Neuroinformatics (2024) 22:193–205
https://doi.org/10.1007/s12021-024-09655-9

RESEARCH

DeepN4: Learning N4ITK Bias Field Correction for T1-weighted Images

Praitayini Kanakaraj[1] · Tianyuan Yao[1] · Leon Y. Cai[2] · Ho Hin Lee[1] · Nancy R. Newlin[1] · Michael E. Kim[1] · Chenyu Gao[3] · Kimberly R. Pechman[4] · Derek Archer[4,5,6] · Timothy Hohman[4,5,6] · Angela Jefferson[4,5,7] · Lori L. Beason-Held[8] · Susan M. Resnick[8] · The Alzheimer's Disease Neuroimaging Initiative (ADNI) · The BIOCARD Study Team · Eleftherios Garyfallidis[9] · Adam Anderson[2,10] · Kurt G. Schilling[11] · Bennett A. Landman[1,11,3,2,10] · Daniel Moyer[1]

Accepted: 19 December 2023 / Published online: 25 March 2024

Abstract
T1-weighted (T1w) MRI has low frequency intensity artifacts due to magnetic field inhomogeneities. Removal of these biases in T1w MRI images is a critical preprocessing step to ensure spatially consistent image interpretation. N4ITK bias field correction, the current state-of-the-art, is implemented in such a way that makes it difficult to port between different pipelines and workflows, thus making it hard to reimplement and reproduce results across local, cloud, and edge platforms. Moreover, N4ITK is opaque to optimization before and after its application, meaning that methodological development must work around the inhomogeneity correction step. Given the importance of bias fields correction in structural preprocessing and flexible implementation, we pursue a deep learning approximation / reinterpretation of the N4ITK bias fields correction to create a method which is portable, flexible, and fully differentiable. In this paper, we trained a deep learning network "DeepN4" on eight independent cohorts from 72 different scanners and age ranges with N4ITK-corrected T1w MRI and bias field for supervision in log space. We found that we can closely approximate N4ITK bias fields correction with naïve networks. We evaluate the peak signal to noise ratio (PSNR) in test dataset against the N4ITK corrected images. The median PSNR of corrected images between N4ITK and DeepN4 was 47.96 dB. In addition, we assess the DeepN4 model on eight additional external datasets and show the generalizability of the approach. This study establishes that incompatible N4ITK preprocessing steps can be closely approximated by naïve deep neural networks, facilitating more flexibility. All code and models are released at https://github.com/MASILab/DeepN4.

Keywords T1-weighted images · Bias field correction · Inhomogeneity · N4ITK · 3D U-Net

Introduction

Structural magnetic resonance imaging (MRI) highlights differences in tissue contrast based on the longitudinal relaxation time of hydrogen protons, making structural images suitable for delineating anatomical structures, abnormalities, and tissue types (Damadian, 1971; Huo et al., 2019; Johnson, 2016). Clinically, structural images are frequently utilized as a reference to monitor the progression of disease and the efficacy of treatments for neurological disorders (Schiffmann & Knaap, 2009). However, structural MRI suffers from intensity inhomogeneity artifacts appearing as a low frequency spatial intensity changes ("bias field") that occur in part due to imperfections in the magnetic fields (Vovk et al., 2007). Correcting for these low frequency artifacts is a necessary preprocessing step in image processing. This helps avoid erroneous results in downstream analyses such as image segmentation, registration, texture analysis, and tissue classification (Gispert et al., 2004; Xu et al., 2022).

There are several frameworks for eliminating the spatially varying bias fields (Song et al., 2017; Vovk et al., 2007). In general, they follow two steps: (1) estimation of the bias field and (2) computing the corrected debiased image. Traditional correction methods can be classified as prospective (Axel et al., 1987; Mihara et al., 1998; Narayana et al., 1988; Simmons et al., 1994; Vovk et al., 2007) or retrospective approaches (Song et al., 2017), with retrospective approaches gaining dominance due to their generalizability, efficiency, and fewer assumptions about the acquisition process (Vovk et al., 2007). Retrospective approaches use acquired images containing anatomical and intensity inhomogeneity information, along with prior knowledge of the

Extended author information available on the last page of the article

imaging object. Retrospective approaches are further divided into filtering (Brinkmann et al., 1998), surface fitting, segmentation (Ashburner & Friston, 2005), and histogram (Sled et al., 1998) stages. Of particular significance to our work is the local histogram-based N3 method, which iteratively estimates the smooth multiplicative field by maximizing the high frequency component of the image intensity distribution (Sled et al., 1998). Our method estimates this spatially varying multiplicative low frequency component with a deep learning network based on the principles of the N3 approach. A refinement of N3 is N4ITK which estimates the field at each iteration using the results of the previous iteration along with a *B*-spline approximation (Tustison et al., 2010), and is widely accepted due to the effectiveness and efficiency of the approach. N4ITK, provided by ITK, is integrated into various neuroimaging analysis tools, including SimpleITK (Yaniv et al., 2018), ANTs (Avants et al., 2009), FreeSurfer (Fischl, 2012), fmriPrep (Esteban et al., 2019), NiPype (Gorgolewski et al., 2011), NeuroNorm (Payares-Garcia & Mateu, 2023), and MRtrix (Tournier et al., 2012). Thus, in the past decade N4ITK has been recognized as the state-of-the-art (SOTA) approach and so we use N4ITK as a starting point for our own model design.

Configuring ITK can be challenging, especially with the unfamiliarity with CMake, and requires careful consideration of compatibility across ITK, operating across ITK, operating systems, compiler versions, and hardware platforms. While ITK is integrated into various neuroimaging analysis tools and is accessible to the public, the compilation process entails the installation of accompanying libraries and software packages. However, in cases where sole interest lies in N4 correction, these additional libraries become superfluous. For example, SimpleITK (Yaniv et al., 2018) encompass Elastix, GTest, Luc, PCRE2, SWIG, and Sphinx, whereas ANTs (Avants et al., 2009) extends to comprise Cppcheck, KWStyle, Slier, Uncrustify, and VTK, alongside ITK. This introduces complexities in terms of licensing and integration (Fig. 1a).

The use of differentiable approaches for end-to-end learning pipelines is an actively evolving area of research (Agrawal et al., 2019; Pineda et al., 2022). Downstream tasks such as segmentation are often performed after inhomogeneity correction. However, optimization of parameters before inhomogeneity correction for outcomes measured afterwards is not easily done; N4ITK is opaque to gradient based optimization. Our paper addresses this problem by constructing an intermediate inhomogeneity correction step that is differentiable to optimize models before and after inhomogeneity correction. This enables the use of loss function based on the characteristics after inhomogeneity correction (Fig. 1b).

Recently, deep learning models have achieved SOTA results in medical image processing tasks. Researchers have proposed deep learning-based methods for bias field correction of MR images (Chuang et al., 2022; Gaillochet et al., 2020; Goldfryd et al., 2021; Simkó et al., 2022; Sridhara et al., 2021; Wan et al., 2019; Xu et al., 2022). Of these, there are two open source sharable models that fit our criteria, and are a feedforward CNN (Simkó et al., 2022; Xu et al., 2022) and an autoencoder (Sridhara et al., 2021). Simkó et al. (2022) used implicit training on convolution neural network (CNNs) for bias field correction (Simkó et al., 2022). Sridhara et al. (2021) used an autoencoder based deep learning architecture to predict bias field that outperforms the conventional N4ITK approach (Sridhara et al., 2021).

