

Harmonisation of scanner-dependent contrast variations in magnetic resonance imaging for radiation oncology, using style-blind auto-encoders

Kavi Fatania^a, Anna Clark^b, Russell Frood^a, Andrew Scarsbrook^a, Bashar Al-Qaisieh^b, Stuart Currie^a, Michael Nix^{b,*}

^a Department of Radiology, St James University Hospital Trust, Beckett Street, Leeds LS9 7TF, UK

^b Leeds Cancer Centre, Bexley Wing, St James University Hospital Trust, Beckett Street, Leeds LS9 7TF, UK

ARTICLE INFO

Keywords:

Magnetic Resonance Imaging (MRI)
harmonisation
Image translation
Oncology
Deep-learning
Radiomics

ABSTRACT

Background and purpose: Magnetic Resonance Imaging (MRI) exhibits scanner dependent contrast, which limits generalisability of radiomics and machine-learning for radiation oncology. Current deep-learning harmonisation requires paired data, retraining for new scanners and often suffers from geometry-shift which alters anatomical information. The aim of this study was to investigate style-blind auto-encoders for MRI harmonisation to accommodate unpaired training data, avoid geometry-shift and harmonise data from previously unseen scanners. **Materials and methods:** A style-blind auto-encoder, using adversarial classification on the latent-space, was designed for MRI harmonisation. The public CC359 T1-w MRI brain dataset includes six scanners (three manufacturers, two field strengths), of which five were used for training. MRI from all six (including one unseen) scanner were harmonised to common contrast. Harmonisation extent was quantified via Kolmogorov-Smirnov testing of residual scanner dependence of 3D radiomic features, and compared to WhiteStripe normalisation. Anatomical content preservation was measured through change in structural similarity index on contrast-cycling (δ SSIM).

Results: The percentage of radiomics features showing statistically significant scanner-dependence was reduced from 41% (WhiteStripe) to 16% for white matter and from 39% to 27% for grey matter. δ SSIM < 0.0025 on harmonisation and de-harmonisation indicated excellent anatomical content preservation.

Conclusions: Our method harmonised MRI contrast effectively, preserved critical anatomical details at high fidelity, trained on unpaired data and allowed zero-shot harmonisation. Robust and clinically translatable harmonisation of MRI will enable generalisable radiomic and deep-learning models for a range of applications, including radiation oncology treatment stratification, planning and response monitoring.

1. Introduction

Quantitative analysis of magnetic resonance imaging (MRI) for diagnosis, prognosis and treatment personalisation in oncology is attractive due to the inherent richness of MRI soft tissue contrast. Radiomics involves quantitative extraction of pre-defined features from medical images [1,2]. First-order features describe statistical properties of voxel intensities within a region-of-interest (ROI). Second-order features describe local spatial relationships (texture) within the ROI and quantify heterogeneity. Higher-order features are extracted from filtered or transformed images. Features can be combined via machine-learning models for diagnostic or prognostic decision support, or therapeutic

design [1]. Radiomic models have been successful in single-centre research settings, but they often lack generalisability and clinical translation has been limited [3], due to non-biological variation in image intensity and scanner-dependent contrast variations [4–7].

Harmonisation aims to transform images from multiple sources to a statistically indistinguishable common contrast space. Carre et al. demonstrated poor radiomic feature consistency for low grade gliomas between scanners with different field strengths [6], but found statistical intensity normalisation (Z-score, Nyul [7] and WhiteStripe [8]) improved feature reproducibility and tumour grading. Z-score normalises using mean intensity and standard deviation of the ROI; WhiteStripe uses normal-appearing white matter to standardise voxel intensities [8].

* Corresponding author.

E-mail addresses: kavi.fatania@nhs.net (K. Fatania), anna.clark11@nhs.net (A. Clark), russellfrood@nhs.net (R. Frood), a.scarsbrook@nhs.net (A. Scarsbrook), bashar.al-qaisieh@nhs.net (B. Al-Qaisieh), stuartcurrie@nhs.net (S. Currie), michael.nix@nhs.net (M. Nix).

<https://doi.org/10.1016/j.phro.2022.05.005>

Received 30 September 2021; Received in revised form 6 May 2022; Accepted 9 May 2022

Available online 17 May 2022

2405-6316/© 2022 The Author(s). Published by Elsevier B.V. on behalf of European Society of Radiotherapy & Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

