHM value, demonstrating the effectiveness of the N4 filter. In the Fig. 3A, the original unfiltered image and the corresponding N4 filtered image with the optimal setting of phantom with TE 60 are depicted, as well as the estimated bias field map. The homogeneous material of the phantom becomes brighter after applying the N4 filter. The bias field map shows the variations of the bias field across the phantom. The difference in the histograms of the image before and after N4 filtering with the optimal settings, is illustrated in Fig. 3B. The distribution of image intensities is significantly narrower in the N4 filtered than the original phantom. The width of the line between the two vertical green lines corresponds to the FWHM, which is marked with orange color.

3.2. Prostate results

The histogram of a prostate MR image before and after the N4 filtering is presented in Fig. 4B to visualize the effect of the N4 bias filter in the image intensity values. A narrower distribution of the high intensity values (i.e., red colored part of histogram) as well as a bigger value in the frequency of the peak are observed in the N4 filtered than the original image. Subsequently, the image voxels that have intensity

values in the fat distribution are presented with red color in Fig. 4C, where the fat is better identified in the N4 filtered than in the original image.

The pipeline was initially performed to the patients from the PROSTATE-DIAGNOSIS dataset, which contains MR images scanned at 1.5 T field with combined surface and endorectal coil. The combination of the use of threshold 0.001, shrink factor 2, fitting level 6, iterations 100 and without mask was identified as optimal for the 52% of the patients, which was the highest percentage (Supporting Table S1). However, this proportion is not adequate large and thus this setting could not be considered representative, before being examined further, for all the patients of this dataset.

Hence, the relative difference between the minimum FWHM of their optimal setting and the FWHM achieved with the setting “mask False, threshold 0.001, shrink factor 2, fitting level 6, iterations 100” was calculated for the rest 48% of the patients and presented in Fig. 5. We further examined the difference between the FWHM of the optimal setting of each patient and the FWHM achieved with the second most frequent setting as well as with other settings that use parameter’s values that lead to lower computational complexity and execution time of N4 bias filter, such as smaller number of iterations, smaller fitting level and larger shrink factor (Fig. 5). The aforementioned setting with the highest frequency led to optimal performance for the rest 48% of the patients, as

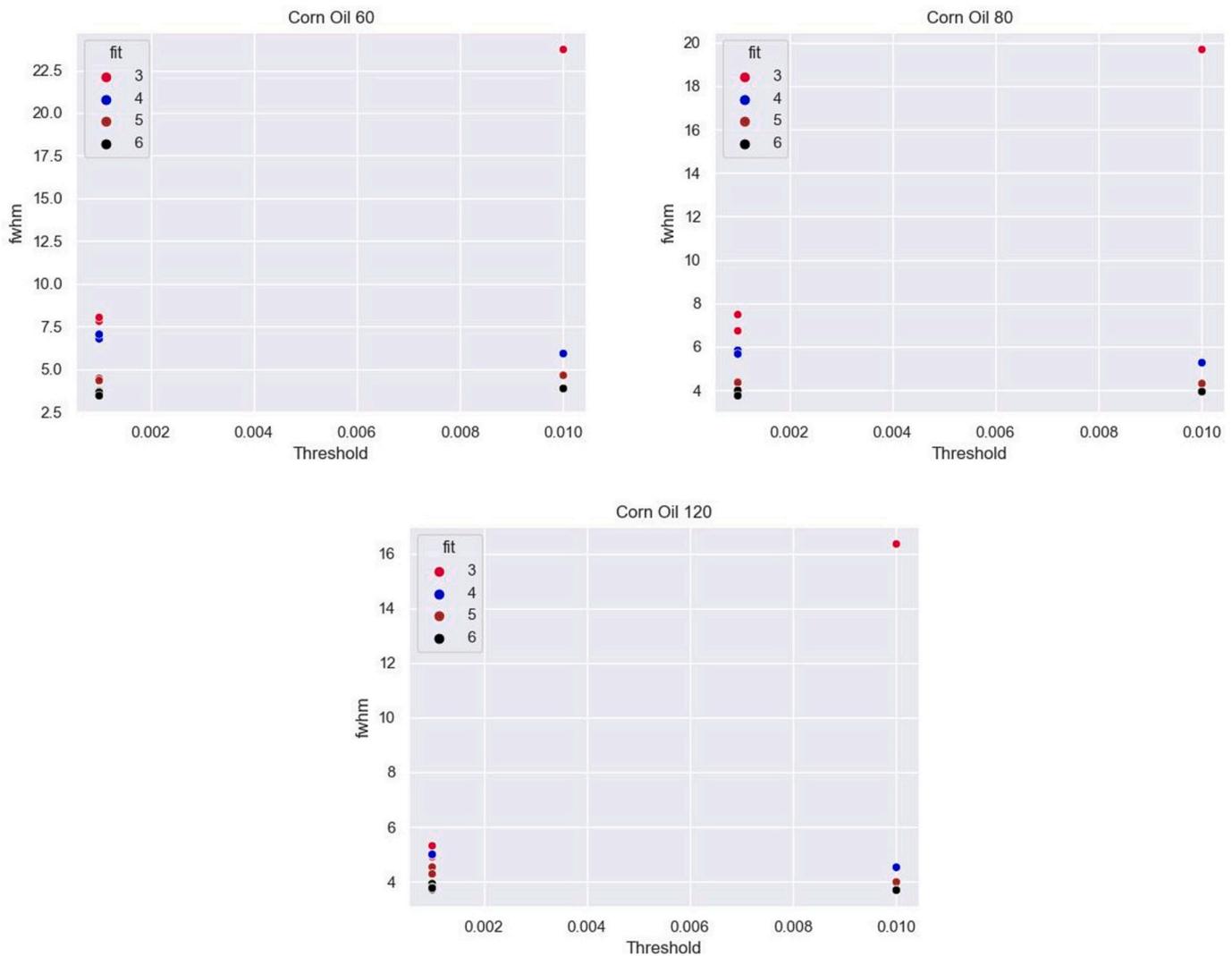


Fig. 2. Scatterplots showing the values of the FWHM achieved by various configurations of the N4 filter that were applied in the corn oil phantoms with the three different TE values. In the first row, from left to right, the scatterplots for the corn oil phantom with TE equal to 60 and 80, respectively, are depicted. In the second row, the scatterplot for the corn oil phantom with TE = 120 is illustrated.

