frontiers | Frontiers in Neuroimaging

TYPE Original Research
PUBLISHED 06 April 2023
DOI 10.3389/fnimg.2023.1072759

OPEN ACCESS

EDITED BY
Jon-Fredrik Nielsen,
University of Michigan, United States

REVIEWED BY
Hamid Osman,
Taif University, Saudi Arabia
Manas Gaur,
University of Maryland, Baltimore County, United States
Deepa Tilwani,
Artificial Intelligence Institute at University of South Carolina, United States in collaboration with reviewer MG

*CORRESPONDENCE
Sairam Geethanath
✉ sairam.geethanath@mssm.edu

SPECIALTY SECTION
This article was submitted to Brain Imaging Methods, a section of the journal Frontiers in Neuroimaging

RECEIVED 17 October 2022
ACCEPTED 15 March 2023
PUBLISHED 06 April 2023

CITATION
Ravi KS, Nandakumar G, Thomas N, Lim M, Qian E, Jimeno MM, Poojar P, Jin Z, Quarterman P, Srinivasan G, Fung M, Vaughan JT Jr and Geethanath S (2023) Accelerated MRI using intelligent protocolling and subject-specific denoising applied to Alzheimer's disease imaging. *Front. Neuroimaging* 2:1072759. doi: 10.3389/fnimg.2023.1072759

Accelerated MRI using intelligent protocolling and subject-specific denoising applied to Alzheimer's disease imaging

Keerthi Sravan Ravi[1,2], Gautham Nandakumar[3], Nikita Thomas[3], Mason Lim[3], Enlin Qian[1,2], Marina Manso Jimeno[1,2], Pavan Poojar[4], Zhezhen Jin[5], Patrick Quarterman[6], Girish Srinivasan[3], Maggie Fung[6], John Thomas Vaughan Jr.[1,2] and Sairam Geethanath[2,4]* for the Alzheimer's Disease Neuroimaging Initiative

[1]Department of Biomedical Engineering, Columbia University in the City of New York, New York, NY, United States, [2]Columbia University Magnetic Resonance Research Center, Columbia University in the City of New York, New York, NY, United States, [3]PMX, Palatine, IL, United States, [4]Department of Diagnostic, Molecular and Interventional Radiology, Accessible MRI Laboratory, Biomedical Engineering and Imaging Institute, Icahn School of Medicine at Mt. Sinai, New York, NY, United States, [5]Mailman School of Public Health, Columbia University in the City of New York, New York, NY, United States, [6]MR Clinical Solutions, GE Healthcare, New York, NY, United States

Magnetic Resonance Imaging (MR Imaging) is routinely employed in diagnosing Alzheimer's Disease (AD), which accounts for up to 60–80% of dementia cases. However, it is time-consuming, and protocol optimization to accelerate MR Imaging requires local expertise since each pulse sequence involves multiple configurable parameters that need optimization for contrast, acquisition time, and signal-to-noise ratio (SNR). The lack of this expertise contributes to the highly inefficient utilization of MRI services diminishing their clinical value. In this work, we extend our previous effort and demonstrate accelerated MRI *via* intelligent protocolling of the modified brain screen protocol, referred to as the Gold Standard (GS) protocol. We leverage deep learning-based contrast-specific image-denoising to improve the image quality of data acquired using the accelerated protocol. Since the SNR of MR acquisitions depends on the volume of the object being imaged, we demonstrate subject-specific (SS) image-denoising. The accelerated protocol resulted in a 1.94× gain in imaging throughput. This translated to a 72.51% increase in MR Value—defined in this work as the ratio of the sum of median object-masked local SNR values across all contrasts to the protocol's acquisition duration. We also computed PSNR, local SNR, MS-SSIM, and variance of the Laplacian values for image quality evaluation on 25 retrospective datasets. The minimum/maximum PSNR gains (measured in dB) were 1.18/11.68 and 1.04/13.15, from the baseline and SS image-denoising models, respectively. MS-SSIM gains were: 0.003/0.065 and 0.01/0.066; variance of the Laplacian (lower is better): 0.104/−0.135 and 0.13/−0.143. The GS protocol constitutes 44.44% of the comprehensive AD imaging protocol defined by the European Prevention of Alzheimer's Disease project. Therefore, we also demonstrate the potential for AD-imaging *via* automated volumetry of relevant brain anatomies. We performed statistical analysis on these volumetric measurements of the hippocampus and amygdala from the GS and accelerated protocols, and found that 27 locations were in excellent agreement. In conclusion, accelerated brain imaging with the potential for AD imaging was demonstrated, and image quality was recovered post-acquisition using DL-based image denoising models.

KEYWORDS

autonomous MRI, deep learning, explainable AI, multi-contrast denoising, MR value

1. Introduction

Dementia affected 50 million people worldwide in 2018, with an estimated economic impact of US$ 1 trillion a year (Patterson, 2018; Banerjee et al., 2020). Alzheimer's Disease (AD) accounts for up to 60–80% of dementia cases and potentially begins upto 20 years before the first symptoms emerge (Bateman et al., 2012). A global trend of longer lifespans has resulted in an increase in the prevalence of dementia/AD (Silva-Spínola et al., 2022). An accurate differential diagnosis of AD is crucial to determine the right course of treatment (Vernooij and van Buchem, 2020). Magnetic Resonance Imaging (MR Imaging) is a powerful imaging modality to obtain valuable information about the brain structure anatomy *via* the acquisition of high-resolution images. It is routinely employed in AD diagnosis. Traditionally, structural MR Imaging (sMRI) is used to exclude treatable and reversible causes of dementia such as brain tumors, subdural hematomas, cerebral infarcts, or hemorrhages (Falahati et al., 2015). The Alzheimer's Disease Neuroimaging Initiative (ADNI) has included sequences in their standardized protocol to specifically image cerebrovascular disease (Fluid Attenuation Inversion Recovery [FLAIR]) and cerebral microbleeds (T_2^* gradient echo) (Weiner et al., 2017). Studies have demonstrated that atrophy of the hippocampus and amygdala volume are reliable indicators of progression from pre-dementia to AD (Simmons et al., 2011). These imaging biomarkers, or image-derived phenotypes (IDP), can be obtained from sMRI. The European Prevention of Alzheimer's Disease (EPAD, https://ep-ad.org/open-access-data/overview) prescribes four core and five advanced sequences for AD imaging. The core sequences are 3D T_1-weighted (T1w), 3D fluid-attenuated inversion recovery (FLAIR), 2D T_2-weighted (T_2w), and 2D T_2^*-weighted (T_2^*w). The advanced sequences are 3D T_2^*w, 3D susceptibility-weighted imaging (SWI), diffusion-weighted imaging (DWI) or dMRI, resting-state functional MR Imaging, and arterial spin labeling. Mehan et al. (2014) reported on the adequacy of a four-sequence protocol consisting of an axial T_1w, axial T_2w FLAIR, axial DWI, and axial SWI images to evaluate new patients with neurological complaints.

Despite MR Imaging's critical utility in neuroimaging for AD, there exist multiple challenges that lower the accessibility of the technology to the general population. MR Imaging is expensive and time-consuming, and subjects with MR-unsafe materials (such as medical device implants, prostheses, etc.) are not eligible for MR Imaging. Considerable research efforts have been directed toward accelerating acquisition times by exploiting the temporal or spatiotemporal redundancies in the images (Tsao and Kozerke, 2012). However, protocol optimization to accelerate MR Imaging requires local expertise. Each sequence involves multiple configurable parameters that need optimization for contrast, acquisition time, and signal-to-noise ratio (SNR). A large number of these combinations exist–for example, 29 million for 12 sequences in a protocol (Block, 2018)—and choosing an optimal combination in real-time is difficult. Since the availability and access to technical training are limited in under-served regions (Geethanath and Vaughan, 2019), this results in a scarcity of local expertise required to operate MR Imaging hardware and perform MR Imaging examinations. These factors, along with other cultural and temporal constraints contribute to the highly inefficient utilization of MR Imaging services diminishing their clinical value (Geethanath and Vaughan, 2019).

This combination of a very high-dimensional optimization space and inadequate local expertise necessitates a data-driven approach to augment the available manpower. Previous works involve machine learning approaches for automated RF pulse design (Shin et al., 2020), sequence design (Zhu et al., 2018), or even a joint framework for sequence generation and data reconstruction (Walker-Samuel, 2019; Loktyushin et al., 2021). We believe that augmenting human expertise by leveraging deep learning (DL) techniques across the MR Imaging pipeline can consistently yield improved MR Value irrespective of where the service is offered or the expertise involved. MR Value is an initiative by the International Society of Magnetic Resonance in Medicine to measure the utility of MR Imaging in the context of constantly evolving healthcare economics (https://www.ismrm.org/the-mr-value-initiative-phase-1/). We based our prior work on this premise and demonstrated preliminary results from MR Value-driven Autonomous MR Imaging, dubbed AMRI (Ravi and Geethanath, 2020; Ravi et al., 2020).

In this work, we extend our previous effort and demonstrate accelerated MR Imaging *via* intelligent protocolling of the modified brain screen protocol (dubbed Gold Standard, GS) employed at our institution. We leverage deep learning-based image denoising to improve the image quality of data acquired using the accelerated protocol. The GS protocol consisted of six sequences: sagittal 3D T_1w, axial 2D T_2w, axial 2D T_2w FLAIR, axial 2D SWI, axial DWI, and axial 2D T_1w. Overall, the GS protocol constitutes 44.44% of the comprehensive EPAD imaging protocol and includes all sequences deemed adequate for neuroimaging by Mehan et al. (2014). Therefore, we also investigate the potential of the accelerated protocol for AD-screening by benchmarking volumetry of the hippocampus and amygdala against measurements from the GS protocol. This volumetry can be used for early detection of atrophy.

In the following sections, we first detail the implementations of intelligent protocolling (Section 2.1), and image denoising using deep learning (Section 2.2). Subsequently, we discuss the image analyses that were performed, including the statistical evaluation technique (Section 2.3) that was recommended by an expert biostatistician with 23 years of experience. Finally, we describe the four experiments performed.

Overall, the contributions of this work are:

1. Demonstrating look-up tables to achieve intelligent protocolling by trading-off image quality for acquisition time.
2. Performing subject-specific image denoising using deep learning to recover image quality post-acquisition.
3. Demonstrating potential for accelerated AD-imaging by evaluating volumetries of AD-related IDPs.