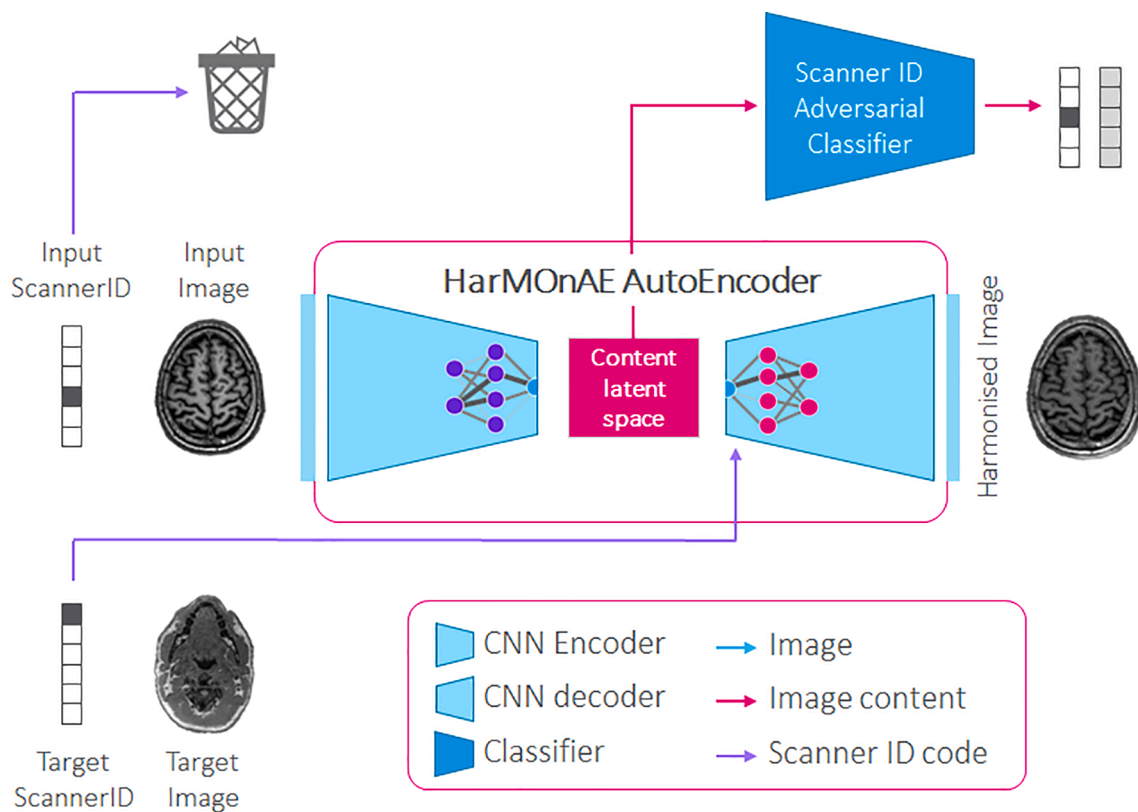


Fig. 1. HarMONAE architecture. An auto-encoder was trained to optimally reconstruct input images from a scanner-independent ‘content’ latent space (z) and an explicitly injected scanner ID (contrast code). Latent space scanner-independent was enforced by adversarially training against a classifier which attempted to establish scanner ID from z . Harmonisation was achieved by exchanging scanner ID. Content preservation on re-encoding harmonised images was encouraged with an L1 loss on z .

Nyul relies on histogram matching, which can distort normal anatomy [6,7] due to the dependence of the reference histogram on the anatomical distribution of the training data [9]. These methods are all sensitive to ROI segmentation quality and require skull stripping which can fail in clinical practice [10]. Intensity normalisation improves radiomic feature reproducibility but is generally insufficient to fully harmonise scanner-dependent contrast [14], which also exhibits spatial dependence.

Removal of scanner effects following radiomic feature extraction has been demonstrated using ComBat [11,12]. Li et al. reported increased radiomic feature robustness in glioblastoma [13]. However, ComBat is performed cohort-wise and must be re-run on addition of new patient data, limiting it to research contexts [11].

Supervised, deep-learning based MRI harmonisation has been demonstrated [14] using U-net convolutional neural networks. The need for paired data (same patients scanned on multiple scanners) makes model training expensive and restricts applicability to the brain, where good MR signal homogeneity and rigid image registration are possible. Harmonisation has been addressed by conditional generative adversarial networks (cGANs), which can synthesise images with a given contrast. The conditioning (input) image defined ‘content’ which should be retained whilst modifying the contrast to that of a target scanner, using a discriminator which predicted whether an image was a ‘true’ target scanner image or a synthesised one. cGANs produced visually impressive ‘style-transfer’ results in non-medical tasks and the cycleGAN method used unpaired training data and unsupervised learning [15]. Baysham et al [16] demonstrated MRI harmonisation using cycleGANs, improving deep-learning predictions for brain age and schizophrenia classification. However, cGANs do not inherently distinguish content from contrast, leading to the possibility that anatomical details are altered to become more like the target scanner dataset, leading to

geometry-shift. cycleGANs, which relied only on cycle-consistency to encourage content preservation were particularly susceptible. Failure to preserve anatomical information occurred if the network learnt a ‘circle-square-circle’ mapping, altering the content in the harmonised domain, but restoring it in cycled images. For robust quantitative analysis of MRI via radiomics or machine-learning, such changes in anatomical or physiological detail on harmonisation are not acceptable.

Disentangled-representation learning attempts to explicitly separate style (contrast) and content of an image, enabling style to be altered, whilst preserving content. CALAMITI [17] used limited paired (intra-scanner) data to learn to disentangle contrast and content for multi-sequence MRI. They then used the resulting content representation for inter-scanner harmonisation, on the assumption that inter-sequence and inter-scanner content representations were interchangeable. Moyer et al. [18] proposed harmonisation of diffusion-weighted MRI through scanner-independent representations using variational auto-encoders for patch-wise harmonisation. The scanner-independent latent representation was constrained by minimizing mutual information with the input scanner-ID. However, an image-based discriminator was needed to improve harmonisation, potentially reintroducing the geometry-shift problems observed in cGANs.

The aim of this study was to demonstrate a simple and generalisable harmonisation approach, HarMONAE, based on style-blind auto-encoders, with adversarial scanner classification on the latent-space, to encourage a scanner-independent content representation. We addressed limitations of previous methods, using unpaired data, allowing zero-shot harmonisation of data from scanners not included in model training and ensuring anatomical detail was preserved at high fidelity.

2. Materials and methods

2.1. Style-blind autoencoders

Auto-encoders learn to compress image information into a ‘latent representation’ using an encoder neural network and decompress it to reconstruct the original image using a decoder. Our encoder was designed to preserve *only content information* in the latent representation, discarding scanner-dependent contrast information. Scanner-ID information was explicitly injected into the decoder to enable auto-encoding to the source-scanner contrast space. For harmonisation, the content-only latent representation was combined with the new target scanner-ID at the decoder input, leading to the desired contrast change.

To make the latent representation scanner-contrast independent, we trained a classifier network to predict the scanner-ID from the latent representation and *penalise the encoder if it succeeded*. This adversarial loss ensured that by learning to accurately reconstruct the images, while removing contrast information to *minimise* the performance of this classifier, the auto-encoder converged to a ‘content-only’ latent representation. This representation could become maximally scanner-independent at no cost to auto-encoding performance, because scanner-ID was explicitly provided to the decoder.