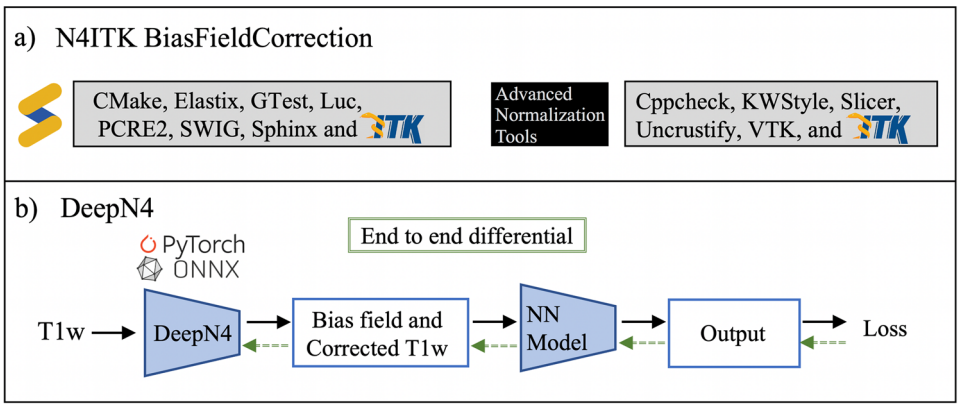

Fig. 1 T1w MRI scans show spatial variations of image intensities, known as bias field effects, caused by the field inhomogeneity. **a** The state-of-the-art framework that models the bias field has external dependencies that complicate integration into imaging pipelines. **b** To address this, we propose DeepN4, a deep learning differentiable end-to end-model that utilizes the PyTorch python library; ONNX allows conversion across a deep learning framework. Our model allows loss function based on post-inhomogeneity correction. Our approach allows for the loss function based on post-inhomogeneity correction. Abbreviation: NN, Neural network

Table 1 Datasets used for training, validation, and testing of DeepN4

Dataset	Subjects	Sessions	T1w Images	Vendor (Scanner)	Field Strength	Resolution (mm)
ADNI (Jack et al., 2008)	799	1–5	1905	Siemens, GE, and Philips (61)	3 T	1 iso 1.2×1 x 1 1.2×1.054 x 1.054 2×1 x 1
BLSA (Shock, 1984)	1151	1–8	2869	Philips (3)	3 T	1.2×1x1
OASIS-3 (LaMontagne et al., 2019)	992	1–6	2452	Siemens (2)	1.5 T, 3 T	1×1 x 1 1.2×1 x 1
NACC	273	1–2	288	Philips (1)	3 T	1.2×1 x 1
BIOCARD (Sacktor et al., 2017)	212	1–4	508	Philips (1)	3 T	1.2×1 x 1
HCP YA (Essen et al., 2013)	1112	1	1112	Siemens (1)	3 T	0.7 iso
HCP Aging (Harms et al., 2018)	664	1	664	Siemens (1)	3 T	0.8 iso
HCP Dev (Harms et al., 2018)	626	1	626	Siemens (1)	3 T	0.8 iso
Learning (Training + Validation)	-	-	9382 (8340 + 1042)	-	-	2 iso
Testing (Withheld)	-	-	1042	-	-	2 iso

iso isotropic, - not applicable

In the present work we provide a simple feedforward deep learning network for estimating the multiplicative field, trained with a direct (non-adversarial) loss term. We show parity or improvement on other open source models, and at a sufficiently high fidelity that further innovation and complexity seem unnecessary (Xu et al., 2022). Thus, we propose a differentiable approach that estimates the smooth bias field while facilitating flexible and portable implementations of the SOTA N4ITK bias field correction from raw T1-weighted (T1w) MRI without complexity (Fig. 1). The model is trained on a large repository of T1w images (Table 1) and then validated on eight external datasets (Table 2) to understand model performance on how the model estimated the low spatial frequency fields from high spatial frequency T1w MRI. Finally, we release and open source all model weights and inference scripts, allowing DeepN4 to be seamlessly integrated into other workflows.

Methods and Materials

We model each bias field as a multiplicative field (Tustison et al., 2010). Rewriting Eq. (1) from N3 paper (Sled et al., 1998), we have

$$a(r) = u(r)b(r) \tag{1}$$

where a is the acquired image, u is the corrected image, b is the bias field, and r is the voxel position of the images. We assume both $u(r)$ and $b(r)$ are greater than zero at all points r. Applying a logarithmic transformation and

Table 2 Datasets used for external validation of DeepN4

Dataset	Subjects	Sessions	T1w Images	Vendor (Scanner)	Field Strength	Resolution (mm)
VMAP (Jefferson et al., 2016)	327	1–4	1074	Philips (1)	3 T	1 iso
KIRBY (Landman et al., 2011)	5	5	5	Philips (1)	3 T	1.2×1 x 1
SCA2 (Mascalchi et al., 2018) (ds001378)	5	5	5	Philips (1)	1.5 T	1 iso
IXI Hammersmith (IXI Dataset - Information eXtraction from images, 2020)	5	5	5	Philips (1)	3 T	1.2×1 x 1
IXI Guys (IXI Dataset - Information eXtraction from images, 2020)	5	5	5	Philips (1)	1.5 T	1.2×1 x 1
MASSIVE (Froeling et al., 2017)	1	5	5	Philips (1)	3 T	1 iso
BRATS (Bakas et al., 2017)	5	5	5	Siemens, GE, and Philips (5)	1.5 T,3 T	1 iso
GLAUCAMO (Miller et al., 2019) (ds001743)	5	5	5	GE (1)	3 T	1 iso

iso isotropic

solving for corrected image, Eq. (1) can be rewritten as $\log(u(r)) = \log(a(r)) - \log(b(r))$. We aim to estimate $\log(b(r))$ using a ral network. The following sections describe how our computing object was constructed, implemented, and trained. This is followed by an overview of the data used.

DeepN4 Architecture

We parameterize the log-transformed bias field by a neural network, i.e.:

$$\log b'(r) = f(a(r)) \tag{2}$$

where function f is *DeepN4*, a generic 3D U-Net network and b' is the predicted bias field image. DeepN4 is a 3D U-Net framework based on the traditional architecture proposed by Ronneberger et al. (2015). The modification made in a well-validated three-dimensional image synthesis network (Schilling et al., 2020) was adapted in this network. It uses Leaky ReLU as activation function and instance batch normalization. The expanding path consists of corresponding transpose convolution layers to regain the spatial dimension of the input image. The convolution and devolution layers' kernels are of size $3 \times 3 \times 3$. The feature maps from the paths are concatenated via skip connections to retain both high-level and low-level features and enhance the accuracy of the model output.

Upon obtaining the predicted bias field $e^{b'(r)}$, we apply smoothing on the predicted bias field $e^{b'(r)}$ (after the voxel-model). This process aims to mitigate high-frequency noise and irregularities present in the predicted bias field, it aids in achieving a consistent and refined correction across the image (Tustison et al., 2010). There are two possible variations of smoothing. The first is a parameterized reconstruction employing *B*-splines to impose smoothness using *B*-spline functions. *B*-spline functions have local support and are numerically stable, making them a powerful tool for smoothing. The *B*-spline approximation from N4ITK is a uniform multivariate *B*-spline object of arbitrary order with resolution increasing at each successive level in the iteration process (Tustison et al., 2010). We perform slicewise smoothing using the *B*-spline model in ANTsPy (Tustison et al., 2021), which is a wrapper of the ITK *B*-spline approximation from ITKN4. *B*-spline was configured with a spline order of 3 and five fitting levels. Alternatively, a second option is approximating the smoothing with an isotropic. We choose a filter (kernel size $19 \times 19 \times 19$ voxels) with standard deviation of 3 voxels which blur the bias field slightly (Simkó et al., 2022). Thus, the experiments can be summarized as DeepN4 NS, DeepN4 B, and DeepN4 G for no smoothing, *B*-spline, or Gaussian smoothing based on the smoothing approach after the U-Net architecture.