2. Methods and materials

This section is organized as follows. Section 2.1 describes intelligent protocolling—accelerating the routine brain screen protocol employed at our institution by consulting a Look Up

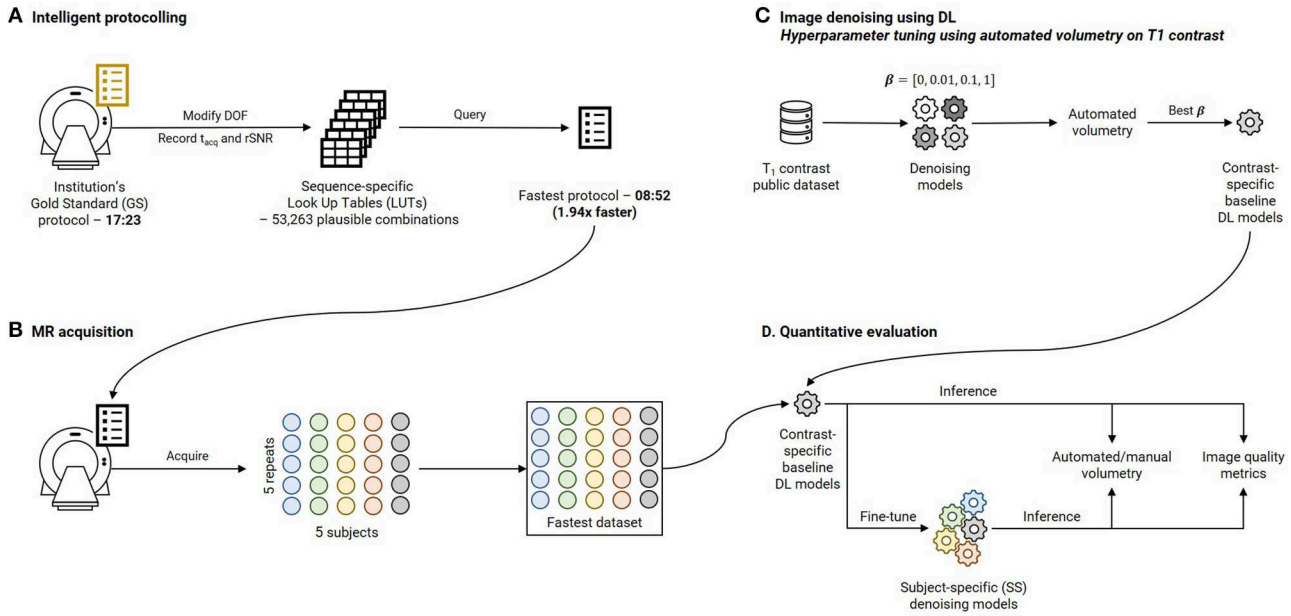

FIGURE 1
Overview of this work. **(A)** Intelligent protocolling: specific acquisition parameters (Degrees Of Freedom, DOF, Table 1) were varied and acquisition duration (t_{acq}) and relative SNR (rSNR) values were recorded from the vendor-provided user interface. All these values were recorded to construct a sequence-specific Look Up Tables (LUTs). The "Fastest" protocol was assembled by querying these LUTs to obtain P_{acq} for each sequence. **(B)** Five subjects were scanned using the Fastest protocol over five repeats to assemble the Fastest dataset. **(C)** Contrast-specific image denoising models were trained to improve image quality of the Fastest dataset. **(D)** Automated volumetry was performed for the T_1-contrast denoising model, and manual volumetry for the remaining contrasts, to evaluate the performances. In addition, quantitative image quality metrics were computed.

Table (LUT). Section 2.2 presents the development of deep learning models to achieve subject-specific (SS) denoising and the explainability of the models. Section 2.3 discusses the quantitative image quality metrics that were computed, and the statistical analysis that was performed. The four experiments that were performed to investigate the hypotheses are detailed in Section 2.5. Finally, Section 2.6 describes visualizing intermediate filter outputs for explainable AI. Figure 1 presents an overview of this work by briefly illustrating the methods involved in intelligent protocolling, data acquisition, DL-based image-denoising, and quantitative evaluation.

2.1. Intelligent acquisition using look-up tables

Table 1 lists the seven GS sequences and their corresponding acquisition parameters and durations. The cumulative acquisition time was **17:23** (minutes:seconds), as per the vendor console's user interface (UI). An experienced clinical application specialist was consulted to collate a list of acquisition parameters that could be varied without compromising image contrast for each sequence in the GS protocol. These acquisition parameters were referred to as degrees of freedom (DOF), also listed in Table 1. Exhaustive combinations of these DOF or a hundred randomly chosen combinations, whichever was smaller, were entered into the vendor console's UI. For each combination (P_{acq}), the acquisition time (t_{acq}), and relative signal-to-noise ratio (rSNR) value were recorded. The P_{acq}, and corresponding t_{acq} and rSNR values were stored in a LUT. These were searched to obtain the optimal P_{acq} yielding the lowest t_{acq}. This procedure was repeated for each sequence in the GS protocol. Once the sequence-specific LUTs were constructed, they were consulted to derive sequence-specific optimal P_{acq} to derive the fastest protocol. The search procedure is described as follows, applicable to each sequence individually:

2.1.1. Compute percentage time allocated

First, the minimum time percentage value (y_1) was computed as the ratio of the shortest sequence acquisition time to the shortest protocol acquisition time (x_1). Similarly, the maximum time percentage value (y_2) was computed from the longest protocol acquisition time (x_2). Now, for an imposed protocol acquisition time constraint (T_{acq}), the percentage time allocated (%TA) to a sequence was derived from the straight line fitting the minimum and maximum time percentage values, as described by Equation 1:

$$Percentage\ time\ allocated\ (\%TA) = \left(\frac{y_2 - y_1}{x_2 - x_1}\right)^{*}(x - x_1) + y_1 \quad (1)$$

2.1.2. Compute weighted rank

The time allocated in seconds for this sequence (t_{acq}) was derived from the %TA value. The LUT was filtered by discarding DOF combinations whose acquisition times exceeded t_{acq}. Of the remaining combinations, weighted differences of rSNR and DOF values with the corresponding default values from the GS protocol were computed. Higher weights were assigned to

TABLE 1 Acquisition parameters and durations of the sequences constituting the gold standard (GS) and the Fastest (LUT-derived) protocols.

		Sequence	Imaging plane (2D/3D)	Flip angle (deg)	Echo time/ repetition time [Inversion time] (ms)	Slices [Slice thickness (mm)]	Acquisition time (minutes: seconds)	DOF	
Gold Standard	1	T_1 MPRAGE	Sagittal (3D)	13	[450]	172 [1.0]	2:44	Num, NEX, ST	
	2	DWI	Axial (2D)		Minimum/5554	47 [3.6]	0:44	Num, ASSET, Dir, ST, TR	
	3	SWI	Axial (2D)	15	Minimum full/Minimum	72 [2.4]	2:30	Num, NEX, ST	
	4	T_2	Axial (2D)	142	/6996	56 [3.0]	2:21	Num, ARC, ST, TR	
	5	T_2 FLAIR	Axial (2D)	160	90/9000 [2477]	56 [3.0]	3:46	Num, ARC, ETL, NEX, ST, TR	
	6	T_1 MPRAGE	Sagittal (3D)	13	[450]	172 [1.0]	2:44	Same as before	
	7	T_1	Axial (2D)	111	24/2846 [1133]	56 [3.0]	2:34	Num, ARC, ETL, NEX, ST, TR	17:23
Expert Express	1	T_1 FLAIR	Sagittal (3D)	111	24/2143.4 [724]	27 [5.0]	0:37	NA	
	2	DWI	Axial (2D)		Minimum/2930		0:24		
	3	T_2^*	Axial (2D)	15	8/346.1		0:44		
	4	T_2	Axial (2D)	142	102/4627		0:34		
	5	T_2 FLAIR	Axial (2D)	160	90/9000 [2473]		1:21		
	6	T_1 FLAIR	Sagittal (3D)	111	24/2143.4 [724]		0:37		
	7	T_2 PROPELLER	Axial (2D)	130	/6301		0:44		
	8	T FLAIR PROPELLER	Axial (2D)	142	/1000 [2365]		2:11		07:12
Fastest	1	T_1 MPRAGE	Sagittal (3D)	13	[450]	172 [1.6]	1:41	NA	
	2	DWI	Axial (2D)		Minimum/7500	32 [3.9]	0:30		
	3	T_2^*	Axial (2D)	15	13.5/580	31 [4.3]	0:34		
	4	T_2	Axial (2D)		121/1204	27 [5.0]	1:30		
	5	T_2 FLAIR	Axial (2D)	160	90/6900 [2191]	45 [3.8]	1:10		
	6	T_1 MPRAGE	Sagittal (3D)	13	[450]	172 [1.6]	1:41		
	7	T_1	Axial (2D)			[4.0]	13:28		08:52

Each protocol's acquisition duration is presented above the arrow-outs. The degrees of freedom (DOF) represent the parameters that were varied to modify the GS protocol without compromising the image contrast. Dir, diffusion directions; Num, number of slices; ARC/ASSET, acceleration options; ETL, echo train length; NEX, number of excitations; ST, slice thickness; TR, repetition time.

DOF values contributing more significantly to the image contrast (Supplementary Table 1). Finally, these weighted differences were summed to obtain a rank for each DOF combination, and the resulting LUT was sorted in ascending order of this rank value. Thus, the combination achieving the lowest rank value had the smallest difference in those DOF values which most significantly contributed to the image contrast.

2.1.3. Obtain optimal combination

For each time constraint, the combination with the lowest rank was chosen as the optimal set of acquisition parameters.

This process was repeated with lower imposed T_{acq} in each iteration until an optimal P_{acq} could not be obtained for every sequence in the GS protocol. In this way, the Fastest protocol was derived by consulting the sequence-specific LUTs. Data acquired utilizing the GS protocol were referred to as the GS dataset. Data acquired from the Fastest protocol for Experiments 1, 2, and 4 (see Section 2.5) were referred to as the Fastest dataset, collectively. The SWI sequence in the GS protocol was replaced by a T_2^*w sequence in the Fastest protocol. The experienced clinical application specialist's express protocol was dubbed the Expert Express (EE) protocol. Data was also acquired from this protocol for comparison (refer Experiment 2 in Section 2.5), referred to as the EE dataset.

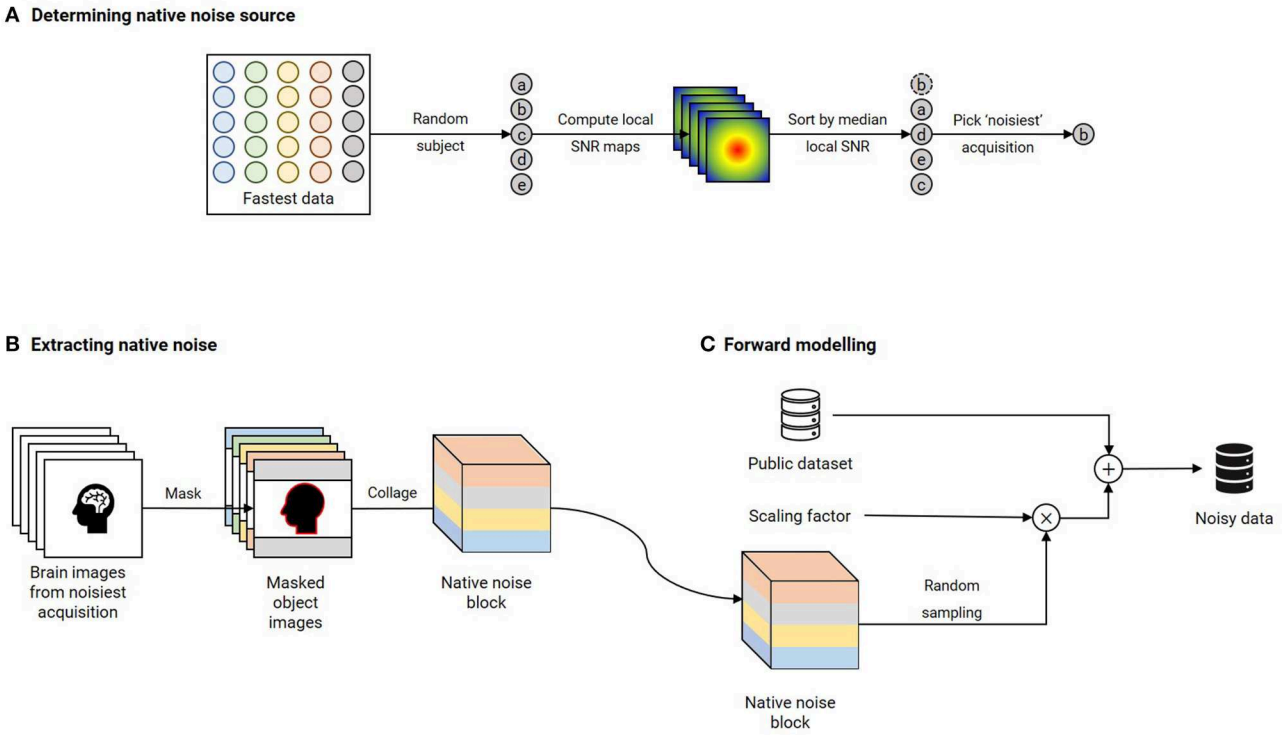

FIGURE 2
Forward modeling of noisy data using native noise for each contrast. **(A)** A random subject's acquisitions are chosen from the Fastest dataset. Brain-masked local SNR maps are computed, and these acquisitions are sorted in ascending order of median local SNR. The acquisition yielding the lowest value is chosen as the "noisiest" acquisition. **(B)** Native noise is extracted from this noisiest acquisition by assigning all non-object voxels as noise. These noise voxels are collaged to obtain a native noise block. **(C)** The noise-free public dataset is corrupted by native noise randomly sampled from the native noise block and scaled by a noise factor. This scaling factor is determined by iteratively incrementing from an initial value of 0.5 until the median brain-masked local SNR of a randomly chosen volume from the public dataset that is corrupted by native noise matches that of the noisiest acquisition.