Due to the lack of an image discriminator, we did not define a particular target contrast for training. We relied on contrast diversity between the training scanners to learn a scanner-independent representation of clinical content. Hence, harmonisation from any training contrast space to any other was possible, as well as zero-shot harmonisation from scanners not included in training. The absence of an image-based discriminator was critical for content fidelity, as HarMO-nAE didn’t attempt to make images ‘similar’ to real examples from a target domain. This avoided the problem of learning anatomical features from the target data and the consequent geometry-shift.

2.2. Architecture and loss

HarMO-nAE (Fig. 1) consisted of an auto-encoder based on the multimodal unsupervised image-to-image translation (MUNIT) architecture [19] for style transfer.

The encoder consisted of two strided 2D convolutional (down-sampling) layers, followed by four residual blocks, each containing two unstrided 2D convolutional layers, with instance normalisation and ReLU activation throughout. The decoder architecture mirrored the encoder, except for the use of adaptive instance normalisation (AdaIN) at each residual block, enabling injection of contrast information via reweighting of feature values, and layer-normalisation to preserve the injected contrast at the transposed convolutional (upsampling) layers. A multi-layer perceptron (MLP) was used to learn parameters for AdaIN from the injected scanner contrast code, enabling the decoder to interpret this information and harmonise the images. Whereas in MUNIT, this ‘style-code’ was learned from the input images by a style-encoder, we directly used the ground-truth scanner-ID.

The scanner classifier took the latent representation as input and consisted of five strided convolutional layers with layer normalisation and ReLU activation. A fully-connected layer with softmax activation predicted source scanner-ID.

The initial convolutional filter depth of 16 was doubled with each downsampling layer in encoder, decoder and classifier.

The HarMO-nAE auto-encoder loss had two key components; 1) L1 loss on the reconstructed image and its first derivatives (to improve sharpness). 2) An adversarial loss against the latent-space scanner-ID classifier, which was computed as:

$$L_{adv} = \langle \text{abs}(-\log(1/n) - CCE(S_{input}, S_z)) \rangle$$

where n was the number of scanner classes, S_{input} was the scanner label and S_z was the predicted label derived from the content latent space.

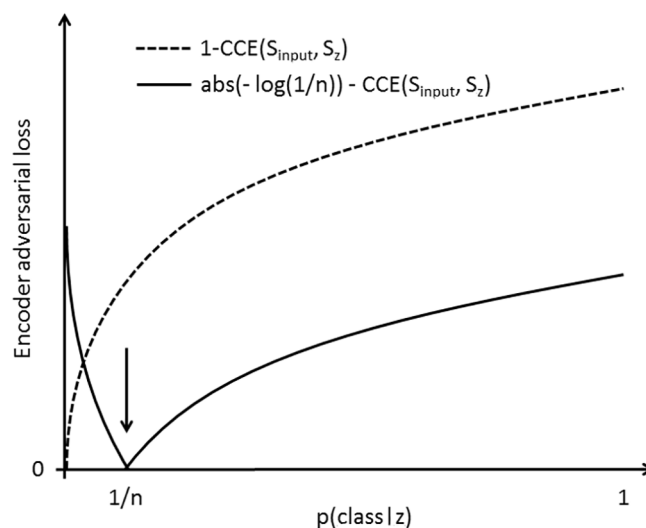


Fig. 2. Behaviour of the adversarial loss term. Complete perplexity occurred at $p(\text{class}|z) = 1/n$. Whereas 1-CCE loss (dashed) would encourage the encoder to introduce domain information from other scanners, our loss (solid) exhibited a sharp minimum at perplexity, encouraging true scanner-independence.

This loss on the auto-encoder increased as the classifier performance increased, encouraging the auto-encoder to remove scanner-dependent information from the latent space. For this multiclass problem, complete classifier perplexity occurred at $(p_{\text{class}}|z) = 1/n$ for the source class, not zero (which would imply the latent space included style information from *another* domain). The $-\log(1/n)$ term caused the generator loss function to become negative at $(p_{\text{class}}|z) = 1/n$, which represented maximal perplexity. The absolute loss (Fig. 2) exhibited a minimum at $(p_{\text{class}}|z) = 1/n$, ensuring the content encoder converged to a scanner-independent representation.

The scanner classifier had a categorical cross entropy loss between scanner-ID and predicted ID, leading to an adversarial mini-max condition when trained with the auto-encoder.

2.3. Training

HarMO-nAE was implemented in Tensorflow 2.5.0 on Python 3.8 and trained on 53995 2D slices from 250 patients, imaged on five scanners (50 per scanner) from the CC359 public dataset [20] of normal T1-weighted brain MRI, for 80 epochs. Auto-encoder and adversarial classifier were trained alternately. Training data were augmented with unit probability, using random translations and rotations drawn from zero-centred normal distributions scaled to have standard deviation of 5 cm and 5 degrees respectively. A total of 25 randomly-selected cases (5 per scanner) were reserved for testing and a further 25 for validation. 60 cases from a 6th scanner, not included in the training, were used for zero-shot validation.

Structural similarity (SSIM, scikit-image 0.19.0) was used to assess the similarity of the input, auto-encoded, harmonised and contrast-cycled images. Due to lack of paired data, we computed delta-SSIM (dSSIM), which was referenced to the auto-encoded image, enabling measurement of the change in similarity on harmonisation (dSSIM-h) and on returning (cycling) the harmonised image to its native contrast space (dSSIM-c). dSSIM-c for perfect harmonisation and content preservation was zero by definition. dSSIM-h was bounded above by the inherent dissimilarity of input and target contrast and was expected to converge on this bound as model performance improved. Hyper-parameters, including filter depths and loss weights were tuned via manual search, based on convergence of dSSIM-h.