The loss function for DeepN4 is defined as $\mathcal{L} = \mathcal{L}_a + \mathcal{L}_b$ where $\mathcal{L}_a$ is L2 loss function between the predicted and the ground truth bias fields, and $\mathcal{L}_b$ is L2 loss between the corrected image from the predicted bias field and the ground truth corrected image. That is:

$$\mathcal{L}_a = \frac{1}{N}\sum_{r=1}^{N}\left(e^{\log b'(r)} - b'(r)\right)^2 \tag{3}$$

$$\mathcal{L}_b = \frac{1}{N}\sum_{r=1}^{N}\left(e^{\log a(r) - \log b'(r)} - u(r)\right)^2 \tag{4}$$

In Eqns. (3) and (4), N denotes the total number of masked voxels in image. The log predicted bias field $b(r)$ is only computed within a brain mask to avoid background intensities of zero. The acquired image is divided by the smoothed bias field $e^{b'(r)}$ to obtain the corrected image $u'(r)$ (Fig. 2).

Furthermore, we compared our proposed DeepN4 models to other open source deep learning based bias field correction methods and Statistical Parametric Mapping (SPM) bias field correction (denoted here as SPMbfc). Specifically, we included the autoencoder model trained by Sridhara et al. using synthetic data from the HCP dataset, which is accessible at https://github.com/Shashank-95/Bias-Field-correction-in-3D-MRIs (Sridhara et al., 2021) and the CNN model implicitly trained on images from BrainWeb by Simkó et al., which is available at https://github.com/attilasimko/bfc (Simkó et al., 2022). The default SPM configuration with 60 mm full width at half maximum and a regularization of 0.001 was used for SPMbfc.

Training Protocol

For the training process, the network is optimized to minimize the loss function in Section "Training Protocol" using the Adam optimizer (Kingma & Ba, 2014) with a learning rate of 0.0001. We trained the model on a NVIDIA-Quadro RTX 5000 GPU with 16 GB of memory. The model was trained on the training cohort with the ground truth consisting of the bias field and the corrected T1w image from N4ITK (Fig. 2). The trained model that performed the best on the validation cohort was chosen to evaluate the test cohort.

Data Overview

The objective of this study was to train a neural network model with diverse datasets obtained from different scanners with different resolutions, and different field strengths to create a robust and generalizable model approximating N4ITK, allowing the model to effectively handle variations in imaging protocols and produce

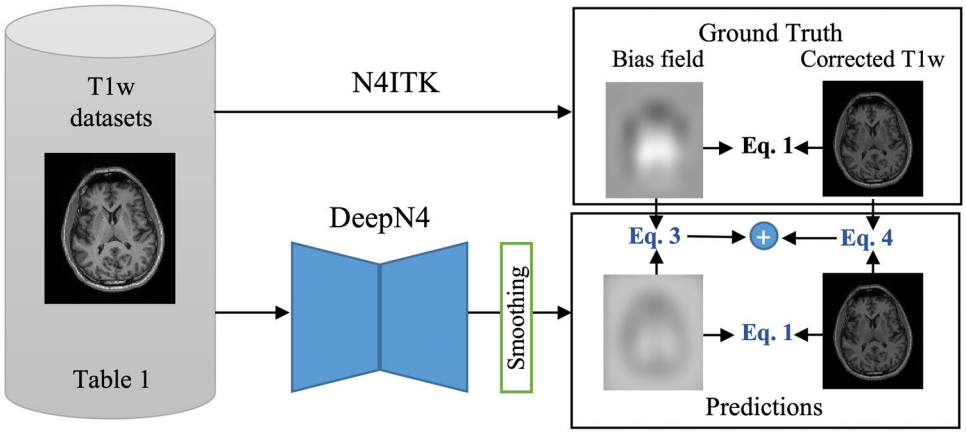

Fig. 2 N4ITK was processed on the large-scale datasets in Table 1 to generate the ground truth bias field and corrected T1w images. All the T1w images in Table 1 were fed into the DeepN4 which outputs the log of predicted bias field. Smoothing is performed on predicted bias field from which the corrected image is obtained. The loss is minimized between the ground truth bias field and corrected T1w image with the predicted bias field and computed corrected T1w image using Eqs. (3) and (4)

accurate results. Consequently, we use de-identified data from eight distinct datasets as listed in Table 1 each with varying subjects, sessions, and scanner vendors. The ADNI cohort (https://adni.loni.usc.edu) began in 2003 as a public–private partnership, led by Principal Investigator, Michael W. Weiner, MD (Jack et al., 2008). The NACC cohort began in 1999 and is comprised of dozens of Alzheimer's Disease Research Centers that collect multimodal AD data (Beekly et al., 2004). The overall intention of the NACC cohort is to collate a large database of standardized clinical/neuropathological data (Beekly et al., 2007; Besser et al., 2018; Weintraub et al., 2009, 2018). There was a total of 10,424 T1w images that we randomly split into 90/5/5% as training, validation, and testing cohorts respectively. These were down-sampled to 2×2×2 mm and padded such that the image dimensions were 128×128×128 across all scans. The scans are normalized with min–max normalization, $(X - X_{min})/(X_{max} - X_{min})$ where X_{max} is the 99th intensity percentile of image X and X_{min} is 0. The normalized image values are clipped with an interval of [0, 1]. These down-sampled, padded, and normalized images are then used as input to the network discussed in this next section. For external validation, we used an additional set of eight external datasets (from sites distinct to those in Table 1), as outlined in Table 2. Seven of these eight datasets are publicly accessible. We evaluated the performance on DeepN4 NS, DeepN4 B, DeepN4 G, Sridhara et al. (2021), and Simkó et al. (2022) models on withheld and external test dataset in Table 1 and Table 2 respectively.

Results

Quantitative Performance

Here, we demonstrate the quantitative evaluation of our model's performance. We began by validating in a simulated environment, computing the peak signal-to-noise ratio (PSNR) for corrected T1w images followed by a comparison with existing models, then we evaluate against SPM bias field correction. Finally, we employed contrast-to-noise ratio (CNR) analysis, to emphasis the model's effectiveness in delineating grey and white matter boundaries across varied datasets.

Validation in Simulation

To ensure robust evaluation of the proposed model, we simulated a controlled environment by introducing a known bias field to five "truth" T1w images. We then compute the PSNR between the truth T1w images and the bias field corrected T1w images from N4ITK, SPMbfc, and DeepN4 models (Fig. 3). We find the median PSNR for images generated by the DeepN4 NS model was 42.56 dB, for DeepN4 B model was 42.67 dB, and for DeepN4 G was 42.79 dB,

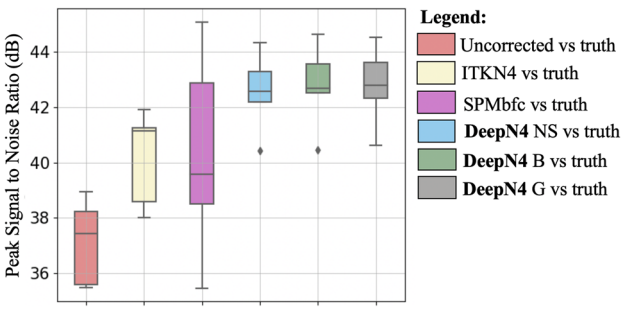

Fig. 3 In simulation, the DeepN4 models performs with PSNR of 42 dB. DeepN4 NS = DeepN4 with no smoothing, DeepN4 G = DeepN4 with Gaussian smoothing, and DeepN4 B = DeepN4 with *B*-spline smoothing

which is a 5 dB increase from uncorrected T1w and 1 dB increase from the SOTA ITKN4. This suggest that the DeepN4 approach was slightly more robust at inhomogeneity correction than ITKN4 and SPMbfc.