2.2. Image denoising using deep learning

Two popular image denoising approaches are to directly predict the denoised image or to obtain the denoised image as the residual of the input noisy image and the predicted noise. We adopt a ResNet-inspired network architecture demonstrated to improve training performance and stability (He et al., 2016), to directly predict the denoised output. Individual contrast-specific denoising models were trained on pairs of noisy-denoised images from publicly available brain MR Imaging datasets (see below). Finally, SS denoising was performed by fine-tuning the models on pairs of noisy-denoised images from the prospectively acquired Fastest dataset.

2.2.1. Datasets, forward simulation, and data splits

Publicly available datasets were utilized to train the contrast-specific image denoising models. T_1 and T_2 contrasts: IXI dataset (https://brain-development.org/ixi-dataset/); T_2^*: ADNI 3 (https://adni.loni.usc.edu); T_2 FLAIR: MSSEG-2 (Commowick et al., 2021) and DWI: AOMICID-1000 (Snoek et al., 2021). Wherever applicable, datasets were filtered to retain only the 3T data. Only the central 50% slices were utilized, and the remaining slices were discarded to avoid either unwanted anatomy or pure background noise. Supplementary Figure 1 presents the search criteria that were utilized to filter the ADNI 3 dataset for relevant results. For DWI, only the b0 images were utilized from the AOMICID-1000 dataset. All these datasets were assumed to be free of MR image artifacts, referred to as "clean images". Figure 2 presents the forward modeling process to generate noise-corrupted data ("noisy images"), described here as follows. First, the object-masked local SNR maps were computed on all acquisitions of an arbitrarily chosen subject from the Fastest dataset (see Experiment 4, Section 2.5). Object masking was based on the technique in Jenkinson (2003) and resulted in all background values being set to zero. All remaining non-zero values were considered to belong to the samples, which were non-skull-stripped brain images. The local SNR maps were computed based on the method reported in Golshan et al. (2013). The volume yielding the lowest median SNR was the "noisiest acquisition" (Figure 2A). Next, motivated by work in Geethanath et al. (2021), Qian et al. (2022), native noise values were extracted from this noisiest acquisition. These noise values were collaged to form a native noise block (Figure 2B). This was randomly sampled to obtain noise values, which were scaled and added to the clean images to obtain noisy images (Figure 2C). The scaling factor was determined using an iterative approach. A volume was chosen at random from the public dataset. Starting with an initial value of 1.0 (corresponding to no scaling), the scaling factor was increased by 0.5 in each iteration until the median object-masked local SNR of the corrupted volume was lesser than that

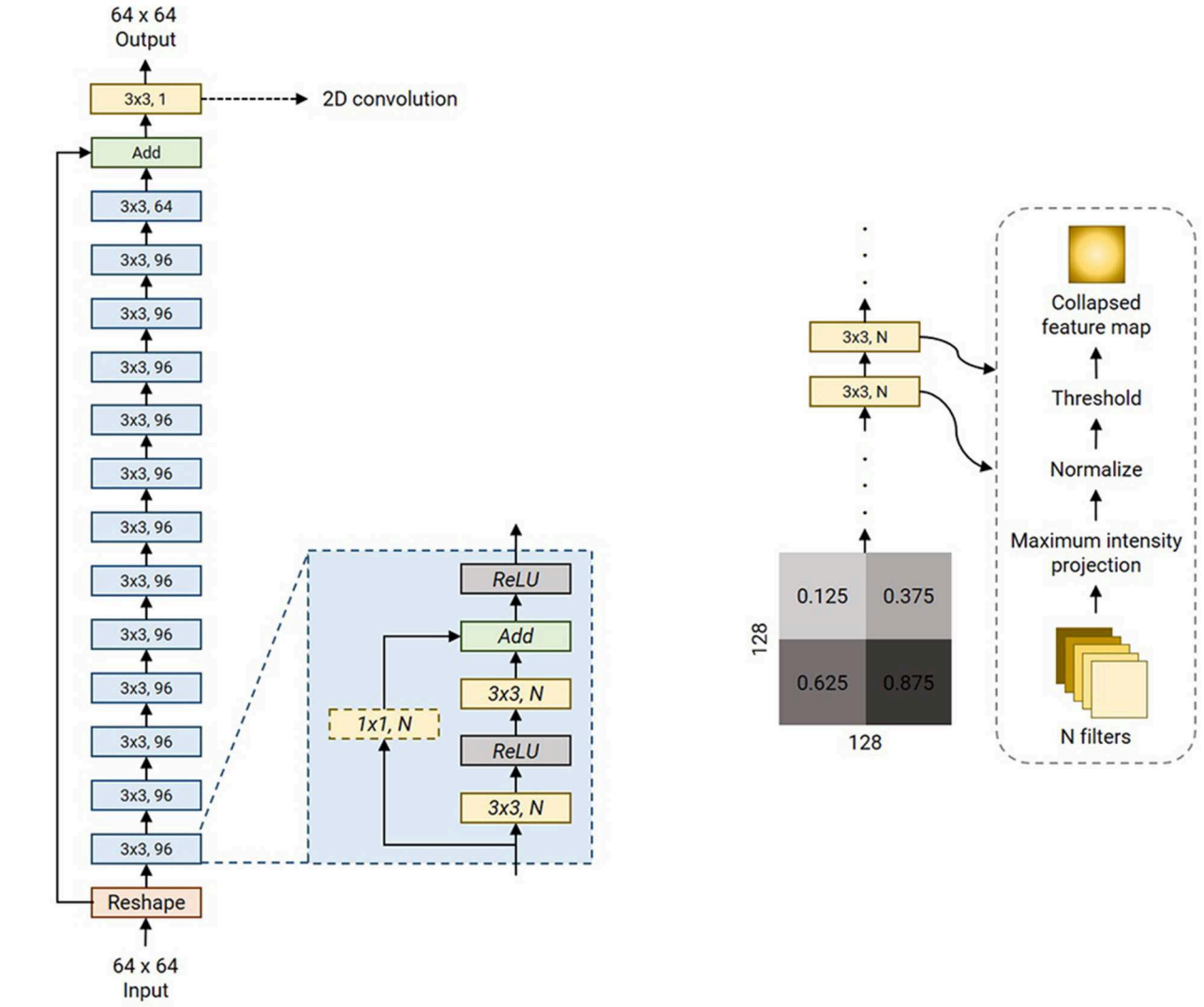

FIGURE 3
Network architecture of the image denoising models and illustration of the explainable AI experiment. **(A)** The denoising model leverages multiple ResBlocks (blue) and minimizes the Mix-L+FTD loss (Section 2.2) using the Adam optimizer (refer Table 2 for initial learning rates). **(B)** The explainable AI experiment involves feeding a 128 × 128 DC biased input image. This is a collage of four 64 × 64 constant-valued image patches, as shown (numbers represent the intensity values). As the input is transformed by each 2D convolution layer, all the filter outputs are collapsed into a single image *via* maximum intensity projection. This is normalized to lie in the range [0, 1.0] and hard-thresholded to only retain values >0.75. All the 2D convolution layers' collapsed feature maps are collaged to investigate the mechanism of the denoising process by the network.

of the noisiest acquisition. We chose median over mean as the guiding measure since it was less affected by skewed distributions. An 85–10–5% subject-wise split was performed to form the train, validation, and test sets.

2.2.2. Network architectures

Figure 3A is an illustration of the network architecture common to all contrast-specific image-denoising models. To predispose the network to learn denoising filters whilst being anatomy agnostic, we adopted a patch-wise approach in this work. Overlapping patches of size 64 × 64 were input to the network. Thirteen ResBlocks leveraged skip connections to improve training performance (He et al., 2016). Each ResBlock consisted of two ReLU-activated (Nair and Hinton, 2010) 3 × 3 2D convolution layers. In case of a mismatch in the number of filters between the previous and current ResBlocks (N_1, N_2), the skip connection included a 1 × 1 2D convolution with N_2 filters. Otherwise, the skip connection was an identity operation. Additionally, an identity skip connection was used to add the input data to the pre-final layer in the overall network. The final layer was a 3 × 3 2D convolution layer with 1 filter. All 2D convolution layers were ReLU-activated, and all development, training, and testing were performed using Keras 2.6/TensorFlow 2.6.2 (Abadi et al., 2015, 2016) libraries.

2.2.3. Loss functions

Zhao et al. report the superiority of their mixed loss (Mix-L) function for image denoising, among other image quality restoration applications (Zhao et al., 2016). This loss is a weighted sum of l_1 and MS-SSIM losses:

$$Mix - L = \alpha\,(1 - MS - SSIM) + (1 - \alpha)\,l_1 \quad (2)$$

... where α was set to 0.84. We modified Mix-L to incorporate a data-consistency term with the measured data in the Fourier domain, referred to as Mix-L+FTD:

$$f = M \odot \left| F(y_{pred}) - F(y_{true}) \right| \quad (3)$$

$$Mix - L + FTD = Mix - L + \beta\, || \frac{f}{||f||_2} ||_2 \quad (4)$$

... where F was the 2D Fourier Transform and M was a mask to only retain the central crop of the k-space of size 16 $\times$ 16. The $\odot$ operator represented the Hadamard product and β determined the trade-off between the denoising and data-consistency errors. We investigated $\beta = [0,\ 0.01,\ 0.1,\ 1]$ in our experiments. To determine the best β, the RMSEs of the volumetric measures ($RMSE_{vol}$) were computed using an automated tool on T_1 denoised outputs. The β yielding the lowest mean $RMSE_{vol}$ was chosen to train the denoising models for the remaining contrasts. Section 2.3 describes the automated T_1 volumetry tool and computing $RMSE_{vol}$ in detail.

2.2.4. Training

All contrast-specific denoising models were trained for 100 epochs with a batch size of 256. The Adam optimizer (Kingma and Ba, 2015) was utilized to minimize the Mix-L+FTD loss with the optimal β, determined as stated above. During training, every input slice was cropped to a 64 $\times$ 64 patch. The bounds for the random crop were manually determined by examining the corresponding public dataset such that the random crops would mostly include brain anatomy. A callback was utilized to save the model achieving the lowest validation loss (corresponding to "best performance"). At the end of the training process, this model was chosen as the best model for evaluation, including to determine the optimal β.

2.2.5. Subject-specific denoising

SS median local SNRs were computed on masked brain volumes from the Fastest dataset to verify the premise of SS denoising. The values were computed only on the central 50% of the slices. Next, the same noise scaling factors were utilized to corrupt each subject's noisiest acquisition from the Fastest dataset with native noise. This data was used to fine-tune the baseline denoising models to achieve SS denoising. This approach posed SS denoising as a self-supervised learning problem, mimicking the noisy-as-clean method demonstrated in Xu et al. (2020). The initial learning rate of the Adam optimizer was reduced to avoid large modifications to the weights which would otherwise harm the learned representations (Table 2).

TABLE 2 Initial learning rates (LRs) for the contrast-specific baseline and subject-specific (SS) denoising models.