Cornoil 60

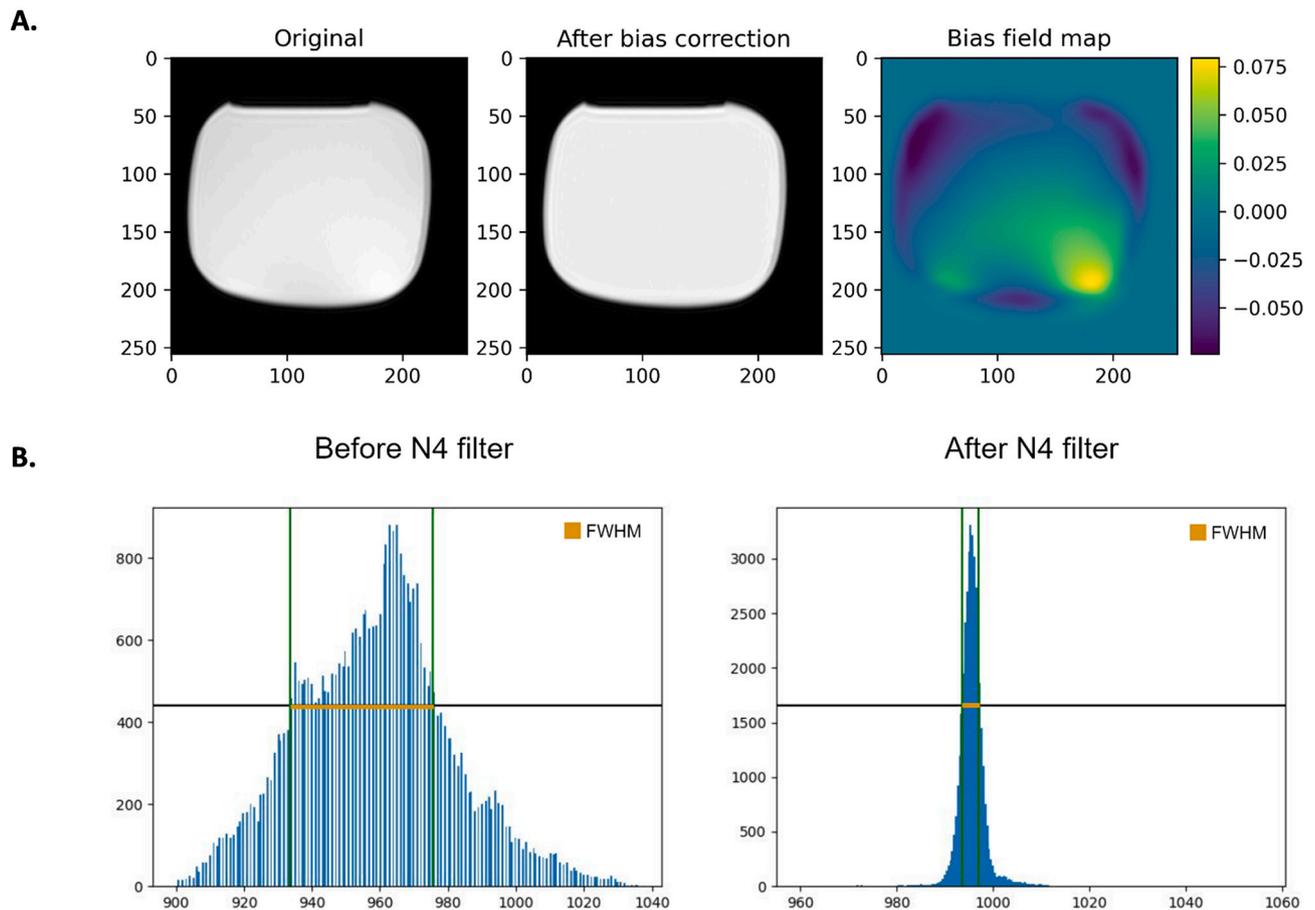


Fig. 3. A. Original image of corn oil phantom with TE 60, N4 filtered image after bias field correction and the corresponding bias field map. B. Histograms of the corn oil phantom with TE 60, before and after N4 filtering.

the FWHM value achieved with this setting was close to the minimum FWHM, having the 75th percentile lower than 20% (Fig. 5B). The same setting with the only difference in the threshold value, i.e., “mask False, threshold 0.0001, shrink factor 2, fitting level 6, iterations 100”, resulted also in small differences in the FWHM value. The FWHM value obtained from these two settings was <20% larger than the minimum FWHM in 75% of the patients.

The analysis was also performed to the patients from the PROSTATEx dataset to assess and identify the optimal N4 filter setting for MR images scanned at 3 T with surface coil. The minimum FWHM in 16% of the patients was achieved using convergence threshold 0.001, shrink factor 2, fitting level 5, number of iterations 25 and without mask (Supporting Table S2). However, this percentage is quite small, indicating that many different configurations were identified as optimum in different patients and thus further investigation was needed.

Hence, the aforementioned exploratory analysis in the values of FWHM obtained from various settings was also performed in the patients of the 3 T PROSTATEx dataset (Fig. 6). Two settings gave a FWHM close to the minimum value. The smaller values of relative difference are achieved by the settings: mask False, threshold 0.001 or 0.0001, shrink factor 2, fitting level 5, iterations 25. The difference in their values was <20% for the largest proportion of patients (Fig. 6A).

In order to validate the efficiency of the derived optimal configurations of N4 filter, the exploratory analysis was performed in two external datasets with scanning conditions similar to the already examined datasets. More specifically, the Prostate-MRI dataset was used to investigate whether the derived optimal configuration for images

scanned at 1.5 T with combined surface and endorectal coil remains optimal for images scanned at 3 T with the same receive coil configuration. The results from the exploratory analysis in the Prostate-MRI dataset showed that the derived optimal configuration “mask False, threshold 0.001, shrink factor 2, fitting level 6, iterations 100” was also optimal for images scanned at higher magnetic field strength of 3 T with combined surface and endorectal coil (Supporting Fig. S1). In this configuration, the median value of the distribution of the relative difference values in FWHM was lower than 20%. Hence, there is one configuration that is optimal for images scanned at either 1.5 T or 3 T with this receive coil setting.