Comparison Against Existing Models

To evaluate the performance of the proposed model, we computed PSNR between N4ITK and DeepN4 corrected T1w images for (a) withheld test dataset of 1042 subjects and (b) external test dataset of 1074 imaging sessions (VMAP dataset), as shown in Fig. 4. In the withheld test set as in Table 1, we find the median PSNR for images generated by the DeepN4 NS model was 48.96 dB, for DeepN4 B model was 49.38 dB, and for DeepN4 G was 49.23 dB (Fig. 4a). For the subjects in VMAP (external) dataset as in Table 2, we find that the median PSNR for the DeepN4 NS model was 42.71 dB, for DeepN4 B was 42.87 dB, and DeepN4 G was 43.43 dB (Fig. 4b). We observe that the DeepN4 models outperforms the existing Sridhara et al. (2021) and Simkó et al. (2022) with notable increase in the median PSNR of 23.21 dB and 3.71 dB respectively in withheld dataset. This indicates that DeepN4 model with access to a large and diverse dataset was able to generalize from the training set while the existing models with limited numbers of training data were not generalizable. Please note that the results from the models were mean shifted to uncorrected T1w image. This adjustment was made to compensate for the global intensity scaling in N4ITK, as the rescale option, which prevents intensity drift at each iteration, was not enabled on by default.

Additionally, we find that accuracy with Gaussian approximation (DeepN4 G) is 0.5 dB higher than *B*-spline regularization (DeepN4 B) on the external datasets and equivalent on the withheld data.. This suggests the straightforward Gaussian approximation can serve as viable substitute for more resource-intensive *B*-spline regularization, which requires ANTsPy (Tustison et al., 2021) package. However, one potential factor contributing to this difference could be the higher number of randomly selected points used in Gaussian smoothing compared to *B*-spline.

The p-value between DeepN4 models was less than 0.0001 with a Bonferroni correction indicating that there was a statistically significant difference between the withheld dataset and the external dataset, reflecting the very large sample size and statistical power to detect small effects. We computed Cohens d, and the effect size was <0.2 (considered small) between the models. This suggest DeepN4 models performed with similar effectiveness.

Comparison Against SPM Bias Field Correction

Figure 5 shows PSNR as Fig. 4, but with the corrected T1w from SPM. We observe a median PSNR for images generated by the DeepN4 NS model was 40.56 dB, for DeepN4 B model was 40.63 dB, and for DeepN4 G was 40.63 dB, outperforming the existing models (Fig. 5a). While the PSNR between SPMbfc T1w and predicted DeepN4 T1w is high, it is approximately 9 dB less compared to the PSNR between ITKN4 T1w and predicted DeepN4 T1w. This indicates that DeepN4 aligns more closely with ITKN4 approach, and it does not capture all the aspects of SPMbfc. For external VMAP test dataset, we find that the median PSNR values 42.70 dB, 42.74 dB, and 43.07 dB for DeepN4 NS, DeepN4 B, and DeepN4 G respectively align with the findings in

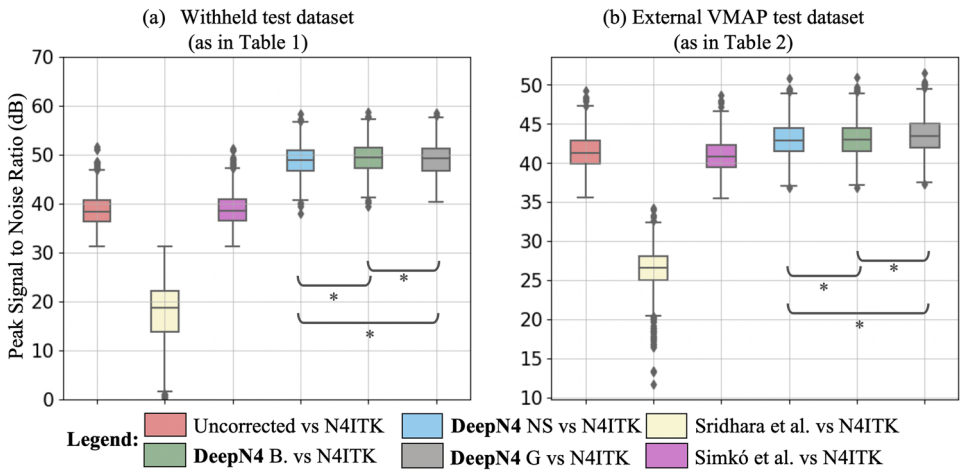

Fig. 4 For both (**a**) and (**b**) DeepN4 models outperform existing models, and the reconstructed image is similar to state-of-the-art N4ITK. Higher PSNR indicates that reconstructed images from DeepN4 models are closer to N4ITK. The observed difference in DeepN4 B and DeepN4 G is effectively the negligible. DeepN4 NS = DeepN4 with no smoothing, DeepN4 G = DeepN4 with Gaussian smoothing, DeepN4 B = DeepN4 with *B*-spline smoothing, and * $p<0.0001$ (Wilcoxon sign rank test with Bonferroni correction)

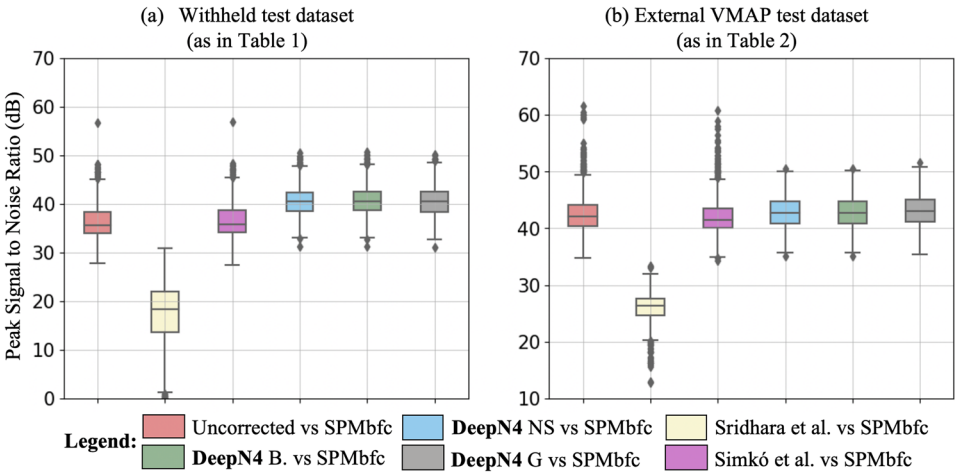

Fig. 5 For both (**a**) and (**b**) DeepN4 models outperform existing models, and the reconstructed image is similar to SPMbfc. Higher PSNR indicates that reconstructed images from DeepN4 models are closer to SPMbfc. SPMbfc = SPM bias field correction, DeepN4 NS = DeepN4 with no smoothing, DeepN4 G = DeepN4 with Gaussian smoothing, DeepN4 NS = DeepN4 with *B*-spline smoothing, and DeepN4 B = DeepN4 with *B*-spline smoothing

Fig. 4b. As the PSNR for uncorrected T1w and SPMbfc is at 41.89 dB, it indicates a lower impact of intensity inhomogeneity. In such cases where the bias field effects are minimal, the model outputs are similar to ITKN4 and SPMbfc.

Evaluation Using Contrast-to-Noise Ratio

The CNR was computed for uncorrected T1w and the corrected ITKN4, SPMbfc, and DeepN4 T1w images for withheld test dataset (Fig. 6a) and external test dataset (Fig. 6b). The CNR calculation was based on the contrast between the white matter (WM) and grey matter (GM) regions obtained from SLANT segmentation, while the noise was determined from the pooled standard derivation across these regions. It is important to note that while this method might overestimate the noise, it remains a relevant measure for assessing CNR, particularly in terms of its ability to delineate the boundary between GM and WM. This approach was preferred over using background noise due to the dataset's diverse origin across multiple sites and acquisition protocols. The noise in such datasets might lack homogeneity in the scan and might not accurately reflect the noise characteristics present in the WM and GM regions.