	Contrast	Initial learning rate	
		Baseline	Subject-specific (SS)
1	T_1	2.5×10^{-4}	1×10^{-5}, 1×10^{-6}
2	T_2	1×10^{-4}	1×10^{-5}
3	T_2 FLAIR	1×10^{-4}	1×10^{-4}
4	T_2^*	2.5×10^{-4}	NA
5	DWI	1×10^{-4}	1×10^{-5}

The baseline denoising models were finetuned on Fastest data to obtain the SS denoising models. Therefore, the initial LRs were reduced to avoid harming the learned representations. For T_1, one subject required a LR of 1×10^{-6} since the loss values did not decrease with an LR of 1×10^{-5}. For T_2 FLAIR, the LR value was not changed since the model did not otherwise converge to lower loss values.

2.3. Image analysis

Thomas et al. (2020) demonstrated an end-to-end pipeline for fully automated mental health screening (Thomas et al., 2020). The authors leveraged a DL model to segment the various subgroups. Further development on the previous work includes a second DL model to segment the brain tissues (white matter, gray matter, cerebrospinal fluid). This second DL model was based on the nnUnet (Isensee et al., 2021), and an evaluation of its performance is presented in Supplementary File 1. We leveraged this tool to perform automated volumetry to measure the performance of the denoising models. HTML reports were generated containing volumetric measures of 27 brain subregions and 3 brain tissues. These were programmatically extracted and tabulated. $RMSE_{vol}$ was calculated as the mean of RMSEs of each of the volumetric measures. A benign White Matter Hyperintensity (WMH) was identified in data acquired from one subject. The free, open-source, and multi-platform 3D Slicer software [https://www.slicer.org/, (Fedorov et al., 2012)] was used to perform manual volumetry of this WMH on the T_2, T_2^*, T_2 FLAIR and DWI data by four different raters with 3–8 years of MR Imaging experience. All volumetries were performed on data acquired for Experiment 4 (see Section 2.5.4).

Additionally, a set of image quality metrics were also computed to evaluate the denoising models. These were: median object-masked local SNR, Peak SNR (PSNR, dB), Multi-scale Structural Similarity Index [MS-SSIM, (Wang et al., 2003)], the variance of the Laplacian, referred to as var-Lap (Pech-Pacheco et al., 2000), and MR Value. While local SNR, PSNR, and MS-SSIM metrics are commonly used to measure image quality, we obtain var-Lap values to measure the amount of blurring. We included this metric in our evaluations since blurring negatively affected the automated volumetry on T_1 (preliminary experiments not reported in this work).

2.4. Statistical analysis

The intra-class correlation coefficient (ICC) was calculated based on the analysis of variance (ANOVA) with repeated measures

to assess the agreement of the volumetric measures amongst the GS, denoised baseline, and denoised SS methods. The ICC values greater than 0.9 indicate excellent agreement, values between 0.75 and 0.9 indicate good agreement, and values between 0.5 and 0.75 indicate moderate agreement.

2.5. Experiments

We performed four experiments to investigate hypotheses regarding the throughput of the Fastest protocol, and the image quality of the Fastest dataset. Supplementary Table 2 lists the experiments performed, the protocols executed, numbers of healthy volunteers imaged, the corresponding claims and hypotheses investigated, and their respective evaluation criteria. In total, 31 brain volumes were acquired from five volunteers across the four experiments. The data acquired from the GS and Fastest protocols are referred to as the GS and Fastest datasets, respectively.

2.5.1. Experiment 1–Throughput

The goal of experiment one was to investigate if the Fastest protocol obtained from the LUT would yield an improvement in throughput. One volunteer was imaged using the GS and Fastest protocols, and a video recording of the entire imaging session was captured. Throughput was computed as the ratio of the table time measurement of the Fastest protocol to that of the GS protocol. In this work, table time is defined as the duration between the scanner bed reaching the center of the bore at the start of the imaging and the scanner bed returning to the home position. We also determined the MR Values of the GS and Fastest protocols, calculated as the ratio of the cumulative median object-masked local SNR values across all contrasts to the protocol's acquisition duration, T_{acq}:

$$\text{MR Value} = \frac{\sum_{c=1}^{\text{Contrasts}} \text{median object} - \text{masked local SNR}_c}{\text{T}_{\text{acq}}} \quad (5)$$

The object-masked local SNR maps were computed on the central 50% slices across all sequences in each protocol.

2.5.2. E2–Image quality

Experiment two quantitatively compared the image quality of the GS, EE, and Fastest datasets using the following metrics: median object-masked local SNR and var-Lap. PSNR and SSIM were not used since they are not reference-less metrics.

2.5.3. E3–SNR recovery

We investigated the feasibility of employing the Fastest protocol. It involved utilizing the image-denoising deep learning models described in Section 2.2 to improve image quality. The metrics described in Section 2.5 were utilized to determine if denoising the Fastest dataset achieved comparable quality to the GS dataset.

2.5.4. E4–Repeatability

A repeatability test to demonstrate the consistency of the quantitative image quality metrics was performed. The GS and Fastest protocols were employed to acquire data from five subjects over five repeats. Automated and manual volumetry were performed on the acquired data as described in Section 2.3.

2.6. Visualizing learned filters for explainable AI

There are no formal definitions for *interpretability* and *explainability* in the field of Artificial Intelligence and in the sub-field of DL (Doshi-Velez and Kim, 2017; Lipton, 2018; Miller, 2019; Aggarwal et al., 2023). However, current explainable AI practices can be cast as a type of model interpretability (Rahman, 2022). Image denoising is a combination of image synthesis and regression, and explainable AI methods do not currently exist for these tasks. Therefore, in this work, we choose to investigate the intermediate outputs of the filters of the 2D convolution layers as a method of explaining the denoising mechanism. The LUT search algorithm is inherently explainable since the ranks are computed as a weighted combination of the DOF. Explainability of the nnUnet utilized in computing the AD-related volumetric measurements is beyond the scope of this work. However, the performance of the nnUnet is available in Supplementary File 1. A 256×256 collage of four panels was assembled. Each of the four 64×64 panels was made up of a single intensity from the following values: $[1.25, 3.75, 6.25, 8.75] \times 10^{-1}$. This collage was corrupted with native noise from an arbitrarily chosen subject's T_1w acquisition. Subsequently, it was denoised using the baseline denoising model for the T_1 contrast. The filter outputs of each 2D convolutional layer were obtained, and maximum intensity projection was performed to achieve dimensionality reduction. Therefore, this collapsed N filter outputs into a single map. This was normalized to lie in the range $[0, 1.0]$. Finally, this collapsed feature map was hard thresholded to only retain values >0.75. Figure 3B briefly illustrates this procedure.

3. Results

3.1. Intelligent acquisition using look-up tables

rSNR was assigned the smallest weight when constructing the LUT since the aim of this work was to achieve acceleration by trading-off SNR which could be recovered post-acquisition *via* deep learning methods. The total duration of the Fastest protocol that was obtained by querying the LUTs was 8:34 (minutes:seconds). This was a 50.71% reduction in acquisition time from the GS protocol, which required 17:23. The EE protocol only required 7:12. It primarily achieved acceleration by employing the 3D T_1 FLAIR sequence. This is not a true 3D acquisition and only required 0:37 when compared to 2:44 and 1:41 for the 3D T_1w sequences in the GS and Fastest protocols, respectively. Figure 4 is a collage of one representative slice of an arbitrarily chosen subject, across contrasts. The rows represent the different datasets–GS, EE, Fastest,

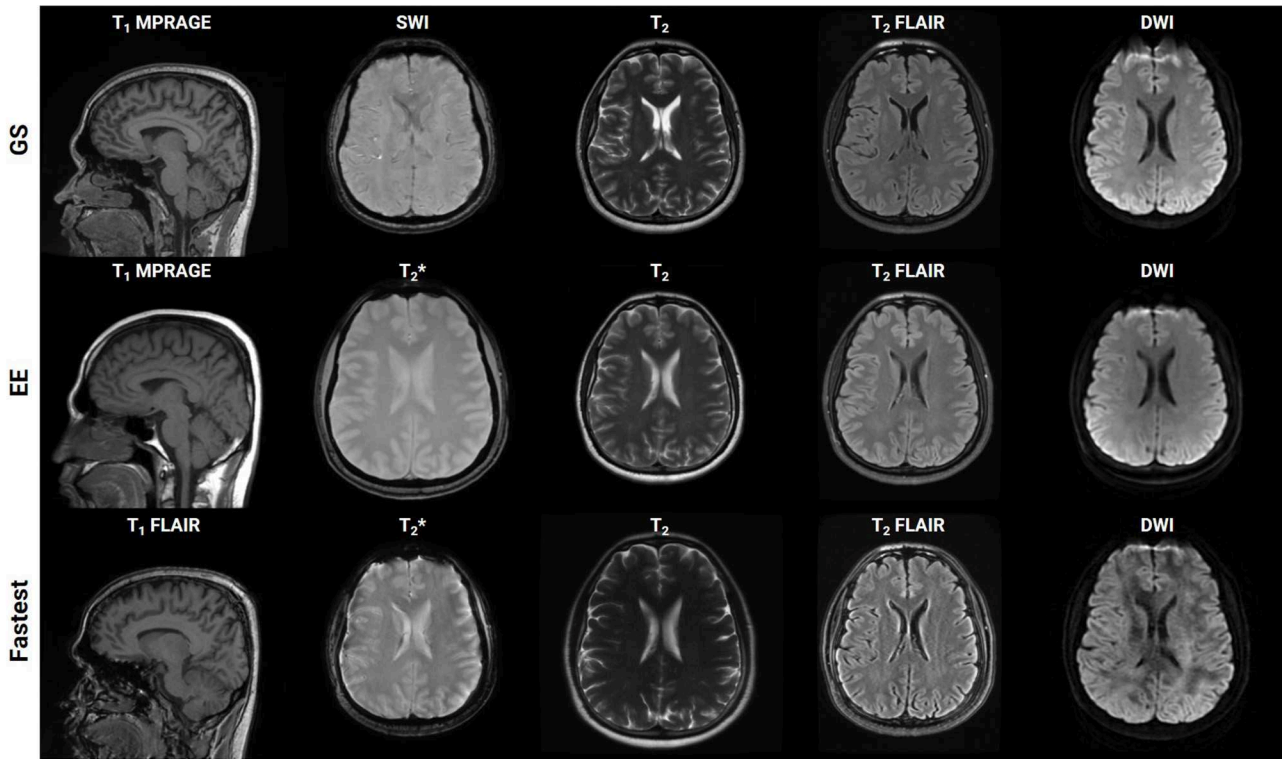

FIGURE 4
Collage of one representative slice of an arbitrarily chosen subject, across contrasts (columns), across the Gold Standard (GS), and Fastest and Expert Express (EE) protocols (rows). The subpanels have been individually windowed.

baseline denoised, and SS denoised. For the Fastest protocol, the sagittal T_1-MPRAGE sequence was accelerated by increasing the slice-thickness from 1.0 mm in the GS protocol to 1.6 mm. Overall, this resulted in increased signal intensities and decreased variance of noise in the Fastest dataset. Therefore, the median local SNR of GS data was lower than that of the Fastest data, as is expected.

3.2. Image denoising using deep learning

3.2.1. Datasets and forward-simulation

The T_1w and T_2w datasets consisted of 185/13,690 and 185/11,760 volumes/slices, respectively. For T_2^* and T_2 FLAIR, this resulted in 89/2,188 and 40/6,792 volumes/slices, respectively. Finally, the DWI dataset contained 81/2,430 volumes/slices. The final noise scaling factors determined using an iterative local SNR-guided approach were as follows. T_1: 1.5; T_2: 2.0; T_2^*: 1.0, T_2 FLAIR: 1.0, DWI: 1.5. Supplementary Figure 2 plots the maximum, minimum and mean (dashed line) local SNR values within a region of interest across the GS and forward modeled datasets, for T_1 contrast. It can be observed that the means of the GS and forward-modeled data are comparable, which validates the iterative local SNR-guided approach to determining the noise scaling factor.