Two subsets of PI-CAI dataset were used to assess the performance of the optimal N4 configurations that were derived for images of the PROSTATEx dataset scanned at 3 T with surface coil. These two validation subsets contained images scanned at 1.5 T and 3 T with the same receive coil configuration. In the subset of the 1.5 T MR images scanned with surface coil, the optimal configurations “mask False, threshold 0.001 or 0.0001, shrink factor 2, fitting level 5, iterations 25” demonstrated good performance, having the 75th percentile close to the relative difference value of 20% (Supporting Fig. S2). However, three other configurations showed even better performance achieving smaller values of relative difference than the derived optimal configuration of PROSTATEx dataset (Supporting Fig. S2). More specifically, the same configuration with the difference of using a number of iterations of 10 and/or fitting level 6 resulted in even better performance of N4 filter. In the subset of the 3 T MR images scanned with surface coil, the derived optimal configurations “mask False, threshold 0.001 or 0.0001, shrink

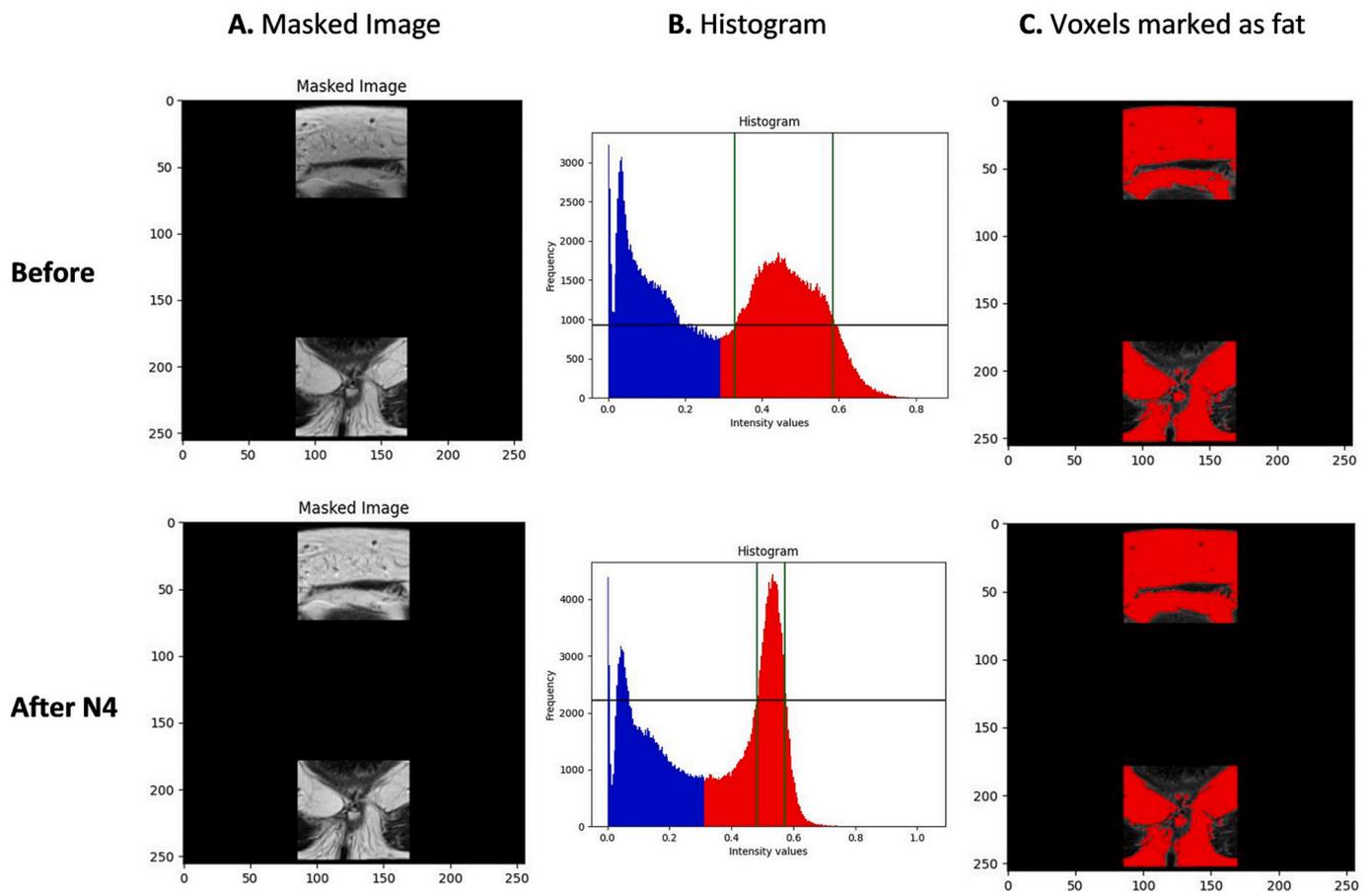


Fig. 4. In the first row, the masked image (A), the corresponding histogram (B) and the voxels of the fat distribution (C) of the original unfiltered image are shown. In the second row, the corresponding images of the N4 filtered image are presented.

factor 2, fitting level 5, iterations 25” from PROSTATEx dataset were also identified as optimum in this subset of PI-CAI dataset (Supporting Fig. S3). Hence, the same configurations were identified as optimal in images scanned at 3 T magnetic field strength with surface coil, originating either from the PROSTATEx or the PI-CAI dataset.

4. Discussion

Prostatic lesion evaluation is either based on multi-parametric MR imaging, comprising T2W, Diffusion-weighted imaging (DWI) and Dynamic Contrast Enhanced (DCE) or can be reduced to a bi-parametric examination comprising only T2W and DWI, as the most indicative acquisitions for disease characterization. Our study focused on T2W images among the T2 as Echo Planar Imaging (EPI)-based DWI sequences are almost always acquired with suppressed fat signal because of the related chemical shift artifact that severely degrades image diagnostic quality.

Periprostatic fat was chosen as the reference tissue to provide the metrics for evaluation of each parameter set, because fat tissue surrounds abdominal organs and it can be found around the target prostate gland. Furthermore, periprostatic fat is a homogeneous tissue expected to have a similar imaging appearance among patients. Lastly, it yields high signal intensity in non-fat suppressed images and thus even small relative changes can have a measurable difference among different tests and can be easily perceived.