In Fig. 6a, we find the T1w images from SPMbfc has a median CNR of 2.08 and 0.04 higher than that of DeepN4 G. This suggest that SPMbfc is slightly better in differentiating between the two tissue types. The CNR of the T1w images from DeepN4 models are similar to that of ITKN4. However, for the external dataset, we find N4ITK corrected T1w images have the highest CNR at 0.84 and the CNR reduces by 0.06 for the DeepN4 predicted T1w images as shown in Fig. 6b.

Qualitative Performance

Figure 7 shows the absolute percent error between the truth T1w image and DeepN4 G predicted T1w in the simulation experiment from Fig. 3. From a visual perspective, the

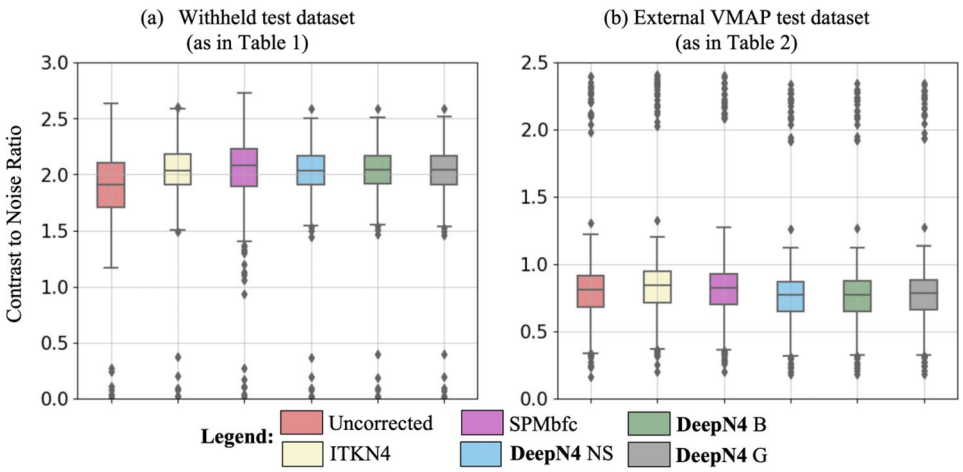

Fig. 6 CNR in SPMbfc and ITKN4 highest for withheld and external test datasets. Negligible difference in DeepN4 T1w CNR from the SOTA ITKN4 and SPMbfc. Higher CNR denotes clearer distinction between the tissue types (here, white matter and gray matter). There are 0.05% of scans with a poor CNR. SPMbfc = SPM bias field correction, DeepN4 NS = DeepN4 with no smoothing, DeepN4 G = DeepN4 with Gaussian smoothing, and DeepN4 B = DeepN4 with *B*-spline smoothing

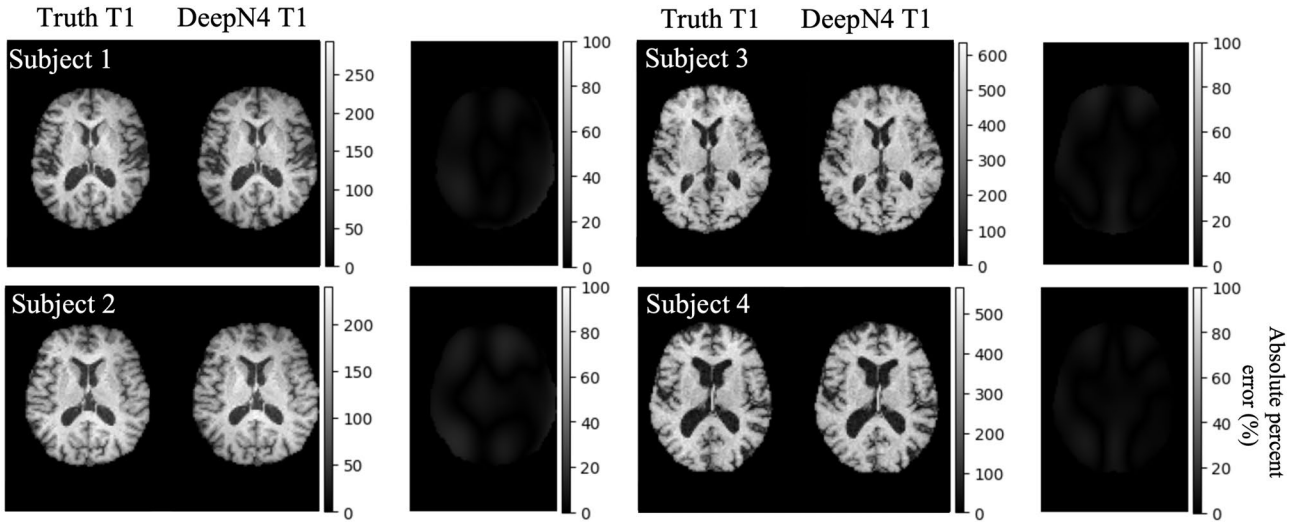

Fig. 7 In simulation, the absolute percent error of truth T1w image to which the bias was introduced and corrected and the DeepN4 G T1 is approximately 20%

DeepN4 corrected T1w strongly resembles the truth, indicating that the approach effectively removes the artificially induced bias field.

Figure 8 shows the visualization of the results from 90th, 50th and 10th percentile images with respect to the DeepN4 G results in the external VMAP dataset (Fig. 4(b) gray) along with intensity profiles for a selected slice. We observe visually (1) noticeable inhomogeneity correction from T1w to N4ITK and (2) DeepN4 corrected T1w are similar to N4ITK T1w images. The intensity profiles of the slices highlighted in blue and orange from uncorrected T1w and DeepN4 corrected T1w respectively, indicate reduction in intensity non-uniformity. Our proposed method is effective at reducing the inhomogeneity.

To demonstrate generalizability, we apply our model across external independent datasets in Table 2. Here, we

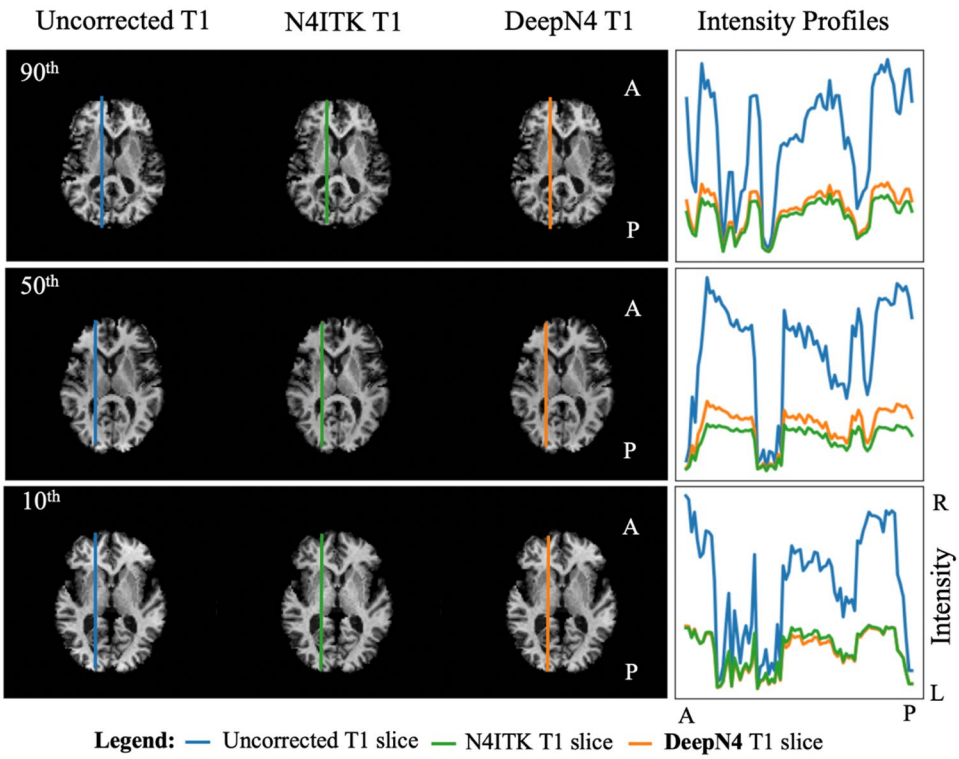

Fig. 8 90th, 50th, and 10th percentile sample are taken from DeepN4 G results in external VMAP dataset. Lower curvature between the intensity along a slice from uncorrected T1w (blue line), N4ITK corrected T1w (green line), and DeepN4 corrected T1w (orange line) denotes more uniformity. The intensity distribution along the slice in DeepN4 and N4ITK have no significant variation in performance across the 90th, 50th, and 10th percentiles sample. A = Anterior, P = Posterior