3.2.2. Loss functions

Supplementary File 2 presents a tabulation of volumetric measures of denoised data obtained from the automated volumetry tool. It compares the values of data denoised using the models trained on $\beta = [0, 0.01, 0.1, 1]$. The model trained on $\beta = 1$ achieved the lowest mean RMSE value, and all subsequent models were trained with the same loss formulation.

3.2.3. Training

Supplementary Figure 3 shows a plot of training and validation losses as a function of epochs for the baseline denoising models, across contrasts. The corresponding approximate training durations are also listed. Figure 5 presents the corresponding mean changes in the image quality metrics computed on the test sets. In each instance, the model with the lowest validation loss was used. The largest gains in PSNR, MS-SSIM, and var-Lap values are observed on the T_2 contrast. Since the network architecture was common to all contrast-specific denoising models, this could be attributed to the larger noise scaling factor during the forward modeling process. Consequently, this might have forced the model to learn to denoise much noisier images during training in comparison with the training data from other contrasts. The lowest gains are observed on the T_2^* contrast.

3.2.4. SS denoising

Figure 6 plots the subject-specific median local SNR values. The subject-dependent variability of SNR validated our rationale for subject-specific denoising. Figure 7 plots the mean changes in PSNR, MS-SSIM, and var-Lap values across contrasts and

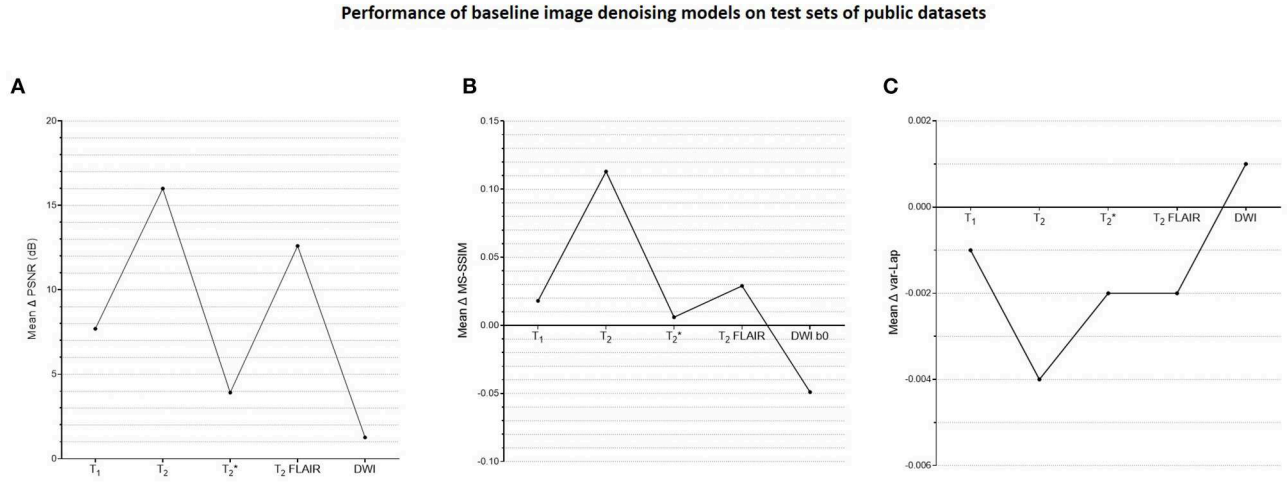

FIGURE 5
Plots of quantitative evaluations of the contrast-specific baseline image denoising models on the test sets. The means of the changes in **(A)** PSNR (dB), **(B)** MS-SSIM, and **(C)** variances of the Laplacian (var-Lap) values are reported. Lowest validation loss models were used in all instances. Higher PSNR and MS-SSIM indicate larger improvements in the image quality of the denoised images. Higher var-Lap values indicate lower loss of sharpness of the denoised images.

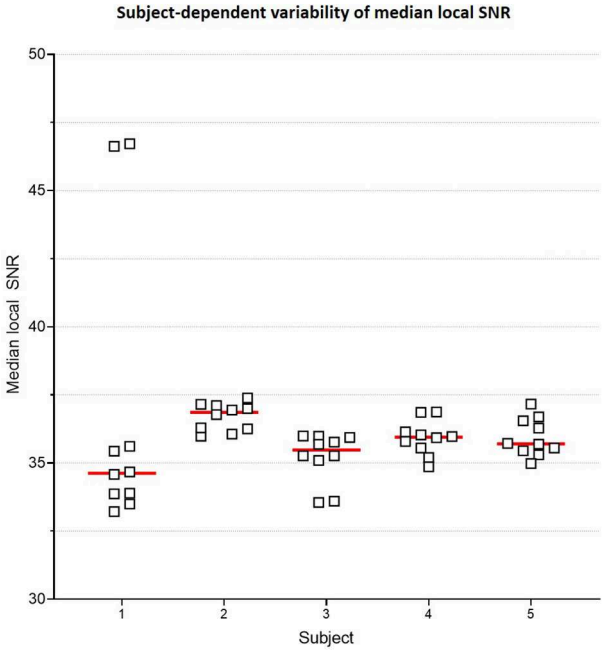

FIGURE 6
Subject-specific median local SNR. Plot comparing median local SNR values for the two T_1 MPRAGE sequences from the Gold Standard protocol, across all 5 subjects and 5 repeats. This subject-dependent variability of SNR is the premise of subject-specific denoising in this work.

subjects. Similar to the baseline denoising models, the largest gains are observed for the T_2 contrast, and the least improvement is observed for the T_2^* contrast. In addition, T_2^* SS model performed significantly poorer than the baseline model, and therefore the results have not been reported. We suspect this is due to the mismatch in the training and fine-tuning datasets. The ADNI-3 training dataset consisted of T_2^* GRE acquisitions with an echo train length (ETL) of 3. On the other hand, the Fastest protocol utilized an ETL of 1 in the T_2^* GRE acquisitions. We attribute the inherent mismatch in signal between the datasets to the poor fine-tuned performance. Additionally, this potentially indicates that the pre-processing steps in this work are inadequate.

3.3. Statistical analysis

Among the four methods for all 30 locations, 27 locations had excellent ICC (> =0.93); 2 had a good ICC (>0.8), 1 had moderate ICC (=0.651). Table 3 lists the individual ICC values for each of the 27 brain subregions and the 3 brain tissues. The 2 locations with good agreement are highlighted in bold, and the 1 location with moderate agreement is highlighted in underline.

3.4. Experiments

3.4.1. E1–Throughput

The cumulative acquisition times for the GS and Fastest protocols as per the vendor UI were 17:23 and 8:34. The practical acquisition times (obtained from the video recording) were 19:12 and 9:52. This discrepancy can be attributed to the time lost during pre-scan calibration and shimming functions. Overall, imaging one healthy volunteer using the Fastest protocol yielded a 1.94x gain in throughput over the GS protocol. Supplementary File 3 presents the timestamps and calculations of durations derived from the video recordings to obtain the final acquisition durations for this experiment. The cumulative median object-masked local SNR values for the GS and Fastest data were 243.354 and 215.767, respectively. Finally, this translates to MR Values of 0.211 and 0.364,

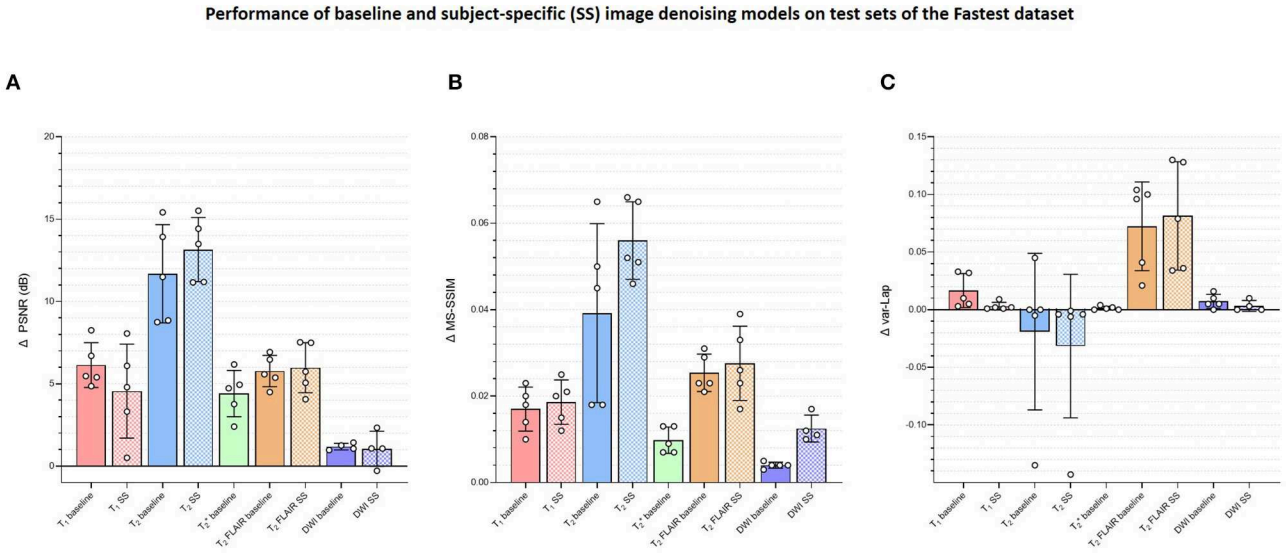

FIGURE 7
Plots comparing the quantitative evaluations of the contrast-specific baseline and subject-specific (SS) image denoising models on the test sets of the Fastest dataset. The means of the changes in **(A)** PSNR (dB), **(B)** MS-SSIM, and **(C)** variances of the Laplacian (var-Lap) values are reported. Lowest validation loss models were used in all instances. Higher PSNR and MS-SSIM indicate larger improvements in the image quality of the denoised images. Higher varLap values indicate improvements in sharpness of the denoised images.

respectively. Overall, employing the Fastest protocol resulted in a 72.51% increase in MR Value. In comparison, the corresponding SNR value for EE data was 264.136. Considering a practical acquisition duration of 07:58, this resulted in an MR Value of 0.552.

3.4.2. E2–Image quality

Figure 8 is a bar graph plotting the median object-masked local SNR and var-Lap values across contrasts, for the GS, EE, and Fastest datasets. The mean values are presented at the bottom of the individual bars. For local SNR, similar performance is observed from the axial DWI and axial T_2 FLAIR sequences, while not for the other sequences. The higher median local SNR values of the T_1 contrast from the EE protocol can be attributed to the Turbo Spin Echo-based FLAIR sequence. The GS and EE protocols also perform better than the Fastest protocol in the T_2 sequence, attributed to the longer repetition times. Overall, the EE protocol yields higher local SNR values due to higher slice thickness: 5 mm across all sequences, as opposed to ranges of 1.0–3.6 mm and 1.6–5.0 mm for GS and Fastest protocols, respectively. For var-Lap, comparable performance is observed only in the T_2 FLAIR contrast. The Fastest protocol performs worse than both GS and EE in T_1, DWI and T_2 contrasts.

3.4.3. E3–SNR recovery

Figure 7 presents the changes in median object-masked local SNR, PSNR, MS-SSIM, and var-Lap values for the baseline and SS denoising models tested on the Fastest datasets. The solid and checker boarded bars correspond to the baseline and SS denoising models, respectively. The T_1 SS denoising model does not improve PSNR over the baseline denoising model, and only modestly improves SSIM. However, it results in a smaller increase in blurriness. For T_2, the SS denoising model yields larger improvements across all metrics. For T_2 FLAIR, similar improvements are observed for PSNR and MS-SSIM, along with an undesirable increase in blurriness–indicating the model potentially denoised by primarily high-pass filtering. The T_2^* SS denoising models deteriorated image quality in every instance, and hence their results are not presented.