The examined MRI prostate images were acquired from scanners with different magnetic field strengths, 1.5 T and 3 T, and different receive coil configurations, surface coil and combined surface and endorectal coil, to investigate whether the different fields and coils lead to different or same optimal configuration of the N4 filter. A decrease

larger or equal than 50% in the FWHM value before and after filtering with the optimal setting is observed for >80% of the patients in each of the examined datasets, indicating substantial improvement in the periprostatic fat distribution (Supporting Table S3-S7). More precisely, the proportion of the patients was approximately 92%, 89%, 81%, 85% and 88% at PROSTATE-DIAGNOSIS, PROSTATEx, Prostate-MRI, 1.5 T PI-CAI and 3 T PI-CAI datasets, respectively.

The analysis in datasets that contain images scanned at either 1.5 T or 3 T but with the same receive coil configuration showed that the optimal configuration is the same. However, the analysis in datasets with different receive coil configuration resulted in a different optimal configuration for the N4 filter. The effect of the different magnetic field strength in images scanned with the same receive coil configuration is observed in the FWHM values. For instance, the 75th percentile of the relative difference distribution achieved by the optimal setting was lower in the 1.5 T than in the 3 T images of the PI-CAI dataset. Lower magnetic field strengths, such as 1.5 T compared to 3 T, results in lower frequency modulation of the corrupting bias field. Hence, MR images scanned with 1.5 T is less corrupted than with 3 T. Thus, the difference in the FWHM value before and after bias field correction is smaller in most N4 configurations in images scanned with the same receive coil configuration at 1.5 T than 3 T. However, the optimal configuration is the same for these images scanned at 1.5 T or 3 T. Hence, the effect of the receive coil configuration prevails over the magnetic field strength in the extraction of the optimal N4 filter's configuration for the bias field correction in MR prostate images.

The optimal settings, which were identified for the corn oil phantom, indicated the minimum requirements for the values of the N4 parameters, as the phantom is a homogeneous material free from anatomy induced challenges. The inherently heterogeneous MR prostate images

PROSTATE-DIAGNOSIS Dataset

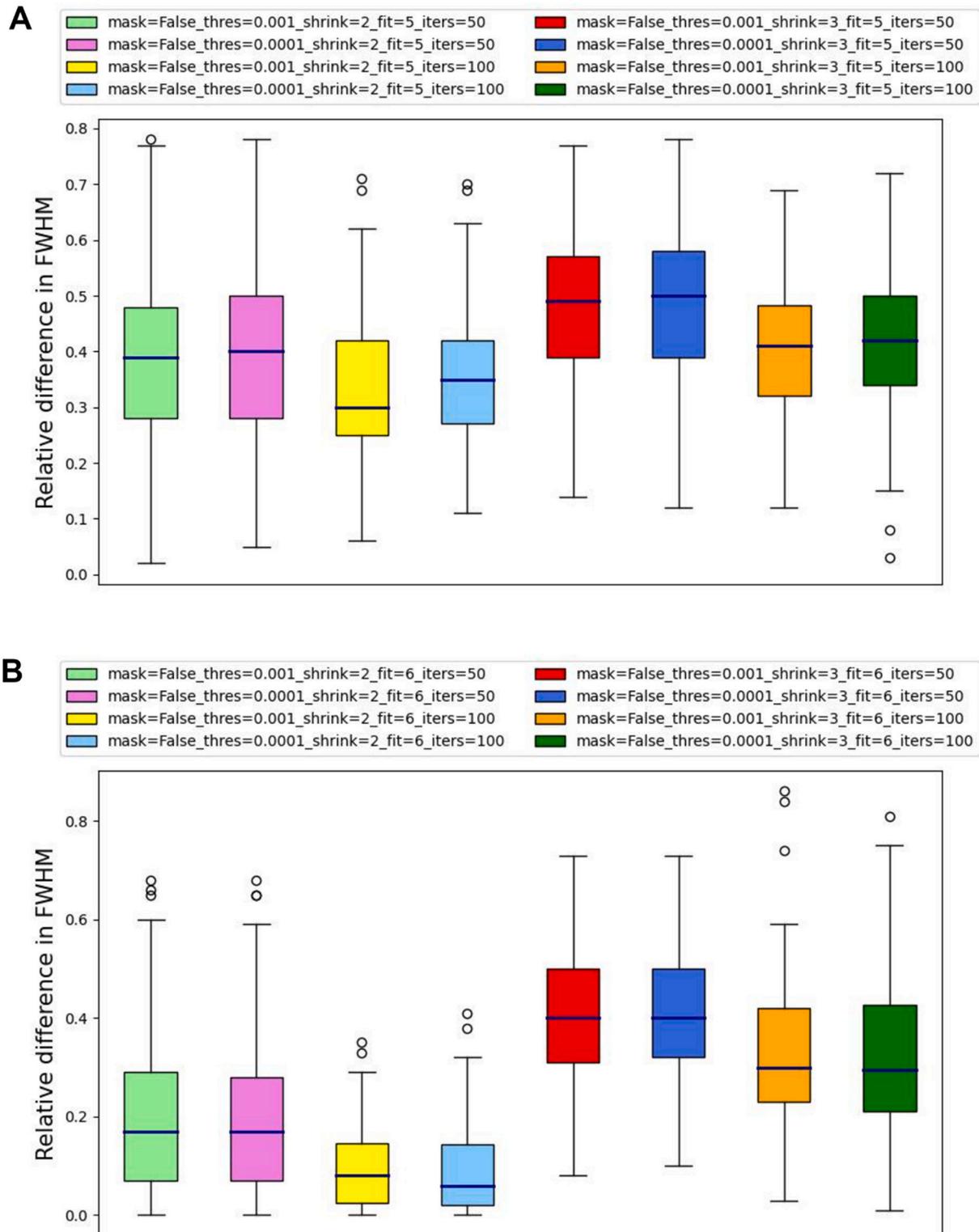


Fig. 5. A. Boxplots showing the relative difference between the FWHM of the optimal setting and the FWHM of various specific settings with fixed fitting level 5. B. Boxplots showing the relative difference between the FWHM of the optimal setting and the FWHM of various specific settings with fixed fitting level 6. These results are obtained from the patients of PROSTATE-DIAGNOSIS dataset. The relative difference is calculated for the patients whose optimal setting is not the specific setting defined in the legend.