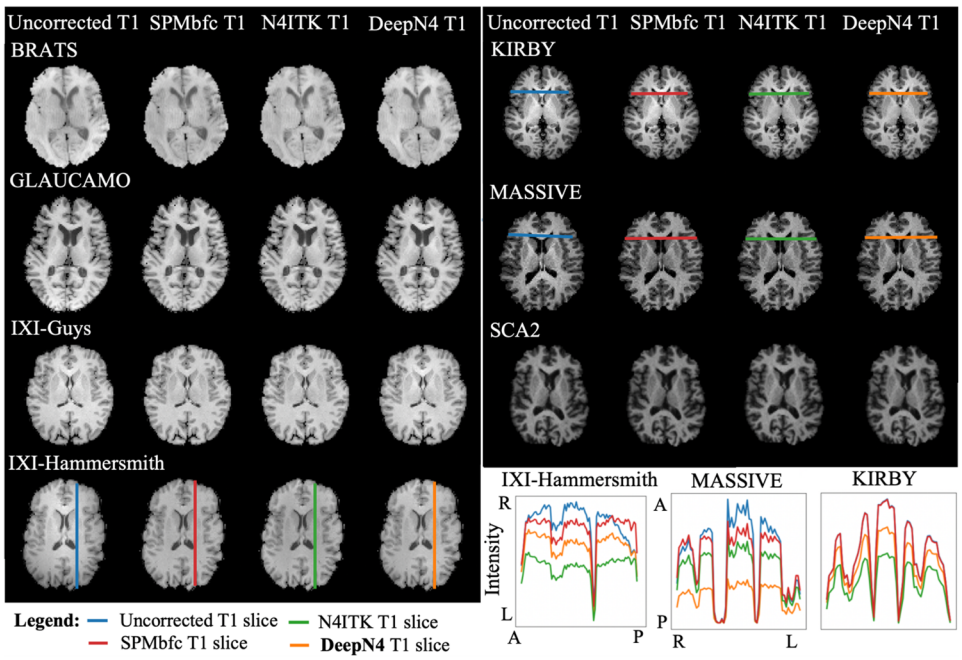

Fig. 9 Here, we show DeepN4 G results plotted against ITKN4, SPMbfc, and the original T1w. DeepN4 results are similar to the ground truth N4ITK (SOTA, but neither easily accessible nor differentiable). Less curvature in the intensity of the selected slices in DeepN4 T1w (orange line) when compared to the uncorrected T1w slice (blue line) is more homogeneous. A = Anterior, P = Posterior

show the resulting images from the sample which had median PSNR value between DeepN4 G and N4ITK and SPMbfc in each of the external dataset (Fig. 9). In all cases, we find the DeepN4 corrected images are similar to the SOTA N4ITK. This suggests the model is well generalizable to images from different cohorts with different characteristics. Note that for Figs. 4 and 5 the DeepN4 results are shown with Gaussian smoothing since the results in Fig. 3 show that performance with Gaussian smoothing and B-spline smoothing are essentially identical.

Conclusion

Adapting the state-of-the-art N4ITK bias field correction in model pipelines is challenging due to its intricate dependency stack. It is not possible to have end-to-end differentiable segmentation models using ITKN4. In this work, we address these concerns by training DeepN4, a generic 3D U-Net with loss functions based on the principals of N4ITK. Thus, we make inhomogeneity correction transparent and amenable to optimization.

Although researchers have proposed various novel deep learning frameworks for bias field correction of MR images, such as BiasNet (Xu et al., 2022), implicit training on CNNs (Simkó et al., 2022), reconstruction algorithms (Gaillochet et al., 2020), and deep learning networks based on generative adversarial net (Chuang et al., 2022), these approaches have lacked flexibility and generalizability. The proposed simple technique has shown the feasibility of estimating and eliminating low frequency inhomogeneities on to rigid high frequency anatomical structures with naïve 3D networks. Our experiments show (1) similar performance to N4ITK both qualitatively and quantitatively and (2) consistent performance on multiple, independent, unrelated external datasets, indicating the generalizability of our model. Thus, our model is easily understandable, efficient in performance, and readily available to incorporate into any existing pipeline with the inference function.

Deep learning models have employed *B*-spline for data augmentation (Chen et al., 2020). A potential direction for future research is to explore the integration of a *B*-spline layer within the neural network and having an iterative network approach.

The pretrained pipeline is available in the form of a container along with source code and can be accessed by following the instructions at https://github.com/MASILab/DeepN4.

Acknowledgements This work was supported by the National Institutes of Health under award numbers R01EB017230, 1K01EB032898, K01-AG073584 and T32GM007347, and in part by the National Center for Research Resources, Grant UL1 RR024975-01, UL1-TR000445 and UL1-TR002243 (Vanderbilt Clinical Translational Science Award), S10-OD023680 (Vanderbilt's High-Performance Computer Cluster for Biomedical Research) and U24-AG074855. The Vanderbilt Institute for Clinical and Translational Research (VICTR) is funded by the National Center for Advancing Translational Sciences (NCATS) Clinical Translational Science Award (CTSA) Program, Award Number 5UL1TR002243-03.

We acknowledge the data provided by several initiatives: ADNI: Data collection and sharing for ADNI were supported by National Institutes of Health Grant U01-AG024904 and Department of Defense (award number W81XWH-12-2-0012). ADNI is also funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug

Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

BLSA: The BLSA is supported by the Intramural Research Program, National Institute on Aging, NIH.

NACC: The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Ann McKee, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

VMAP: Study data were obtained from the Vanderbilt Memory and Aging Project (VMAP). VMAP data were collected by Vanderbilt Memory and Alzheimer's Center Investigators at Vanderbilt University Medical Center. This work was supported by NIA grants R01-AG034962 (PI: Jefferson), R01-AG056534 (PI: Jefferson), R01-AG062826 (PI: Gifford), and Alzheimer's Association IIRG-08-88733 (PI: Jefferson).

BIOCARD: The BIOCARD study is supported by a grant from the National Institute on Aging (NIA): U19-AG03365. The BIOCARD Study consists of 7 Cores and 2 projects with the following members: (1) The Administrative Core (Marilyn Albert, Corinne Pettigrew, Barbara Rodzon); (2) the Clinical Core (Marilyn Albert, Anja Soldan, Rebecca Gottesman, Corinne Pettigrew, Leonie Farrington, Maura Grega, Gay Rudow, Rostislav Brichko, Scott Rudow, Jules Giles, Ned Sacktor); (3) the Imaging Core (Michael Miller, Susumu Mori, Anthony Kolasny, Hanzhang Lu, Kenichi Oishi, Tilak Ratnanather, Peter vanZijl, Laurent Younes); (4) the Biospecimen Core (Abhay Moghekar, Jacqueline Darrow, Alexandria Lewis, Richard O'Brien); (5) the Informatics Core (Roberta Scherer, Ann Ervin, David Shade, Jennifer Jones, Hamadou Coulibaly, Kathy Moser, Courtney Potter); the (6) Biostatistics Core (Mei-Cheng Wang, Yuxin Zhu, Jiangxia Wang); (7) the Neuropathology Core (Juan Troncoso, David Nauen, Olga Pletnikova, Karen Fisher); (8) Project 1 (Paul Worley, Jeremy Walston, Mei-Fang Xiao), and (9) Project 2 (Mei-Cheng Wang, Yifei Sun, Yanxun Xu.