3.4.4. E4–Repeatability

Figure 9 presents the plots of volumetric measures obtained from the automated tool for T_1 contrast. The top row plots values of White Matter (WM) and Gray Matter (GM), and the bottom row corresponds to measures of two IDPs for AD–hippocampal and amygdala volumes. For each of these anatomies, a representative slice with the corresponding masks overlaid is illustrated in the figure inset. Figure 10 is a plot of manual volumetric measures for T_2, T_2^*, T_2 FLAIR and DWI contrasts.

3.5. Visualizing learned filters for explainable AI

Figure 11 is a collage of intermediate layer outputs obtained from denoising a DC-biased input using the baseline image denoising model for the T_1 contrast. The model appears to perform denoising (akin to low-pass filtering) in the earlier layers. Each MaxPool2D layer halves the spatial dimension, leading to reduced resolution in the later layers (refer network architecture in Figure 3A). In these layers, the model appears to be performing

TABLE 3 Individual inter-class agreement coefficient (ICC) values for each of the 27 subregions and 3 tissues.

	Subregion/tissue	ICC
1	Amygdala	0.992
2	Basal ganglia	0.971
3	Cerebellum	0.995
4	Cerebrospinal fluid	0.651
5	Frontal	0.987
6	Frontal/parietal	0.928
7	Gray matter	0.985
8	Headfat	0.951
9	Hippocampus	0.993
10	Insular	0.996
11	Left amygdala	0.99
12	Left caudate	0.996
13	Left cortical white matter	0.997
14	Left hippocampus	0.989
15	Left pallidum	0.945
16	Left putamen	**0.884**
17	Left thalamus	0.954
18	Limbic	0.987
19	Occipital	0.961
20	Parietal	0.98
21	Right amygdala	0.989
22	Right caudate	0.993
23	Right cortical white matter	0.997
24	Right hippocampus	0.991
25	Right pallidum	0.972
26	Right putamen	0.946
27	Right thalamus	0.99
28	Temporal	0.96
29	Temporal/occipital	**0.817**
30	White matter	0.992

ICC values greater than 0.9 indicate excellent agreement, values between 0.75 and 0.9 indicate good agreement (highlighted in bold), and values between 0.5 and 0.75 indicate moderate agreement (highlighted in underline).

low-frequency denoising (high-pass filtering). Overall, no brain-specific anatomy is identifiable across any of the intermediate layer outputs (as desired), potentially attributed to the patch-wise approach adopted in this work. Supplementary Figure 4 present representative layer outputs for five other threshold values: 0.50, 0.60, 0.70, 0.80, and 0.90. In all cases, a similar pattern of low-pass filtering in the earlier layers and high-pass filtering the later layers is observed. However, for 0.50, 0.60, the resulting intermediate outputs contain excessive high-frequency content (Supplementary Figures 4, 5). On the other hand, for threshold values 0.80 and 0.90, a large number of values are zeroed-out, and hence the resulting outputs do not convey any relevant information. Between threshold values 0.70 and 0.75, we chose 0.75 since we were able to better observe the denoising mechanism (Supplementary Figure 6, Figure 11).

4. Discussion and conclusion

The LUT search to accelerate the GS protocol was automated. In comparison, designing the EE protocol required human expertise and manual hours. The LUT approach is also scalable–automated recording of acquisition times and rSNR values from the vendor UI for different P_{acq} can potentially enable the construction of high-dimensional LUTs. Subsequently, high-dimensional constrained search techniques can be explored to arrive at different P_{acq}. Our LUT search formulation also allows optimizing for different criteria. We optimized for shorter acquisition durations whilst trading-off SNR. However, this can easily be modified to any other criteria by suitably modifying the weights described in Supplementary Table 1. Or, the LUT search can involve finding optimal P_{acq} that satisfies an imposed acquisition time constraint, as demonstrated in our previous work (Ravi and Geethanath, 2020; Ravi et al., 2020). Furthermore, since domain expertise is involved in setting the weights for the DOF, the LUT search is inherently explainable.

Initially, we trained our image-denoising models for T_1 and T_2 contrasts on the Human Connectome Project dataset (HCP, http://www.humanconnectomeproject.org/). Preliminary results (not reported in this work) indicated poor accuracy on the automated volumetry (high $RMSE_{vol}$), although the denoising performance was good. We attributed this to HCP data's superior image quality–HCP data were acquired on Siemens Prisma 3 T scanners with 80 mT/m gradient strength and 200 T/m/s slew rate. A 3D T_1 MPRAGE sequence was utilized with isotropic resolution and repetition/echo times = 2,530/1.15 ms. Therefore, the iterative local SNR-guided approach resulted in a higher noise scaling factor to degrade the HCP data to match the median local SNR with that of the Fastest dataset. We suspect that the denoising models trained on this data caused excessive blurring, which subsequently affected the automated T_1 volumetry. Therefore, we chose to proceed with the IXI dataset for T_1 contrast, and also for T_2 contrast to potentially mitigate a similar issue.

For T_2^*, the SS denoising models failed to demonstrate better performance than the baseline denoising models. The baseline model was trained on T_2^* data of the ADNI 3 dataset corrupted by native noise extracted from SWI data. However, fine-tuning the baseline model involved training on T_2^* data corrupted by native noise extracted from T_2^* data itself. We suspect this sequence-specific noise distribution could have impacted the training process of the SS models.

4.1. Limitations and future work

4.1.1. Intelligent protocolling

The T_1 MPRAGE sequence in the Fastest protocol achieved shorter scan durations due to higher slice thickness (1.6 vs. 1 mm). Future work could involve exploring the impact of interpolating

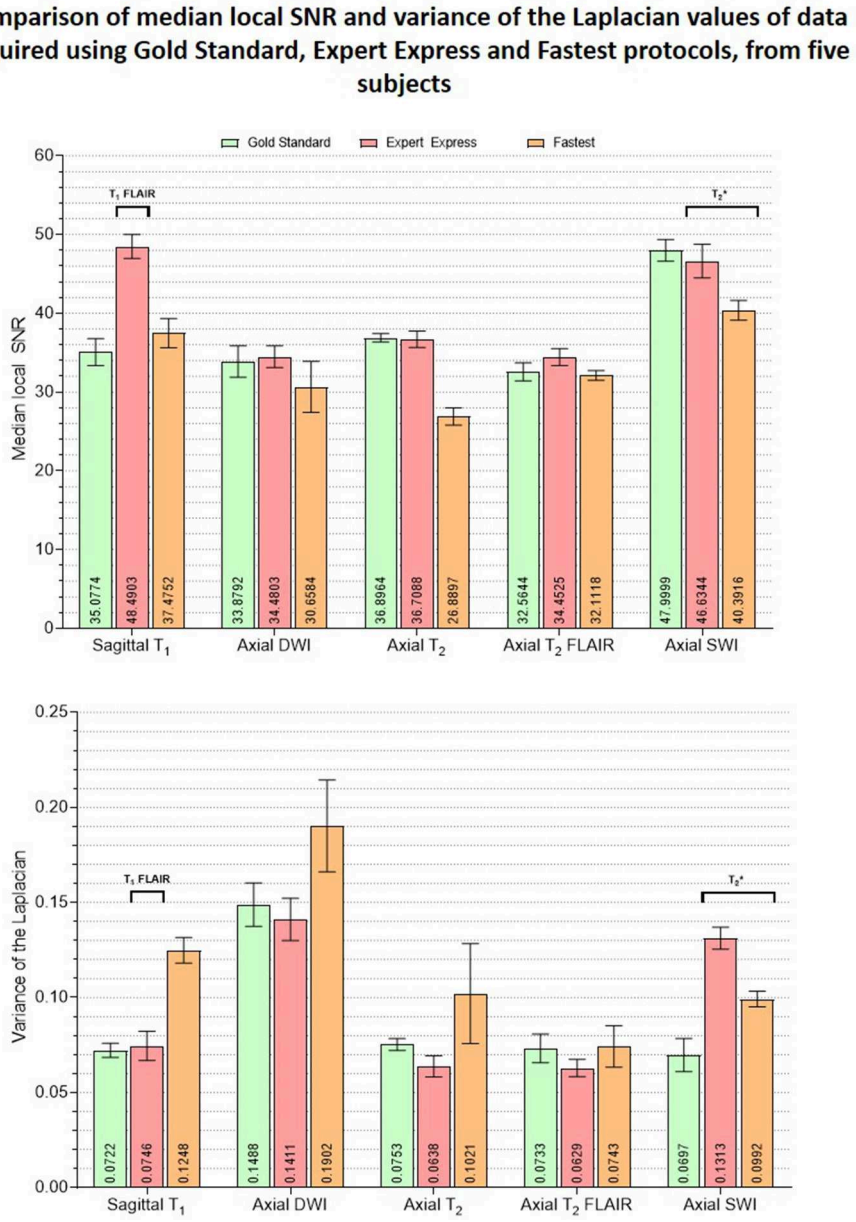

FIGURE 8
Plots comparing the median object-masked local SNR and variance of the Laplacian values computed on the GS, EE and Fastest datasets, for matched contrasts. The mean values are presented at the bottom of the individual bars. The EE protocol employed a T1 FLAIR sequence while the GS and Fastest protocols leveraged T1 MPRAGE sequences instead. Similarly, T_2^* sequences are utilized in the EE and Fastest protocols instead of a SWI sequence as in the GS protocol.

anisotropic data to achieve isotropic voxel resolutions on the accuracy of automated volumetry (Deoni et al., 2022). Although our LUT search formulation was designed to avoid modifying image contrast, the Fastest dataset marginally deviates from the GS dataset's contrast. For example, this can be observed in the T_2 FLAIR contrast in Figure 4. Potentially, Virtual Scanner (Tong et al., 2019) and its digital twinning capability (Tong et al., 2021) can be leveraged to design a physics-informed LUT optimization approach.

4.1.2. Data distribution

For detecting AD, volumetry from T_1-MPRAGE sequence is crucial. The denoising models were evaluated on a small and healthy cohort of five volunteers. Their performance on denoising pathological data has not been investigated. While we have demonstrated the value of denoising in improving the accuracy of volumetry, the robustness of the denoising models on out-of-distribution data has not been considered. A thorough evaluation will be required to assess the quality of data acquired