We implemented an additional image-based discriminator, as used in MUNIT [19], commonly used in cGAN methods and adopted by Moyer

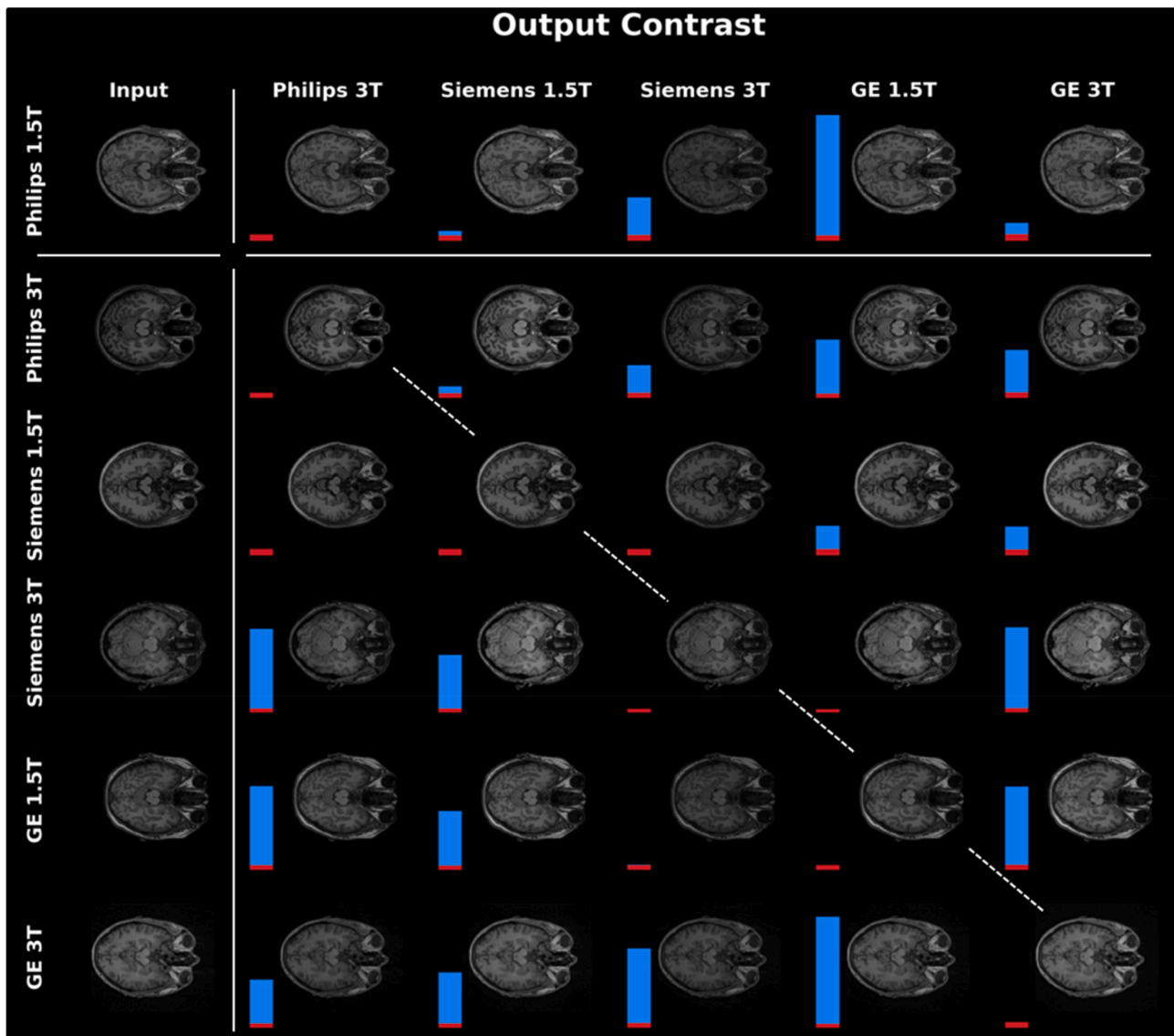


Fig. 3. Harmonisation from all scanners (input data – far left) to all five trained output contrasts. Note Philips 1.5 T was excluded from training, to demonstrate zero-shot harmonisation. Degree of harmonisation, dSSIM-h given by blue bar size, degree of content-loss, dSSIM-c given by red bar size. Images on the diagonal are auto-encoded (no harmonisation, dSSIM-h = 0).

et al. [18]. This discriminator was optionally used during training, to study its impact on geometry-shift and harmonisation.

2.4. Radiomic analysis

Following HarMOnAE training and harmonisation of validation data, radiomic analysis was performed on unharmonised and harmonised images. All images were processed in R-Studio for OS X (v 1.4.1106). N4 bias-field correction [21] was applied using ‘ANTsR’ [22]. FSL tools [23] (FMRIB library) were implemented using ‘neuroconductor’ [24] and ‘FSLR’ [25]. Skull stripping (FSL-BET) used option “B” to remove neck voxels [26,27]. White and grey matter masks were produced using FSL FAST [28]. WhiteStripe normalisation was optionally applied using the ‘WhiteStripe’ package in R. 2D slices were stacked to create 3D volumes and mask alignment checked visually. Default parameters were used throughout.

3D textural radiomic features were extracted from a) unharmonised; b) WhiteStripe normalised, and c) HarMOnAE harmonised images, with the open source ‘pyradiomics’ package in Python (v3.0.1) [29]. Feature extraction [30] was performed with 64 bins, with voxels resampled

isotropically (2 mm) using ITK b-spline interpolation [31]. 75 textural features were extracted: grey-level co-occurrence matrix (GLCM, 24), grey-level run length matrix (GLRLM, 16), neighbouring grey-tone difference matrix (NGTDM, 5), grey-level dependence matrix (GLDM, 14) and grey-level size zone matrix (GLSZM, 16).