PROSTATEx Dataset

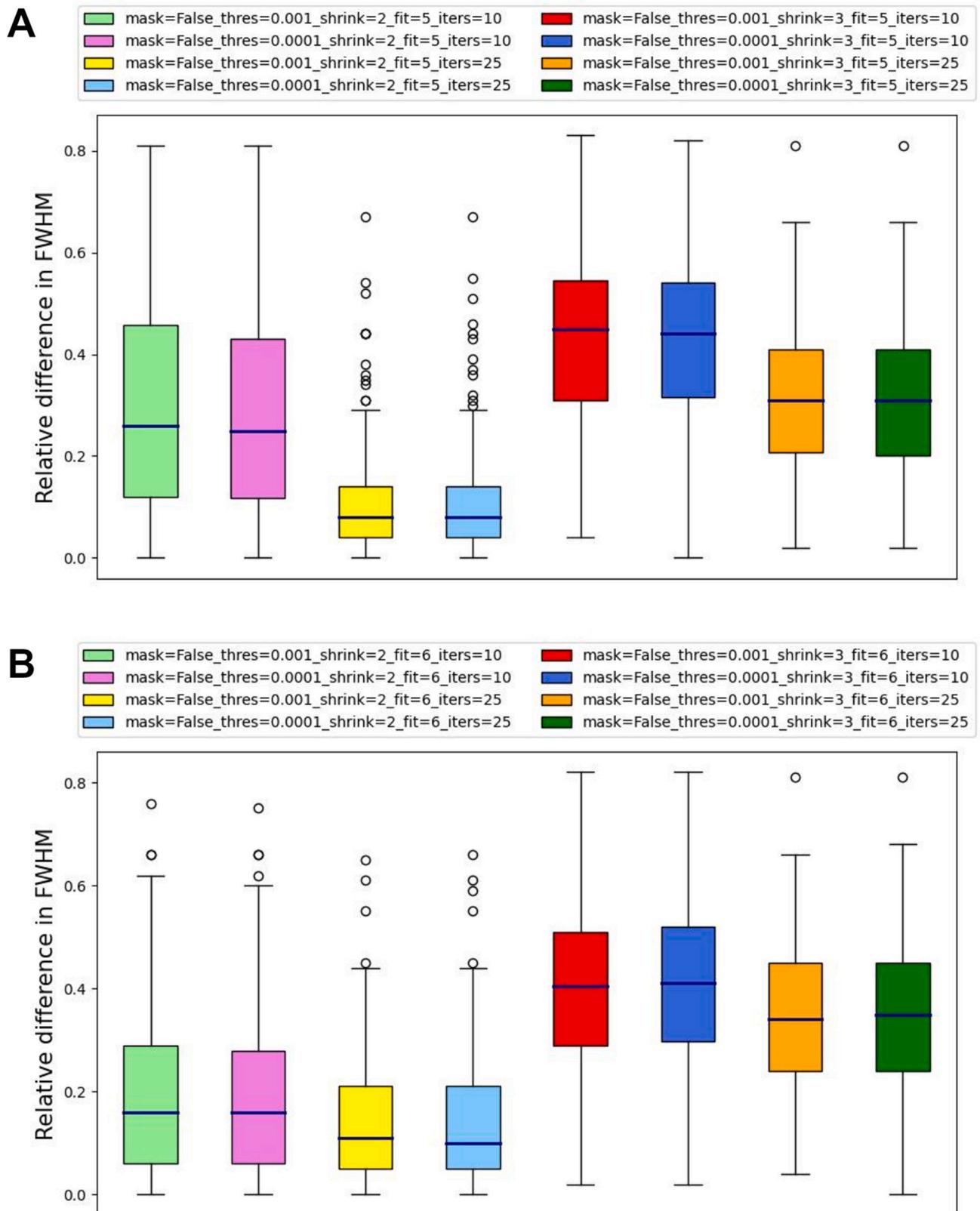


Fig. 6. A. Boxplots showing the relative difference between the FWHM of the optimal setting and the FWHM of various specific settings with fixed fitting level 5. B. Boxplots showing the relative difference between the FWHM of the optimal setting and the FWHM of various specific settings with fixed fitting level 6. These results are obtained from the patients of PROSTATEx dataset. The relative difference is calculated for the patients whose optimal setting is not the specific setting defined in the legend.

from two different receive coil configurations and both 1.5 T and 3 T magnet resulted in more “strict” values in the N4 parameters to achieve the minimum FWHM, as intuitively expected.

4.1. Convergence threshold

The effect of the convergence threshold was examined to identify the optimal value that should be used for the bias field correction of prostate images. In the phantom, there is no need to use a strict value for the convergence threshold, since the minimum FWHM is achieved using the largest among the examined values of threshold. On the other hand, a smaller threshold value 0.001 was considered as optimal from the analysis in the patients from all datasets. However, the exploratory analysis for the comparison of the FWHM achieved by the optimal and some other settings (Figs. 5 and 6) demonstrated that only a smaller threshold 0.0001 can also be used, preserving the values of the rest parameters stable in each case, which increases the computational complexity and the required execution time of the N4ITK filter.

4.2. Number of iterations

The number of iterations is also crucial for the performance and the execution time of the N4 bias filter. The algorithm is an iterative process, which makes it computationally expensive. In the phantom, the minimum FWHM can be achieved using a small number of iterations, such as 5. More precisely, the minimum FWHM is achieved independently of the number of iterations, showing that increasing the number of iterations, and thus the computational complexity, do not improve the result of the N4 filter. However, a large number of iterations, such as 100, is required when the N4 filter is applied in the heterogeneous MR prostate images scanned with a combined surface and endorectal coil. The use of a smaller number of iterations, which lead to faster execution of the N4 filter, resulted in larger values of difference in the FWHM value. However, 25 iterations are adequate for the bias field correction in MR images scanned with a surface coil. In these images, the use of a larger number of iterations has also good results, but increases the computational complexity without improving the performance of the N4 filter. However, some configurations with 10 iterations have also good results only for images scanned at 1.5 T with surface coil.