OASIS-3: Data were provided in part by OASIS for the OASIS-3 cohort: Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly. HCP: Data were provided [in part] by the Human Connectome Project, WU- Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) – Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

The BIOCARD Study Team – Data used in preparation of this article were derived from BIOCARD study data, supported by grant U19 –AG033655 from the National Institute on Aging. The BIOCARD study team did not participate in the analysis or writing of this report, however, they contributed to the design and implementation of the study. A listing of BIOCARD investigators may be accessed at: https://www.biocard-se.org/public/Core%20Groups.html.

Author Contributions Conceptualization: Praitayini Kanakaraj, Eleftherios Garyfallidis, Bennett A. Landman and Daniel Moyer. Data curation: Praitayini Kanakaraj, Nancy R. Newlin, Michael E. Kim, Chenyu Gao, Kimberly R. Pechman, Derek Archer, Timothy Hohman, Angela Jefferson, Lori L. Beason-Held, Susan M. Resnick, The Alzheimer's Disease Neuroimaging Initiative (ADNI), The BIOCARD Study Team. Formal analysis: Praitayini Kanakaraj, Tianyuan Yao, and Daniel Moyer. Funding acquisition: Bennett A. Landman. Methodology: Praitayini Kanakaraj, Leon Y. Cai, Ho Hin Lee, and Daniel Moyer. Resources: Bennett A. Landman. Software: Praitayini Kanakaraj, Leon Y. Cai and Tianyuan Yao. Supervision: Daniel Moyer. Validation: Praitayini Kanakaraj, Eleftherios Garyfallidis, Kurt G. Schilling, and Adam Anderson. Writing – original draft: Praitayini Kanakaraj. Writing – review & editing: Nancy R. Newlin, Michael E. Kim, Chenyu Gao, Kimberly R. Pechman, Derek Archer, Timothy Hohman, Angela Jefferson, Lori L. Beason-Held, Susan M. Resnick, Adam Anderson, Kurt G. Schilling, Bennett A. Landman, and Daniel Moyer.

Availability of Data and Material The data used for the analysis of this article are confidential due to privacy or other restrictions.

Code Availability The code used for this article is available from https://github.com/MASILab/DeepN4.

Declarations

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process During the preparation of this work the authors used ChatGPT 3.5, an AI language model developed by OpenAI, in order to assist in rephrasing the text in this paper for clarification. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publication Not applicable.

Competing Interest The authors declare no competing interests.

References

Agrawal, A., Amos, B., Barratt, S., Boyd, S., Diamond, S., & Kolter, J. Z. (2019). Differentiable convex optimization layers. *Advances in neural information processing systems* (Vol. 32).

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage, 26*(3), 839–851.

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalization tools (ANTS). *Insight J, 2*(365), 1–35.

Axel, L., Costantini, J., & Listerud, J. (1987). Intensity correction in surface-coil MR imaging. *AJR American Journal of Roentgenology, 148*(2), 418–420.

Bakas, S., Akbari, H., Sotiras, A., et al. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data, 4*(1), 1–13.

Beekly, D. L., Ramos, E. M., Lee, W. W., et al. (2007). The National Alzheimer's Coordinating Center (NACC) database: The uniform data set. *Alzheimer Disease & Associated Disorders, 21*(3), 249–258.

Beekly, D. L., Ramos, E. M., van Belle, G., et al. (2004). The national Alzheimer's coordinating center (NACC) database: An Alzheimer disease database. *Alzheimer Disease & Associated Disorders, 18*(4), 270–277.

Besser, L. M., Kukull, W. A., Teylan, M. A., et al. (2018). The revised National Alzheimer's Coordinating Center's Neuropathology Form—available data and new analyses. *Journal of Neuropathology & Experimental Neurology, 77*(8), 717–726.

Brinkmann, B. H., Manduca, A., & Robb, R. A. (1998). Optimized homomorphic unsharp masking for MR grayscale inhomogeneity correction. *IEEE Transactions on Medical Imaging, 17*(2), 161–171.

Chen, C., Qin, C., Qiu, H., Ouyang, C., Wang, S., Chen, L., & Rueckert, D. (2020). Realistic adversarial data augmentation for MR image segmentation. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23* (pp. 667–677). Springer International Publishing.

Chuang, K.-H., Wu, P.-H., Li, Z., Fan, K.-H., & Weng, J.-C. (2022). Deep learning network for integrated coil inhomogeneity correction and brain extraction of mixed MRI data. *Scientific Reports, 12*(1), 8578.

Damadian, R. (1971). Tumor detection by nuclear magnetic resonance. *Science, 171*(3976), 1151–1153.

Esteban, O., Markiewicz, C. J., Blair, R. W., et al. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods, 16*(1), 111–116.

Fischl, B. (2012). *FreeSurfer. Neuroimage, 62*(2), 774–781.

Froeling, M., Tax, C. M., Vos, S. B., Luijten, P. R., & Leemans, A. (2017). "MASSIVE" brain dataset: Multiple acquisitions for standardization of structural imaging validation and evaluation. *Magnetic Resonance in Medicine, 77*(5), 1797–1809.

Gaillochet, M., Tezcan, K. C., & Konukoglu, E. (2020). Joint reconstruction and bias field correction for undersampled MR imaging. *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 44–52). Cham: Springer International Publishing.

Gispert, J. D., Reig, S., Pascau, J., Vaquero, J. J., García-Barreno, P., & Desco, M. (2004). Method for bias field correction of brain T1-weighted magnetic resonance images minimizing segmentation error. *Human Brain Mapping, 22*(2), 133–144.

Goldfryd, T., Gordon, S., & Raviv, T. R. (2021). *Deep semi-supervised bias field correction of Mr images* (pp. 1836–1840). IEEE.

Gorgolewski, K., Burns, C. D., Madison, C., et al. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics, 5*, 13.

Harms, M. P., Somerville, L. H., Ances, B. M., et al. (2018). Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage, 183*, 972–984.

Huo, Y., Xu, Z., Xiong, Y., et al. (2019). 3D whole brain segmentation using spatially localized atlas network tiles. *NeuroImage, 194*, 105–119.

IXI Dataset - Information eXtraction from images. (2020). Biomedical Image Analysis Group, Imperial College London. https://brain-development.org/ixi-dataset/

Jack, C. R., Jr., Bernstein, M. A., Fox, N. C., et al. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine, 27*(4), 685–691.

Jefferson, A. L., Gifford, K. A., Acosta, L. M. Y., et al. (2016). The Vanderbilt Memory & Aging Project: Study design and baseline cohort overview. *Journal of Alzheimer's Disease, 52*(2), 539–559.

Johnson, K. A. (2016). *Basic proton MR imaging: tissue signal characteristics*. Harvard Medical School. Archived from the original on, 03-05.

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*. arXiv preprint. arXiv:14126980

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., et al. (2019). OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *MedRxiv*, 2019.12. 13.19014902.

Landman, B. A., Huang, A. J., Gifford, A., et al. (2011). Multi-parametric neuroimaging reproducibility: A 3-T resource study. *NeuroImage, 54*(4), 2854–2866.

Mascalchi, M., Marzi, C., Giannelli, M., et al. (2018). Histogram analysis of DTI-derived indices reveals pontocerebellar degeneration and its progression in SCA2. *PLoS ONE, 13*(7), e0200258.

Mihara, H., Iriguchi, N., & Ueno, S. (1998). A method of RF inhomogeneity correction in MR imaging. *Magnetic Resonance Materials in Physics, Biology and Medicine, 7*(2), 115–120.