Due to the absence of paired data, we performed unpaired statistical testing to determine similarity between radiomic feature distributions. Parametric testing involved Welch’s two-tailed unpaired *t*-test for unequal variances, to determine the similarity of radiomic feature distributions deriving from given source scanner and the remaining scanners. Non-parametric 2-sided Kolmogorov-Smirnov (KS) tests were also used to determine similarity of cumulative distribution functions from a given source scanner and the remaining scanners. For successful harmonisation, no difference was expected between these distributions. $p < 0.05$ was considered significant, representing a difference in radiomic distributions and harmonisation failure for that particular radiomic feature and scanner. Performance was assessed for ‘zero-shot’ harmonisation of data from a previously unseen scanner. All statistical testing was performed in R.

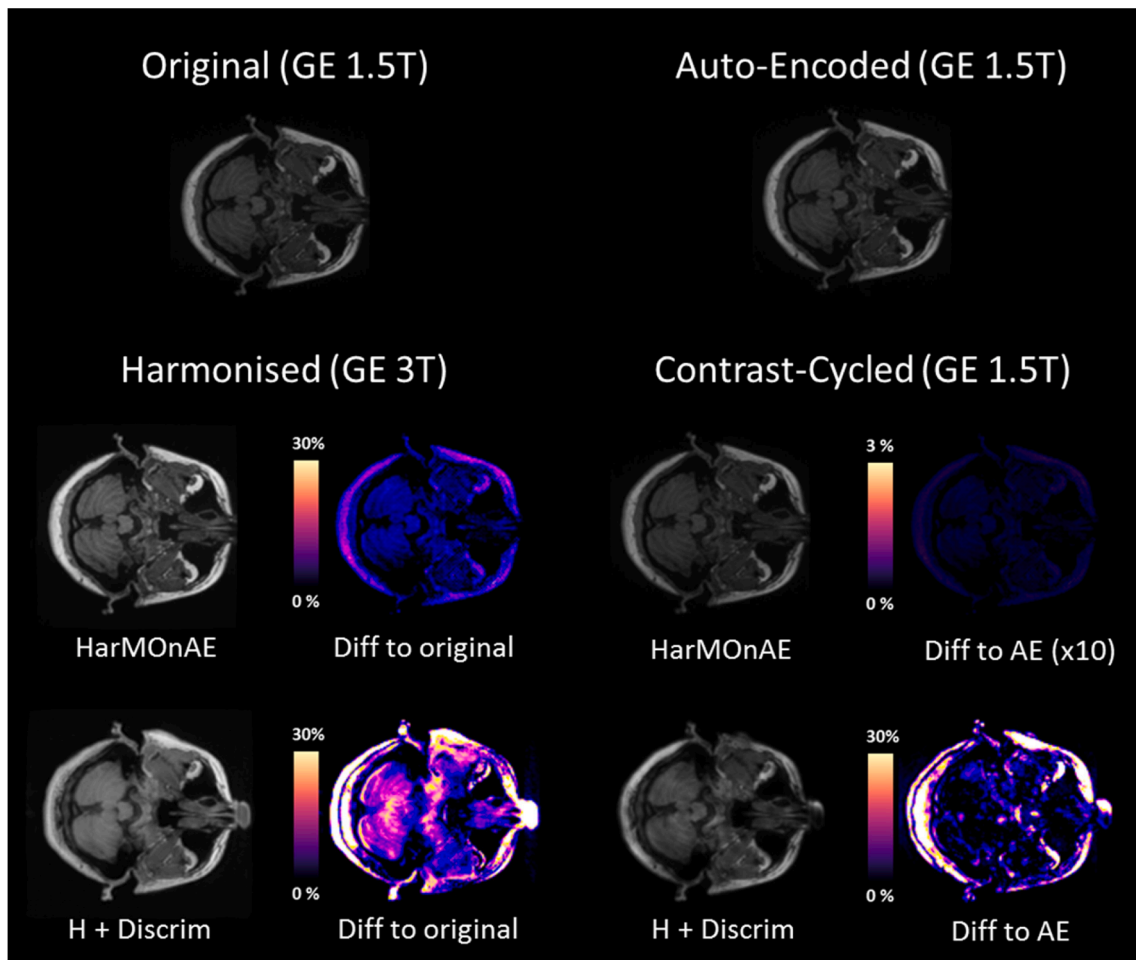


Fig. 4. Example of geometry-shift when using an image-based discriminator. Top: Input and auto-encoded images were indistinguishable. Middle: HarMONAE altered contrast without perturbing anatomical features and contrast cycling resulted in minimal difference to the auto-encoded image. Bottom: Use of an additional image discriminator on the target domain induced severe geometry-shift, altering cerebellum anatomy and hallucinating a nose. Contrast cycling failed to remove the geometric perturbations. Color maps represent percentage change in greyscale value from reference image (top).

Table 1
Harmonisation of radiomic features, measured by 2-sided KS test. Comparison of Unharmonised, WhiteStripe normalised and HarMONAE.

Harmonisation type (target contrast)	Mean no. (%) significantly different radiomic features*		
	White matter	Grey matter	
Unharmonised	49 (64.9)	48 (63.3)	
WhiteStripe	31 (40.9)	29 (39.1)	
HarMONAE	(Philips 3.0 T)	19 (24.7)	25 (33.3)
	(Siemens 1.5 T)	15 (20.2)	24 (31.6)
	(Siemens 3.0 T)	12 (16.4)	20 (26.9)
	(GE 1.5 T)	15 (19.6)	19 (24.7)
	(GE 3.0 T)	20 (26.2)	18 (24.0)

*Mean features calculated by dividing total number of significantly different features ($p < 0.05$) by number of scanners (6).