4.3. Shrink factor

The minimum FWHM was achieved using the shrink factor of 2 and thus this value is recommended. The use of shrink factor 3 resulted in substantial increase in the FWHM value in all the examined MR images. Hence, the use of a shrink factor of 2 is strongly recommended.

4.4. Fitting level

The number of the parameter fitting level, which is an advancement of the N4 compared to N3, defines the number of the resolution levels that the N4 filter uses to calculate the bias field, allowing for multi-resolution approximation. The scale of the inhomogeneity depends on the wavelength of the radiofrequency field at the field strength of the scanner. For 3 T scanners the frequency is 128 MHz, which in free space corresponds to a wavelength of about 2 m and is doubled for 1.5 T scanners. However, inside the human body these dimensions for the dielectric effects are reduced to 20 cm for 3 T and double for 1.5 T. Hence, it becomes obvious that the spline density requirements are more demanding for higher field strengths, as well as that the bias field corruption is obvious in higher field, while it may remain subtle or unperceived for 1.5 T field. However, another factor that affects the bias field corruption is the use of an endorectal coil. This coil is used to increase the signal-to-noise ratio (SNR), but it simultaneously increases the intensity near the coil, resulting in increased intensity inhomogeneity in the image [42]. In the phantom analysis, the use of a

fitting level 5 or 6 resulted in smaller values of FWHM, setting the lower boundaries for efficient bias field correction. In the patients analyses with images acquired by a surface coil, the fitting level 5 is optimum resulting to the minimum FWHM. The configurations with a fitting level 6 and small number of iterations (10 or 25) resulted in good performance, but worse than the optimal configuration with fitting level 5. In the patient analysis with images scanned with a combined surface and endorectal coil, the fitting level 6 resulted in the minimum FWHM. In these datasets, the fitting level 5 leads to substantially larger values of FWHM at images from both 1.5 T and 3 T and thus it is not recommended when an endorectal coil is used during image acquisition. The use of an endorectal coil and higher magnetic field strengths require the use of smaller values of spline distance in N4ITK due to the increased bias field corruption. The effect of the endorectal coil prevailed over the magnetic field strength and thus a fitting level 6, which results in a smaller spline distance than a fitting level 5, is required when an endorectal coil is used independently of the magnetic field strength.

The FWHM of the periprostatic fat distribution is proposed as a novel quantitative metric to evaluate the performance of the algorithm. The smaller the FWHM value, the better the performance of the N4 algorithm, implying a more homogeneous representation for the periprostatic fat tissue. This work presents a comprehensive overview of the effect of the parameters of the well-established N4ITK filter using an automatic quantitative metric. The exploratory analysis was performed on MR prostate images acquired from both 1.5 T and 3 T magnetic field strength and different receive coil configurations (surface coil and combination of surface and endorectal coil) to account for the variability induced by those factors highly relevant to the bias field corruption. In addition to those factors, each patient's specific anatomical characteristics contribute differently to the final intensity inhomogeneity of the MR image. Thus, the use of a large number of patients and validation datasets also contributed to form a valid set of datasets that embraces the expected variability of bias field manifestation in terms of its severity and location in the image. To the best of our knowledge, this is the first study that assesses the impact of the parameters' values on the performance of the N4 filter investigating a large number of different settings and proposes specific optimal configurations of the state-of-the-art N4ITK filter for the bias field correction in the heterogeneous MR prostate images.

However, this study could be more generalizable if a broader image cohort was used, imaged with different sequence types, contrasts, acquisition parameters and receive coil schemes. For instance, provided that fat suppression techniques are avoided, MR abdominal images from gradient echo T1 acquisitions as in a DCE acquisition could be included in future work. Moreover, the experimental set up with the phantom can be extended to examine the effect of other coil types (e.g., phased array coils), sequences (e.g., gradient echo) or contrasts (e.g., T1 or proton density) on a homogeneous sample. A major limitation of the N4 filter is that it is computationally expensive, resulting in large execution times. Thus, we constrained the exploratory values of the number of iterations and fitting level to the maximum values of 100 and 6, respectively, to explore possible N4 configurations that are feasible to be used in the research studies regarding the time complexity. Moreover, the set of values for each parameter was selected by using the default values the authors of the N4ITK filter suggest and scaling them up and down. Furthermore, addressing field inhomogeneities is a key element for addressing lack of reproducibility in radiomics studies. Hence, the impact of the N4 method with the derived optimal setting on the robustness and stability of radiomic analyses will be examined in future work. Moreover, the N4-corrected images will be used to develop Artificial Intelligence (AI) tools to non-invasively detect prostate cancer and predict the aggressiveness of the cancer. Machine learning models using the radiomic features extracted from the N4-filtered images and deep learning models using as input the N4-filtered images will be developed. The efficiency of the N4 filter will be investigated on assessing whether the application of the N4 filter enhances the performance of the AI

models instead of using raw data.

5. Conclusion

In this study, the performance of the state-of-the-art, for bias field correction, N4ITK filter with various configurations was quantitatively evaluated to identify the optimal set of parameters' values and thus improve the bias field correction in MR prostate images. The FWHM of the periprostatic fat distribution was automatically calculated to quantify the impact of the N4 filter on the correction of the intensity inhomogeneities. The periprostatic fat tissue was used as reference tissue to evaluate the different configurations of the N4 filter due to its beneficial position around the prostate and its magnetic properties. A large number of possible different configurations of N4 filter was applied to 1.5 T and 3 T MRI prostate images acquired by a surface coil only or a combination of surface and endorectal coil to identify the setting with the best performance. The exploratory analysis indicated that the optimal values of the parameters of the N4ITK filter for bias field correction in MRI prostate images acquired by a surface coil are convergence threshold 0.001, shrink factor 2, fitting level 5, number of iterations 25 and without mask (i.e. using the default mask). In MR images scanned by a combination of surface and endorectal coil, the optimal configuration is convergence threshold 0.001, shrink factor 2, fitting level 6, number of iterations 100 and without mask (i.e. using the default mask), independently of the magnetic field strength. The same configurations with the only difference of using a convergence threshold 0.0001 while keeping the rest of the parameters unchanged, could also be used for the bias field correction of MR prostate images; however, it leads to higher computational complexity and execution time of the filter. The N4-corrected images using the derived optimal configurations will be used to assess the reproducibility of the radiomic features and develop AI tools for the detection and the prediction of the prostate cancer aggressiveness in a future study.