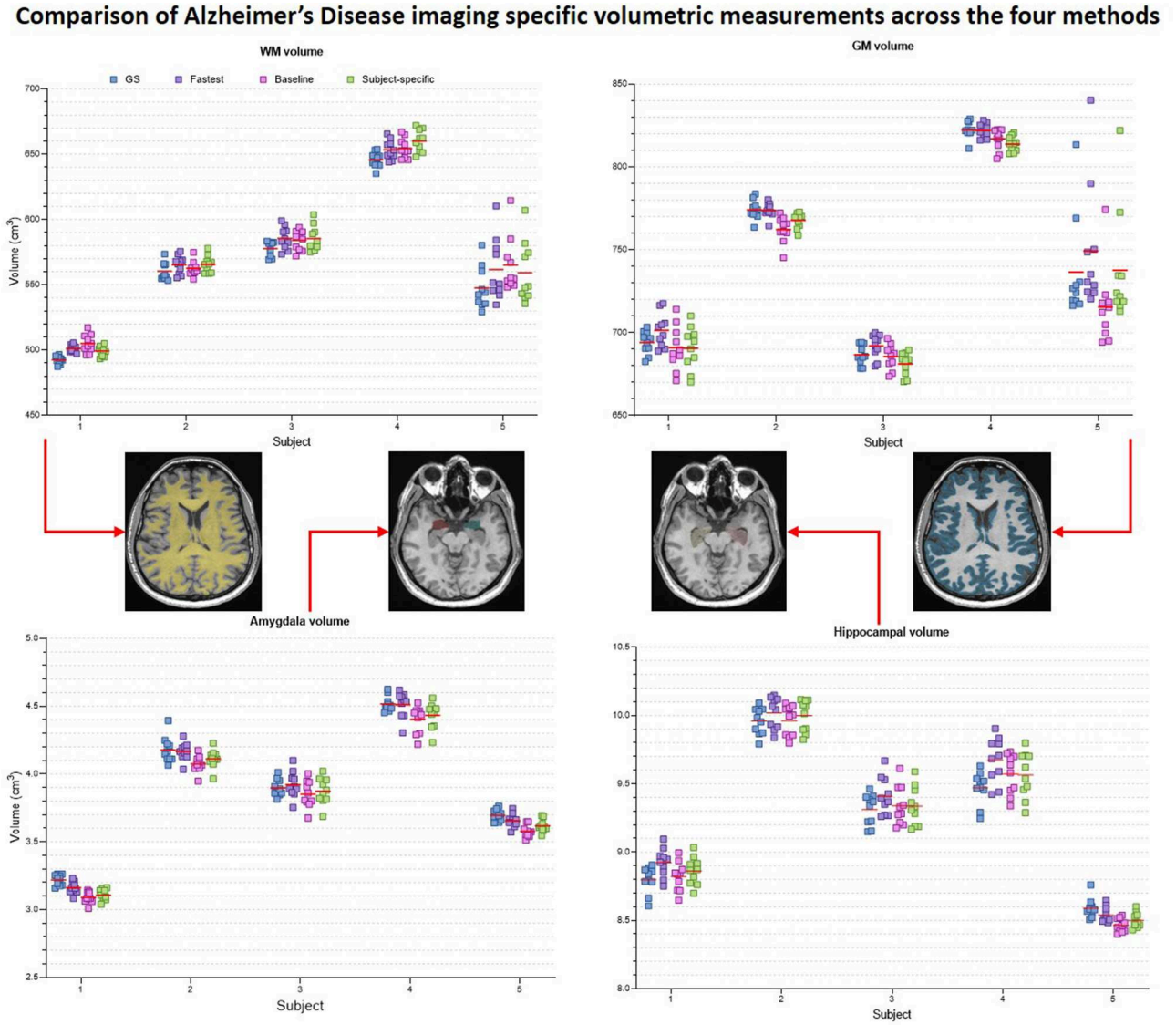

FIGURE 9
Volumetric measures of White Matter (WM), Gray Matter (GM), Amygdala and Hippocampus obtained from baseline deep learning and subject-specific deep learning (SS) denoising methods. Values across all 5 repeats are reported. Inset: Representative tissue, lobes and subregions masks.

from pathological subjects and denoising using our models. Alternatively, the training dataset could include pathological data to improve the models' generalization capabilities. Datasets which have not undergone extensive preprocessing and/or stringent quality control are valuable during the native noise extraction process of our workflow. Future iterations could involve training on a multi-site, multi-vendor, real-world dataset such as RadImageNet (Mei et al., 2022).

4.1.3. Evaluation metrics

This work utilizes a combination of referenceless (local SNR, var-Lap) and reference-based (PSNR, SSIM) image quality metrics. The referenceless metrics were borrowed from the broader computer vision community, and might not be ideal for evaluating methods in medical imaging. In particular, since the var-Lap metric is on an arbitrary scale, it does not allow performance comparisons without controlling for the testing dataset. The reference-based metrics inherently require a gold standard (GS), and hence do not lend themselves to evaluation on real-world data which, by nature, do not have reference data.

4.1.4. Inference on denoising models

While the patch-wise implementation enables flexibility of input data sizes, this approach significantly increases the inference durations-—4× on 256 × 256 input and 8× on 512 × 512 input,

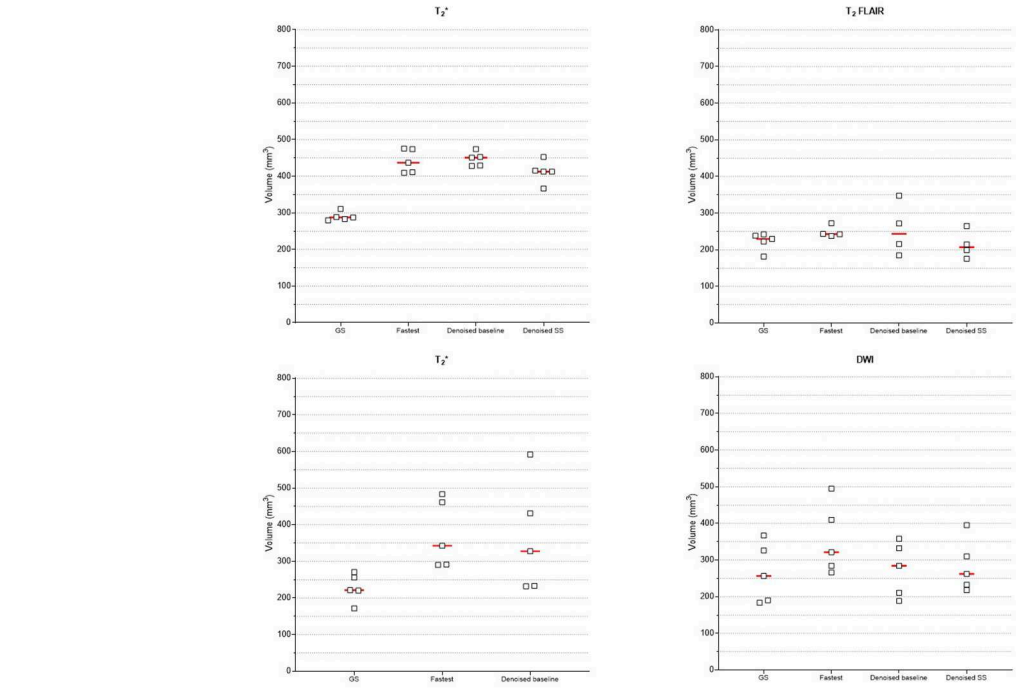

FIGURE 10
Volumetric measures of incidental benign White Matter Hyperintensity (WMH) finding from one subject. Four different raters with 3–8 years of MR experience manually measured WMH volumes in data from T_2, T_2^*, T_2 FLAIR and DWI contrasts.

when compared with full input size approaches. Furthermore, the preprocessing step of converting a full-size image into 64 × 64 patches adds an overhead that is directly proportional to the dimensions of the input image. Currently, our denoising models require approximately 3.679 seconds per slice if the input image dimensions are 512 × 512, and 2.575 seconds per slice if the input image dimensions are 256 × 256. Potentially, this could approximately be reduced 0.459 seconds and 0.321 seconds per slice, respectively, if a full input size were instead adopted. The denoising models are also not implemented in an end-to-end workflow—currently, the data needs to be transferred to a designated system *via* physical storage media. Future work will potentially involve streamlining file I/O to further accelerate DL denoising durations and packaging the pipeline to be tested for deployment at beta site.

4.1.5. Explainable AI

While there exist multiple methods that aid in interpretability classification models (Zhou et al., 2016; Selvaraju et al., 2017; Shrikumar et al., 2017; Smilkov et al., 2017), image-to-image model outputs are difficult to explain. Explainable AI techniques such Concept Activation Vectors (CAV) (Clough et al., 2019) allow probing the latent spaces of convolutional models to determine which human-friendly concepts the models are most sensitive to. However, this technique can only be applied to network architectures that include a bottleneck layer, such as U-Nets (Ronneberger et al., 2015), autoencoders [or variants thereof, such as variational-quantized autoencoders (Van Den Oord et al., 2017)]. To the best of our knowledge, there is no prior work on applying CAVs to investigate the performance of image-denoising models. Future work could involve exploring these network architectures to leverage CAVs for explainability.

4.2. Conclusion

This work demonstrates an end-to-end framework tailored for AD imaging. The framework involved implementing a LUT to shorten the acquisition duration of an existing brain imaging protocol that was employed at our institution, by sacrificing image quality. Accelerated brain imaging using this faster protocol was demonstrated, and image quality was recovered post-acquisition using DL-based image denoising models. Furthermore, MR Imaging physics dictates that the amount of signal captured relates to the volume of the subject being imaged, as this directly affects the size of the proton population. This variability of SNR depending on subject size motivated the authors to implement and demonstrate subject-specific image denoising. Code to reproduce methods, and pre-trained models will be shared upon fair request. An earlier version of code to search look-up tables is publicly available at: https://github.com/imr-framework/amri-ip/tree/ISMRM_2020.

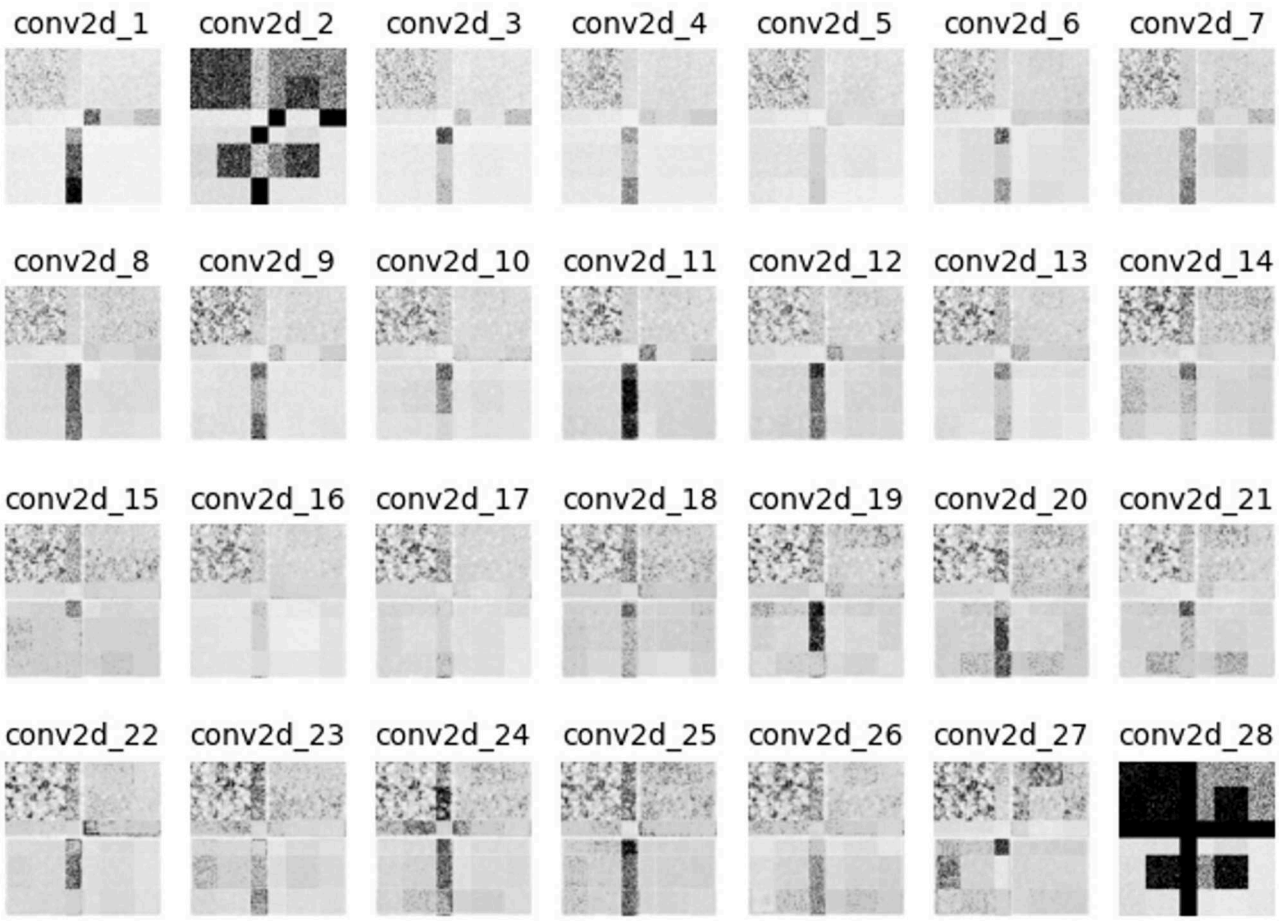

FIGURE 11
Collage of intermediate 2D convolutional layer outputs. The filter outputs from each layer were collapsed into a single image *via* maximum intensity projection. Subsequently, this image was normalized to lie in the range [0, 1.0] and then thresholded to only retain values >0.75.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

KR contributed to the conception and design of the study, performed data acquisition and method development, and wrote the manuscript and created the figures and tables. GN, NT, and ML contributed to the method development. EQ, MJ, and PP contributed to the data acquisition and method development. ZJ contributed to the design of the study and performed the statistical analysis. PQ, MF, GS, and JV contributed to the design of the study. PQ and MF contributed to the method development. SG contributed to the conception and design of the study and the method development. All authors contributed to the manuscript revision, read, and approved the submitted version.