3. Results

Examples of MRI harmonisation from each scanner contrast, including zero-shot harmonisation from a previously unseen scanner (Philips 1.5 T), to all other scanner contrasts included in training are shown in Fig. 3. The extent of contrast change on harmonisation, measured by dSSIM-h varied between contrast spaces [mean 0.0156, s.d. 0.0154]. SSIM change on contrast cycling (re-harmonisation back to the original contrast space), dSSIM-c was found to be small and independent

of input and target scanner-ID [mean 0.0021, s.d. 0.0004]. No anatomical changes (geometry-shift) were observed visually with the HarMONAE model, which was consistent with these dSSIM-c values.

Using the optional image discriminator during training, dSSIM-c increased by a factor of 3, to 0.0062, indicating geometry-shift (Fig. 4) on contrast cycling and lack of content preservation.

The number of statistically indistinguishable radiomic features was increased by HarMONAE, compared to unharmonised and WhiteStripe normalised images. For white matter, 65% of features were significantly different on 2-sided KS testing in unharmonised images and 41% in WhiteStripe normalised images, whereas HarMONAE ranged between 16 and 26% depending on target contrast (Table 1 and Supplementary tables 1–6). For grey matter radiomics, 63%, 39% and 24–33% of features were significantly different in unharmonised, WhiteStripe normalised and HarMONAE images, respectively. The differences between the distributions of p-values for the 2-sided KS test and 2-tailed Welch test are illustrated in Fig. 5a and 5b. The distribution of radiomic feature values per scanner in white and grey matter for raw and harmonised images (Fig. 5c) following standardisation of radiomic feature values, showed improved consistency of 3D radiomic feature values post-harmonisation, across all six scanners. HarMONAE outperformed WhiteStripe normalisation, for white and grey matter, on both parametric (Welch’s t) and non-parametric (Kolmogorov-Smirnov) tests. Performance varied somewhat between target scanners but was consistently improved over WhiteStripe and zero-shot harmonisation

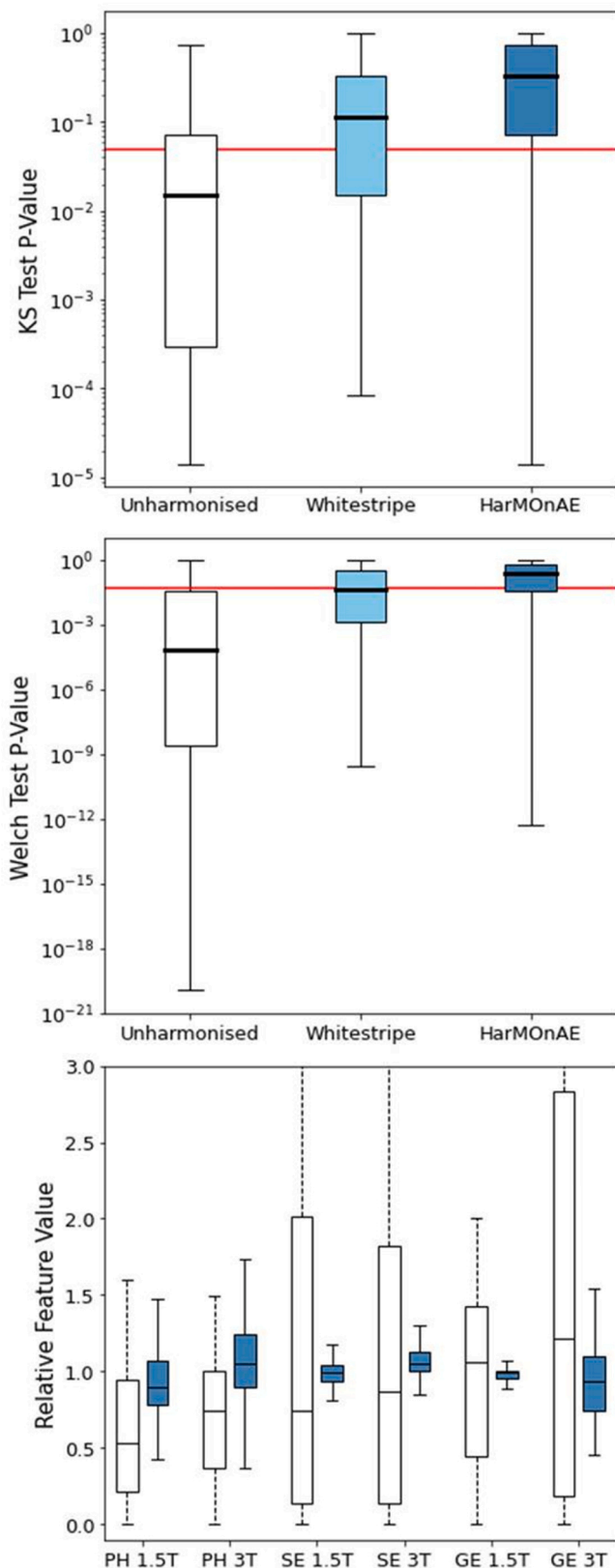


Fig. 5. a) 2-sample Kolmogorov-Smirnov test p-value distribution for 75 2nd order 3D textural radiomic features. $p > 0.05$ (red line) represented statistical indistinguishability of features; b) 2-sample Welch test p-value distribution as in a); c) relative 3D radiomic feature values for white matter, pre- (open/dashed) and post- (filled/solid) HarMONAE.

performance was comparable to that for in-training scanners (Table 1).

4. Discussion

This study demonstrated the potential of style-blind autoencoders for unpaired, multi-source MRI harmonisation. HarMONAE successfully harmonised T1-w normal brain imaging from 5 scanners from three manufacturers with two field strengths, without distorting anatomical details and further, was able to harmonise images from a 6th source scanner, unseen during training, with comparable performance.

In neuro-oncology, prognosis prediction [32], non-invasive identification of genetic and molecular changes within tumours [33], and discrimination of treatment effects and tumour progression [34] have been reported using predictive radiomic models in single-centre studies. O⁶-methylguanine methyltransferase (MGMT) gene promoter methylation determines sensitivity to alkylating chemotherapy agents including temozolomide in GBM, and can be predicted with high accuracy from textural features extracted from T2-weighted MRI [35]. In patients treated with stereotactic radiosurgery (SRS) for metastases, accurate distinction between true disease progression and radio-necrosis was difficult radiologically, but possible with radiomics-based classification [36,37].