Although the analysis focused on fat distribution, the target organ for reliable quantification is the prostate gland, being an organ very commonly affected by oncologic and non-oncologic pathologies requiring accurate evidence-based classification. The use of the proposed methodology can be translated to any other abdominal organ surrounded by visceral fat. Furthermore, similarly to the prostate, the breast region presents similar challenges and thus the benefit of the N4 filter can be examined in such a challenging imaging field containing fat-air interfaces.

CRedit authorship contribution statement

Aikaterini Dovrou: Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft. **Katerina Nikiforaki:** Conceptualization, Validation, Investigation, Writing – Original Draft, Writing – Review & Editing. **Dimitris Zaridis:** Data curation. **Georgios C. Manikis:** Writing - Review & Editing. **Eugenia Mylona:** Writing - Review & Editing. **Nikolaos Tachos:** Writing - Review & Editing. **Manolis Tsiknakis:** Writing - Review & Editing. **Dimitrios I. Fotiadis:** Writing – Review & Editing. **Kostas Marias:** Supervision, Writing – Review & Editing.

Declaration of Competing Interest

None.

Acknowledgement

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 952159 (ProCancer-I). This work reflects the author's view. The Commission is not responsible for any use that may be made of the information it contains.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.mri.2023.03.012>.

References

- [1] Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* 2021;71:7–33. <https://doi.org/10.3322/CAAC.21654>.
- [2] Boesen L, Chabanova E, Løgager V, Balslev I, Mikines K, Thomsen HS. Prostate cancer staging with extracapsular extension risk scoring using multiparametric MRI: a correlation with histopathology. *Eur Radiol* 2015;25:1776–85. <https://doi.org/10.1007/s00330-014-3543-9>.
- [3] Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 2017;14:749–62. <https://doi.org/10.1038/nrclinonc.2017.141>.
- [4] Hoult DI. The principle of reciprocity in signal strength calculations - a mathematical guide. *Concepts Magn Reson* 2000;12:173–87. [https://doi.org/10.1002/1099-0534\(2000\)12:4<173::AID-CMR1>3.0.CO;2-Q](https://doi.org/10.1002/1099-0534(2000)12:4<173::AID-CMR1>3.0.CO;2-Q).
- [5] Song S, Zheng Y, He Y. A review of methods for bias correction in medical images. *Biomed Eng Rev* 2017;1. <https://doi.org/10.18103/BME.V3I1.1550>.
- [6] Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans Med Imaging* 2010;29:1310–20. <https://doi.org/10.1109/TMI.2010.2046908>.
- [7] Hsu SH, Cao Y, Lawrence TS, Tsien C, Feng M, Grodzki DM, et al. Quantitative characterizations of ultrashort echo (UTE) images for supporting air-bone separation in the head. *Phys Med Biol* 2015;60:2869–80. <https://doi.org/10.1088/0031-9155/60/7/2869>.
- [8] Fang L, Wang X. Brain tumor segmentation based on the dual-path network of multi-modal MRI images. *Pattern Recognit* 2022;124. <https://doi.org/10.1016/j.patcog.2021.108434>.
- [9] Ullah F, Ansari SU, Hanif M, Ayari MA, Chowdhury MEH, Khandakar AA, et al. Brain mr image enhancement for tumor segmentation using 3d u-net. *Sensors* 2021;21:1–14. <https://doi.org/10.3390/s21227528>.
- [10] Wang M, Yang J, Chen Y, Wang H. The multimodal brain tumor image segmentation based on convolutional neural networks. In: 2017 2nd IEEE Int Comput Intell Appl (ICCIA), Beijing, China; 2017. p. 336–9. <https://doi.org/10.1109/CIAPP.2017.8167234>.
- [11] Saman S, Narayanan SJ. Active contour model driven by optimized energy functionals for MR brain tumor segmentation with intensity inhomogeneity correction. *Multimed Tools Appl* 2021;80:21925–54. <https://doi.org/10.1007/s11042-021-10738-x>.
- [12] Nguyen AA, Onishi N, Carmona-bozo J, Li W, Kornak J, Newitt DC, et al. Post-processing bias field inhomogeneity correction for assessing background parenchymal enhancement on breast MRI as a quantitative marker of treatment response. *Tomography* 2022;8:891–904. <https://doi.org/10.3390/tomography8020072>.
- [13] Gitto S, Bologna M, Corino VDA, Emili I, Albano D, Messina C, et al. Diffusion-weighted MRI radiomics of spine bone tumors: feature stability and machine learning-based classification performance. *Radiol Med* 2022. <https://doi.org/10.1007/s11547-022-01468-7>.
- [14] Chang E, Joel MZ, Chang HY, Du J, Khanna O, Omuro A, et al. Comparison of radiomic feature aggregation methods for patients with multiple tumors. *Sci Rep* 2021;11:1–7. <https://doi.org/10.1038/s41598-021-89114-6>.
- [15] Sled J, Zijdenbos A, Evans A. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans Med Imaging* 1998;17:87–97. <https://doi.org/10.1109/42.668698> [PMID: 9617910].
- [16] Vovk U, Pernuš F, Likar B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans Med Imaging* 2007;26:405–21. <https://doi.org/10.1109/TMI.2006.891486>.
- [17] Arnold JB, Liow JS, Schaper KA, Stern JJ, Sled JG, Shattuck DW, et al. Qualitative and quantitative evaluation of six algorithms for correcting intensity nonuniformity effects. *Neuroimage* 2001;13:931–43. <https://doi.org/10.1006/nimg.2001.0756>.
- [18] Leger S, Löck S, Hietschold V, Haase R, Böhme HJ, Abolmaali N. Physical correction model for automatic correction of intensity non-uniformity in magnetic resonance imaging. *Phys Imaging Radiat Oncol* 2017;4:32–8. <https://doi.org/10.1016/j.phro.2017.11.003>.
- [19] Zin YC, Zheng W, Chee MWL, Zagorodnov V. Evaluation of performance metrics for bias field correction in MR brain images. *J Magn Reson Imaging* 2009;29:1271–9. <https://doi.org/10.1002/jmri.21768>.
- [20] Li Y, Ammari S, Balleyguier C, Lassau N, Chouzenoux E. Impact of preprocessing and harmonization methods on the removal of scanner effects in brain mri radiomic features. *Cancers (Basel)* 2021;13:1–22. <https://doi.org/10.3390/cancers13123000>.
- [21] Saint Martin MJ, Orihac F, Aki P, Khalid F, Nioche C, Buvat I, et al. A radiomics pipeline dedicated to breast MRI: validation on a multi-scanner phantom study. *Magn Reson Mater Phys Biol Med* 2021;34:355–66. <https://doi.org/10.1007/s10334-020-00892-y>.
- [22] Moradmand H, Aghamiri SMR, Ghaderi R. Impact of image preprocessing methods on reproducibility of radiomic features in multimodal magnetic resonance imaging in glioblastoma. *J Appl Clin Med Phys* 2020;21:179–90. <https://doi.org/10.1002/acm2.12795>.