Miller, N., Liu, Y., Krivochenitser, R., & Rokers, B. (2019). Linking neural and clinical measures of glaucoma with diffusion magnetic resonance imaging (dMRI). *PLoS ONE, 14*(5), e0217011.

Narayana, P., Brey, W., Kulkarni, M., & Sievenpiper, C. (1988). Compensation for surface coil sensitivity variation in magnetic resonance imaging. *Magnetic Resonance Imaging, 6*(3), 271–274.

Payares-Garcia, D., Mateu, J., & Schick, W. (2023). NeuroNorm: An R package to standardize multiple structural MRI. *Neurocomputing, 550*, 126493.

Pineda, L., Fan, T., Monge, M., et al. (2022). Theseus: A library for differentiable nonlinear optimization. *Advances in Neural Information Processing Systems, 35*, 3801–3818.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234–241). Springer International Publishing.

Sacktor, N., Soldan, A., Grega, M., Farrington, L., Cai, Q., Wang, M. C., & Albert, M. (2017). *The BIOCARD index: a summary measure to predict onset of mild cognitive impairment (P1. 095)*. AAN Enterprises.

Schiffmann, R., & van der Knaap, M. S. (2009). Invited article: An MRI-based approach to the diagnosis of white matter disorders. *Neurology, 72*(8), 750–759.

Schilling, K. G., Blaber, J., Hansen, C., et al. (2020). Distortion correction of diffusion weighted MRI without reverse phase-encoding scans or field-maps. *PLoS ONE, 15*(7), e0236418.

Shock, N. W. (1984). *Normal human aging: The Baltimore longitudinal study of aging (No. 84)*. US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Aging, Gerontology Research Center.

Simkó, A., Löfstedt, T., Garpebring, A., Nyholm, T., & Jonsson, J. (2022). MRI bias field correction with an implicitly trained CNN. *International Conference on Medical Imaging with Deep Learning* (pp. 1125–1138). PMLR.

Simmons, A., Tofts, P. S., Barker, G. J., & Arridge, S. R. (1994). Sources of intensity nonuniformity in spin echo images at 1.5 T. *Magnetic Resonance in Medicine, 32*(1), 121–128.

Sled, J. G., Zijdenbos, A. P., & Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Transactions on Medical Imaging, 17*(1), 87–97.

Song, S., Zheng, Y., & He, Y. (2017). A review of methods for bias correction in medical images. *Biomedical Engineering Review, 1*(1).

Sridhara, S. N., Akrami, H., Krishnamurthy, V., & Joshi, A. A. (2021). Bias field correction in 3D-MRIs using convolutional autoencoders. *Medical Imaging 2021: Image Processing* (Vol. 11596, pp. 671–676). SPIE.

Tournier, J. D., Calamante, F., & Connelly, A. (2012). MRtrix: Diffusion tractography in crossing fiber regions. *International Journal of Imaging Systems and Technology, 22*(1), 53–66.

Tustison, N. J., Avants, B. B., Cook, P. A., et al. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging, 29*(6), 1310–1320.

Tustison, N. J., Cook, P. A., Holbrook, A. J., et al. (2021). The ANTsX ecosystem for quantitative biological and medical imaging. *Scientific Reports, 11*(1), 9068.

Van Essen, D. C., Smith, S. M., Barch, D. M., et al. (2013). The WU-Minn human connectome project: An overview. *NeuroImage, 80*, 62–79.

Vovk, U., Pernus, F., & Likar, B. (2007). A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Transactions on Medical Imaging, 26*(3), 405–421.

Wan, F., Smedby, Ö., & Wang, C. (2019). Simultaneous MR knee image segmentation and bias field correction using deep learning and partial convolution. *Medical Imaging 2019: Image Processing* (Vol. 10949, pp. 61–67). SPIE.

Weintraub, S., Besser, L., Dodge, H. H., et al. (2018). Version 3 of the Alzheimer Disease Centers' neuropsychological test battery in the Uniform Data Set (UDS). *Alzheimer Disease and Associated Disorders, 32*(1), 10.

Weintraub, S., Salmon, D., Mercaldo, N., et al. (2009). The Alzheimer's disease centers' uniform data set (UDS): The neuropsychological test battery. *Alzheimer Disease and Associated Disorders, 23*(2), 91.

Xu, Y., Wang, Y., Hu, S., & Du, Y. (2022). Deep convolutional neural networks for bias field correction of brain magnetic resonance images. *The Journal of Supercomputing, 78*(16), 17943–17968.

Yaniv, Z., Lowekamp, B. C., Johnson, H. J., & Beare, R. (2018). SimpleITK image-analysis notebooks: A collaborative environment for education and reproducible research. *Journal of Digital Imaging, 31*(3), 290–303.

Authors and Affiliations

Praitayini Kanakaraj[1] · Tianyuan Yao[1] · Leon Y. Cai[2] · Ho Hin Lee[1] · Nancy R. Newlin[1] · Michael E. Kim[1] · Chenyu Gao[3] · Kimberly R. Pechman[4] · Derek Archer[4,5,6] · Timothy Hohman[4,5,6] · Angela Jefferson[4,5,7] · Lori L. Beason-Held[8] · Susan M. Resnick[8] · The Alzheimer's Disease Neuroimaging Initiative (ADNI) · The BIOCARD Study Team · Eleftherios Garyfallidis[9] · Adam Anderson[2,10] · Kurt G. Schilling[11] · Bennett A. Landman[1,11,3,2,10] · Daniel Moyer[1]

✉ Praitayini Kanakaraj
praitayini.kanakaraj@vanderbilt.edu

Tianyuan Yao
tianyuan.yao@vanderbilt.edu

Leon Y. Cai
leon.y.cai@vanderbilt.edu

Ho Hin Lee
ho.hin.lee@vanderbilt.edu

Nancy R. Newlin
nancy.r.newlin@vanderbilt.edu

Michael E. Kim
michael.kim@vanderbilt.edu

Chenyu Gao
chenyu.gao@vanderbilt.edu

Kimberly R. Pechman
kimberly.r.pechman@vumc.org

Derek Archer
derek.archer@vumc.org

Timothy Hohman
timothy.j.hohman@vanderbilt.edu

Angela Jefferson
angela.jefferson@vanderbilt.edu

Lori L. Beason-Held
heldlo@grc.nia.nih.gov

Susan M. Resnick
resnicks@grc.nia.nih.gov

The Alzheimer's Disease Neuroimaging Initiative (ADNI)
clpreys@mghihp.edu

The BIOCARD Study Team
biocard@jhu.edu

Eleftherios Garyfallidis
elef@indiana.edu

Adam Anderson
adam.anderson@vanderbilt.edu

Kurt G. Schilling
kurt.g.schilling.1@vumc.org

Bennett A. Landman
bennett.landman@vanderbilt.edu

Daniel Moyer
daniel.moyer@vanderbilt.edu

[1] Department of Computer Science, Vanderbilt University, 400 24th Ave S, Nashville, TN 37240, USA

[2] Department of Biomedical Engineering, Vanderbilt University, Nashville, TN, USA

[3] Department of Electrical and Computer Engineering, Vanderbilt University, Nashville, TN, USA

[4] Vanderbilt Memory and Alzheimer's Center, Vanderbilt University Medical Center, Nashville, TN, USA

[5] Department of Neurology, Vanderbilt University Medical Center, Nashville, TN, USA

[6] Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN, USA

[7] Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA

[8] Laboratory of Behavioral Neuroscience, National Institute On Aging, National Institutes of Health, Baltimore, MD, USA

[9] Intelligent Systems Engineering, Indiana University, Bloomington, IN, USA

[10] Department of Radiology and Radiological Services, Vanderbilt University Medical Center, Vanderbilt University Medical, Nashville, TN, USA

[11] Vanderbilt University Institute for Imaging Science, Vanderbilt University Medical Center, Nashville, TN, USA