Generalisability and clinical translation were severely limited by inconsistent MRI scanner contrast. Unsupervised harmonisation, with zero-shot capability for unseen scanners would change this situation. Previous methods have failed to meet all these criteria and may be susceptible to geometry-shift, resulting in unacceptable alteration of anatomical or physiological information. HarMONAE met these criteria, producing excellent radiomic feature harmonisation and anatomical content preservation. The network learned to identify and extract salient content information only. Hence, the trained network could perform zero-shot harmonisation as the unknown scanner-dependent contrast information was discarded, with relevance in real-world clinical scenarios, where new scanners may come online, or patient images may be sent from locations with unknown scanners. By comparing statistical properties of images via unpaired equivalence testing of radiomic feature value distributions, we showed that HarMONAE outperforms WhiteStripe, the reference normalisation method for brain imaging [6,8].

Relative change of SSIM on contrast-cycling (dSSIM-c) demonstrated that HarMONAE did not degrade image content significantly more than auto-encoding itself. dSSIM-c of < 0.0025 indicated excellent reconstruction of both image content and source-scanner contrast. Absolute SSIM has been used previously to measure harmonisation quality, using paired (travelling-subject) images. However, imperfect pairing of images renders this metric useful only for relative method comparison using the same dataset. CALAMITI[17] achieves mean absolute SSIM = 0.884 for travelling subject assessment, given un-harmonised SSIMs of 0.803–0.871, implying change of SSIM on harmonisation (dSSIM-h) of 0.01 to 0.08, dependent on scanner. Despite different datasets this was broadly comparable to our dSSIM-h values (0.005–0.054, Fig. 3).

As HarMONAE was an unsupervised method requiring only unpaired data, and could accommodate multiple scanner contrasts, it potentially also enables harmonisation of body MRI, where well-registered paired data are unavailable. The absence of an image-based discriminator in HarMONAE circumvented geometry-shift, by removing the adversarial loss term which encouraged *image* similarity to a target distribution. The experimental addition of such a discriminator to HarMONAE increased mean dSSIM-c with visibly reduced image content fidelity. By highly weighting the adversarial loss, dramatically visually degraded harmonised images were observed (Fig. 4), with ‘hallucinated’ false anatomical features, highlighting the problem of anatomical fidelity loss.

The current work had certain limitations, particularly in that the zero-shot scanner data tested was not extremely dissimilar to the training scanner data. Further work will be needed to determine the

performance of HarMONAE as unseen images deviate further from the mean contrast space of the training cohort. Having studied normal T1-w brain MRI, the impact of pathology and mobile organs in other body sites, as well as extension to other modalities, remain open questions which we intend to address in future. One challenge with pathological imaging is that cohorts associated with particular scanners may have common findings (e.g. oncology specific scanners), biasing the content distribution of the training dataset. However, whereas image-discriminator based methods might suffer from geometry-shift in such a scenario, HarMONAE is expected to be more robust to systematic anatomical variance between scanner cohorts, due to the absence of an image discriminator.

In summary, HarMONAE was found to be a simple and flexible deep-learning architecture for image harmonisation based on style-blind autoencoders, harmonising images from multiple sources including unseen scanners, without domain-adaptation or retraining. It relied on a learnt disentanglement of anatomical content and scanner-dependent contrast in the latent representation. It was shown to be immune to geometry-shift, which can plague cGAN-based approaches reliant on image-based discriminators. Our approach outperformed WhiteStripe normalisation, including for zero-shot harmonisation and showed similar relative image similarity change (dSSIM-h) to other state of the art methods. Practical, robust, and general harmonisation methods such as HarMONAE will enable quantitative deep-learning and radiomic analysis for personalised radiotherapy and oncology.

Declaration of Competing Interest

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

Acknowledgements

This work was funded by the Cancer Research UK Radiation Research Centre of Excellence at the University of Leeds: C19942/A28832.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.phro.2022.05.005>.