- [23] Bologna M, Corino V, Mainardi L. Technical note: virtual phantom analyses for preprocessing evaluation and detection of a robust feature set for MRI-radiomics of the brain. *Med Phys* 2019;46:5116–23. <https://doi.org/10.1002/mp.13834>.
- [24] Boyes RG, Gunter JL, Frost C, Janke AL, Yeatman T, Hill DLG, et al. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *Neuroimage* 2008;39:1752–62. <https://doi.org/10.1016/j.neuroimage.2007.10.026>. Intensity.
- [25] Zheng W, Chee MWL, Zagorodnov V. Improvement of brain segmentation accuracy by optimizing non-uniformity correction using N3. *Neuroimage* 2009;48:73–83. <https://doi.org/10.1016/j.neuroimage.2009.06.039>.
- [26] Gispert JD, Reig S, Pascual J, Vaquero JJ, García-Barreno P, Desco M. Method for bias field correction of brain T1-weighted magnetic images minimizing segmentation error. *Hum Brain Mapp* 2004;22:133–44. <https://doi.org/10.1002/hbm.20013>.
- [27] Liao L, Lin T, Li B. MRI brain image segmentation and bias field correction based on fast spatially constrained kernel clustering approach. *Pattern Recogn Lett* 2008;29:1580–8. <https://doi.org/10.1016/j.patrec.2008.03.012>.
- [28] Chen Y, Zhao B, Zhang J, Zheng Y. Automatic segmentation for brain MR images via a convex optimized segmentation and bias field correction coupled model. *Magn Reson Imaging* 2014;32:941–55. <https://doi.org/10.1016/j.mri.2014.05.003>.
- [29] Ioannidis GS, Nikiforaki K, Kalaitzakis G, Karantanas A, Marias K, Maris TG. Inverse Laplace transform and multiexponential fitting analysis of T2 relaxometry data: a phantom study with aqueous and fat containing samples. *Eur Radiol Exp* 2020;4:28. <https://doi.org/10.1186/s41747-020-00154-5>.
- [30] Nikiforaki K, Ioannidis GS, Lagoudaki E, Manikis GH, de Bree E, Karantanas A, et al. Multiexponential T2 relaxometry of benign and malignant adipocytic tumours. *Eur Radiol Exp* 2020;4. <https://doi.org/10.1186/s41747-020-00175-0>.
- [31] Bloch BN, Jain A, Jaffe CC. Data from PROSTATE-DIAGNOSIS. *Cancer Imaging Arch* 2015. <https://doi.org/10.7937/K9/TCIA.2015.FOQEUJVT>.
- [32] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging* 2013;26:1045–57. <https://doi.org/10.1007/s10278-013-9622-7>.
- [33] PROSTATE-DIAGNOSIS. The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. <https://wiki.cancerimagingarchive.net/display/Public/PROSTATE-DIAGNOSIS#3277254e31de0388cfa43f6af738c8e20831bd1>; 2015 (accessed February 3, 2022).
- [34] SPIE-AAPM-NCI PROSTATEx Challenges (PROSTATEx). The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. <https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=23691656#23691656e9a7bb92dcee4b419511436cc3a364b3>; 2017 (accessed October 15, 2021).
- [35] Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. ProstateX challenge data. *Cancer Imaging Arch* 2017. <https://doi.org/10.7937/K9TCIA.2017.MURS5CL>.
- [36] Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging* 2014;33:1083–92. <https://doi.org/10.1109/TMI.2014.2303821>.
- [37] Choyke P, Turkbey B, Pinto P, Merino M, Wood B. Data from PROSTATE-MRI. *Cancer Imaging Arch* 2016. <https://doi.org/10.7937/K9/TCIA.2016.6046GUDv>.
- [38] PROSTATE-MRI. The Cancer Imaging Archive (TCIA) Public Access - Cancer Imaging Archive Wiki. <https://wiki.cancerimagingarchive.net/display/Public/PROSTATE-MRI#3277260bb280e5de24945ab904c074b2c168664>; 2016 (accessed July 13, 2022).
- [39] Saha A, Twilt JJ, Bosma JS, Van Ginneken B, Yakar D, Elschot M, et al. Artificial Intelligence and Radiologists at Prostate Cancer Detection in MRI: The PI-CAI Challenge. 2022. <https://doi.org/10.5281/zenodo.6522364>.
- [40] The PI-CAI Challenge. Public Training and Development Dataset. <https://zenodo.org/record/6624726#.YxIX1XZBzIU>; 2022 (accessed July 13, 2022).
- [41] Zaridis D, Mylona E, Tachos N, Marias K, Tsiknakis M, Fotiadis DI. A deep learning-based cropping technique to improve segmentation of prostate's peripheral zone. In: *BIBE 2021 - 21st IEEE Int Conf Bioinforma Bioeng Proc*; 2021. <https://doi.org/10.1109/BIBE52308.2021.9635576>.
- [42] Lui D, Modhafar A, Glaister J, Wong A, Haider MA. Monte Carlo bias field correction in endorectal diffusion imaging. *IEEE Trans Biomed Eng* 2014;61:368–80. <https://doi.org/10.1109/TBME.2013.2279635>.