References

- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* 2015;278:563–77. <https://doi.org/10.1148/radiol.2015151169>.
- van Timmeren J, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—“how-to” guide and critical reflection. *Insights Imaging* 2020; 11:91. <https://doi.org/10.1186/s13244-020-00887-2>.
- Pinto dos Santos D, Dietzel M, Baessler B. A decade of radiomics research: are images really data or just patterns in the noise? *Eur Radiol* 2021;31:2–5. <https://doi.org/10.1007/s00330-020-07108-w>.
- Da-Ano R, Visvikis D, Hatt M. Harmonization strategies for multicenter radiomics investigations. *Phys Med Biol* 2020;65:24. <https://doi.org/10.1088/1361-6560/aba798>.
- Yang F, Dogan N, Stoyanova R, Ford JC. Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth. *Phys Medica* 2018;50:26–36. <https://doi.org/10.1016/j.ejmp.2018.05.017>.
- Carré A, Klausner G, Edjlali M, Lerousseau M, Briand-Diop J, Sun R, et al. Standardization of brain MR images across machines and protocols: bridging the gap for MRI-based radiomics. *Sci Rep* 2020;10:1–16. <https://doi.org/10.1038/s41598-020-69298-z>.
- Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* 2000;19:143–50. <https://doi.org/10.1109/42.836373>.
- Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage Clin* 2014;6:9–19. <https://doi.org/10.1016/j.nicl.2014.08.008>.
- Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. *Proc. SPIE 10949, Medical Imaging 2019: Image Processing*, 109493H, 2019; <https://doi.org/10.1117/12.2513089>.
- Hoebel K V., Patel JB, Beers AL, Chang K, Singh P, Brown JM, et al. Radiomics Repeatability Pitfalls in a Scan-Rescan MRI Study of Glioblastoma. *Radiol Artif Intell* 2021;3:e190199. <https://doi.org/10.1148/ryai.20210190199>.
- Orlhac F, Lecler A, Savatovski J, Goya-Outi J, Nioche C, Charbonneau F, et al. How can we combat multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol* 2021;31:2272–80. <https://doi.org/10.1007/s00330-020-07284-9>.
- Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;8:118–27. <https://doi.org/10.1093/biostatistics/kxj037>.
- 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>.
- Dewey BE, Zhao C, Reinhold JC, Carass A, Fitzgerald KC, Sotirchos ES, et al. DeepHarmony: A deep learning approach to contrast harmonization across scanner changes. *Magn Reson Imaging* 2019;64:160–70. <https://doi.org/10.1016/j.mri.2019.05.041>.
- Zhu, J-Y, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv:1703.10593 [cs.CV]* <https://doi.org/10.48550/arXiv.1703.10593>.
- Bashyam VM, Doshi J, Erus G, Srinivasan A, Abdulkadir A, Habes M, et al. Medical image harmonization using deep learning based canonical mapping: Toward robust and generalizable learning in imaging. *arXiv:2010.05355 [eess.IV]* <https://doi.org/10.48550/arXiv.2010.05355>.
- Zuo L, Dewey BE, Aaron Carass A, Yihao Liu Y, Yufan He Y, Peter A. Calabresi PA, et al. Information-Based Disentangled Representation Learning for Unsupervised MR Harmonization. In: Feragen, A., Sommer, S., Schnabel, J., Nielsen, M. (eds) *IPMI 2021. Lect Notes Comp Sci*, vol 12729. https://doi.org/10.1007/978-3-030-78191-0_27.
- Moyer D, Ver Steeg G, Tax CMW, Thompson PM. Scanner invariant representations for diffusion MRI harmonization. *Magn Reson Med* 2020;84:2174–89. <https://doi.org/10.1002/mrm.28243>.
- Huang X, Liu M, Belongie S, Kautz J. Multimodal Unsupervised Image-to-Image Translation. *arXiv:1804.04732 [cs.CV]* <https://doi.org/10.48550/arXiv.1804.04732>.
- Souza R, Lucena O, Garrafa J, Gobbi D, Saluzzi M, Appenzeller S, et al. An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *Neuroimage* 2018;170:482–94. <https://doi.org/10.1016/j.neuroimage.2017.08.021>.
- 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>.
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 2011;54:2033–44. <https://doi.org/10.1016/j.neuroimage.2010.09.025>.
- Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL *Neuroimage* 2012;62:782–90. <https://doi.org/10.1016/j.neuroimage.2011.09.015>.
- Muschelli J, Gherman A, Fortin JP, Avants B, Whitcher B, Clayden JD, et al. Neuroconductor: An R platform for medical imaging analysis. *Biostatistics* 2019; 20:218–39. <https://doi.org/10.1093/biostatistics/kxx068>.
- Muschelli J, Sweeney E, Lindquist M, Crainiceanu C. fsr: Connecting the FSL Software with R. *R J* 2015;7:163–75. <https://doi.org/10.32614/rj-2015-013>.
- Popescu V, Battaglini M, Hoogstrate WS, Verfaillie SCJ, Sluiter IC, van Schijndel RA, et al. Optimizing parameter choice for FSL-Brain Extraction Tool (BET) on 3D T1 images in multiple sclerosis. *Neuroimage* 2012;61:1484–94. <https://doi.org/10.1016/j.neuroimage.2012.03.074>.
- Zhang Y, Brady M, Smith S. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans Med Imaging* 2001;20:45–57. <https://doi.org/10.1109/42.906424>.
- Smith SM. Fast robust automated brain extraction. *Hum Brain Mapp* 2002;17: 143–55. <https://doi.org/10.1002/hbm.10062>.
- van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017;77:e104–7. <https://doi.org/10.1158/0008-5472.CAN-17-0339>.
- Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative 2016. <https://doi.org/10.11719/candat.2016.08.1>.
- Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of simpleITK *Front Neuroinform* 2013;7:1–14. <https://doi.org/10.3389/fninf.2013.00045>.
- Kickingederer P, Burth S, Wick A, Götz M, Eidel O, Schlemmer H-P, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival with Improved Performance over Established Clinical and Radiologic Risk Models. *Radiology* 2016;280:880–9. <https://doi.org/10.1148/radiol.2016160845>.
- Rathore S, Akbari H, Rozycki M, Abdullah KG, Nasrallah MP, Binder ZA, et al. Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1. *Sci Rep* 2018;8:1–12. <https://doi.org/10.1038/s41598-018-22739-2>.
- Akbari H, Rathore S, Bakas S. Quantitative image analysis and machine learning techniques for distinguishing true progression from pseudoprogression in patients with glioblastoma. *Neuro Oncol* 2018;20:191–2. <https://doi.org/10.1093/neuonc/ny148.794>.

- [35] Korfiatis P, Kline TL, Coufalova L, Lachance DH, Parney IF, Carter RE, et al. MRI texture features as biomarkers to predict MGMT methylation status in glioblastomas. *Med Phys* 2016;43:2835–44. <https://doi.org/10.1118/1.4948668>.
- [36] Peng L, Parekh V, Huang P, Lin DD, Sheikh K, Baker B, et al. Distinguishing True Progression From Radionecrosis After Stereotactic Radiation Therapy for Brain Metastases With Machine Learning and Radiomics. *Int J Radiat Oncol Biol Phys* 2018;102:1236–43. <https://doi.org/10.1016/j.ijrobp.2018.05.041>.
- [37] Forghani R. Precision Digital Oncology: Emerging Role of Radiomics-based Biomarkers and Artificial Intelligence for Advanced Imaging and Characterization of Brain Tumors. *Radiol Imaging Cancer* 2020;2:e190047.