Funding

PMX funded this work as a part of the "Autonomous MRI for Dementia" project (grant number: 22-0674-00001-01, PI: SG). PMX is responsible for the translation of research outcomes. The authors acknowledge support from the Biomedical Engineering and Imaging Institute, Dept. of Diagnostic, Molecular and Interventional Radiology, Icahn School of Medicine, Mt. Sinai, New York, NY, USA. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award

number W81XWH-12-2-0012). ADNI is 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.

Acknowledgments

The authors acknowledge grant support from PMX, Chicago, IL (22-0674-00001-01-PD, "Autonomous MRI for Dementia screening", PI: SG). Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). For up-to-date information, see www.adni-info.org.

Conflict of interest

GN, NT, ML, and GS were employed by PMX. PQ and MF were employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declare that this study received funding from PMX. The funder had the following involvement in the study: data analysis and preparation of the manuscript.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnimg.2023.1072759/full#supplementary-material

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems.* Available online at: http://tensorflow.org (accessed March 29, 2023).

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "TensorFlow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016.*

Aggarwal, K., Jimeno, M. M., Ravi, K. S., Gonzalez, G., and Geethanath, S. (2023). Developing and deploying deep learning models in brain MRI: a review. *arXiv Prepr* arXiv230101241.

Banerjee, D., Muralidharan, A., Hakim Mohammed, A. R., and Malik, B. H. (2020). Neuroimaging in dementia: a brief review. *Cureus.* 12, e8682. doi: 10.7759/cureus.8682

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., et al. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N. Engl. J. Med.* 367, 795–804. doi: 10.1056/NEJMoa1202753

Block, K. T. (2018). "Creating New Value Through Innovation," in *ISMRM Workshop on High-Value MRI.*

Clough, J. R., Oksuz, I., Puyol-Antón, E., Ruijsink, B., King, A. P., Schnabel, J. A., et al. (2019). "Global and local interpretability for cardiac MRI classification," in *Artificial Intelligence and Lecture Notes in Bioinformatics* (Lecture Notes in Computer Science). doi: 10.1007/978-3-030-32251-9_72

Commowick, O., Cervenansky, F., Cotton, F., and Dojat, M. (2021). "MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure," in *MICCAI 2021-24th International Conference on Medical Image Computing and Computer Assisted Intervention* 126.

Deoni, S. C. L., O'Muircheartaigh, J., Ljungberg, E., Huentelman, M., and Williams, S. C. R. (2022). Simultaneous high-resolution T2-weighted imaging and quantitative T 2 mapping at low magnetic field strengths using a multiple TE and multi-orientation acquisition approach. *Magn. Reson. Med.* 88, 1273–1281. doi: 10.1002/mrm.29273

Doshi-Velez, F., and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv Prepr* arXiv170208608.

Falahati, F., Fereshtehnejad, S. M., Religa, D., Wahlund, L. O., Westman, E., Eriksdotter, M., et al. (2015). The use of MRI, CT and lumbar puncture in dementia diagnostics: Data from the svedem registry. *Dement. Geriatr. Cogn. Disord.* 39, 81–91. doi: 10.1159/000366194

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., et al. (2012). 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging.* 30, 1323–1341. doi: 10.1016/j.mri.2012.05.001

Geethanath, S., Poojar, P., Ravi, K. S., and Ogbole, G. (2021). MRI denoising using native noise," in *ISMRM and SMRT Annual Meeting and Exhibition, nd.*

Geethanath, S., and Vaughan, J. T. (2019). Accessible magnetic resonance imaging: A review. *J. Magn. Reson. Imaging.* 49, e65–77. doi: 10.1002/jmri.26638

Golshan, H. M., Hasanzadeh, R. P. R., and Yousefzadeh, S. C. (2013). An MRI denoising method using image data redundancy and local SNR estimation. *Magn. Reson. Imaging.* 31, 1206–1217. doi: 10.1016/j.mri.2013.04.004

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2016.90

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods*. 18, 203–211. doi: 10.1038/s41592-020-01008-z

Jenkinson, M. (2003). Fast, automated, N-dimensional phase-unwrapping algorithm. *Magn. Reson. Med*. 49, 193–197. doi: 10.1002/mrm.10354

Kingma, D. P., and Ba, J. L. (2015). "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*. 16, 31–57. doi: 10.1145/3236386.3241340

Loktyushin, A., Herz, K., Dang, N., Glang, F., Deshmane, A., Weinmüller, S., et al. (2021). MRzero - Automated discovery of MRI sequences using supervised learning. *Magn. Reson. Med*. 86, 709–724. doi: 10.1002/mrm.28727

Mehan, W. A., González, R. G., Buchbinder, B. R., Chen, J. W., Copen, W. A., Gupta, R., et al. (2014). Optimal brain MRI protocol for new neurological complaint. *PLoS ONE*. 9, e110803. doi: 10.1371/journal.pone.0110803

Mei, X., Liu, Z., Robson, P. M., Marinelli, B., Huang, M., Doshi, A., et al. (2022). RadImageNet: An open radiologic deep learning research dataset for effective transfer learning. *Radiol. Artif. Intell*. 4, e210315. doi: 10.1148/ryai.210315

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell*. 267, 1–38. doi: 10.1016/j.artint.2018.07.007

Nair, V., and Hinton, G. E. (2010). "Rectified linear units improve Restricted Boltzmann machines," in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*.

Patterson, C. (2018). *World Alzheimer Report 2018 - The State of the Art of Dementia Research: New frontiers*. Alzheimer's Dis Int London, UK.

Pech-Pacheco, J. L., Cristóbal, G., Chamorro-Martinez, J., and Fernández-Valdivia, J. (2000). "Diatom autofocusing in brightfield microscopy: a comparative study," in *Proceedings 15th International Conference on Pattern Recognition, Vol. 3* (IEEE), 314–317.

Qian, E., Poojar, P., Vaughan, J. T., Jin, Z., and Geethanath, S. (2022). Tailored magnetic resonance fingerprinting for simultaneous non-synthetic and quantitative imaging: A repeatability study. *Med. Phys*. 49, 1673–1685. doi: 10.1002/mp.15465

Rahman, M. M. (2022). *Deep interpretability methods for neuroimaging*. Dissertation, Georgia State University.

Ravi, K. S., and Geethanath, S. (2020). Autonomous magnetic resonance imaging. *Magn. Reson. Imaging*. 73, 177–185. doi: 10.1016/j.mri.2020.08.010

Ravi, K. S., Geethanath, S., Quarterman, P., Fung, M., and Vaughan, J. T. (2020). Intelligent Protocolling for Autonomous MRI," in *ISMRM & SMRT Virtual Conference and Exhibition*.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 234–41. doi: 10.1007/978-3-319-24574-4_28

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D., et al. (2017). "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proceedings of the IEEE International Conference on Computer Vision* 618–26. doi: 10.1109/ICCV.2017.74

Shin, D., Ji, S., Lee, D., Lee, J., Oh, S. H., Lee, J., et al. (2020). Deep Reinforcement Learning Designed Shinnar-Le Roux RF Pulse Using Root-Flipping: DeepRFSLR. *IEEE Trans. Med. Imag*. 39, 4391–4400. doi: 10.1109/TMI.2020.3018508

Shrikumar, A., Greenside, P., and Kundaje, A. (2017). "Learning important features through propagating activation differences," in *34th International Conference on Machine Learning, ICML 2017*.

Silva-Spínola, A., Baldeiras, I., Arrais, J. P., and Santana, I. (2022). The road to personalized medicine in Alzheimer's Disease: the use of artificial intelligence. *Biomedicines*. 10, 315. doi: 10.3390/biomedicines10020315

Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., et al. (2011). The AddNeuroMed framework for multi-centre MRI assessment of Alzheimer's disease: Experience from the first 24 months. *Int. J. Geriatr. Psychiatry*. 26, 75–82. doi: 10.1002/gps.2491

Smilkov, D., Thorat, N., Kim, B., Viégas, F., and Wattenberg, M. (2017). Smoothgrad: removing noise by adding noise. *arXiv Prepr* arXiv170603825.

Snoek, L., van der Miesen, M. M., Beemsterboer, T., van der Leij, A., Eigenhuis, A., Steven Scholte, H., et al. (2021). The Amsterdam Open MRI Collection, a set of multimodal MRI datasets for individual difference analyses. *Sci. Data*. 8, 85. doi: 10.1038/s41597-021-00870-6

Thomas, N., Perumalla, A., Rao, S., Thangaraj, V., Ravi, K. S., Geethanath, S., et al. (2020). "Fully Automated End-to-End Neuroimaging Workflow for Mental Health Screening," in Proceedings - IEEE *20th International Conference on Bioinformatics and Bioengineering, BIBE 2020*. doi: 10.1109/BIBE50027.2020.00109

Tong, G., Geethanath, S., Jimeno, M., Qian, E., Ravi, K., Girish, N., et al. (2019). Virtual Scanner: MRI on a Browser. *J. Open Source Softw*. 4, 1637. doi: 10.21105/joss.01637

Tong, G., Vaughan, J. T., and Geethanath, S. (2021). "Virtual Scanner 2.0 enabling the MR digital twin," in *Proceedings of 2021 ISMRM and SMRT Annual Meeting and Exhibition 2021 ISMRM and SMRT Annual Meeting and Exhibition*.

Tsao, J., and Kozerke, S. (2012). MRI temporal acceleration techniques. *J. Magn. Reson. Imaging*. 36, 543–560. doi: 10.1002/jmri.23640

Van Den Oord, A., Vinyals, O., and Kavukcuoglu, K. (2017). "Neural discrete representation learning," in *Advances in Neural Information Processing Systems*.

Vernooij, M. W., and van Buchem, M. A. (2020). "Neuroimaging in dementia," in *Dis Brain, Head Neck, Spine 2020–2023 Diagnostic Imaging* 131–42. doi: 10.1007/978-3-030-38490-6_11

Walker-Samuel, S. (2019). "Using deep reinforcement learning to actively, adaptively and autonomously control of a simulated MRI scanner," in *Proceedings 27th Annual Meeting of ISMRM*.

Wang, Z., Simoncelli, E. P., and Bovik, A. C. (2003). "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, Vol. 2* (IEEE), 1398–1402.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., et al. (2017). Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's Dement*. 13, e1–85. doi: 10.1016/j.jalz.2016.11.007

Xu, J., Huang, Y., Cheng, M.-M., Liu, L., et al. (2020). Noisy-as-Clean: Learning Self-Supervised Denoising From Corrupted Image. *IEEE Trans. Image Process*. 29, 9316–9329. doi: 10.1109/TIP.2020.3026622

Zhao, H., Gallo, O., Frosio, I., and Kautz, J. (2016). Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imaging*. 3, 47–57. doi: 10.1109/TCI.2016.2644865

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). Learning Deep Features for Discriminative Localization," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2016.319

Zhu, B., Liu, J., Koonjoo, N., Rosen, B. R., and Rosen, M. S. (2018). "AUTOmated pulse SEQuence generation (AUTOSEQ) using Bayesian reinforcement learning in an MRI physics simulation environment," in *Proceedings of Joint Annual Meeting ISMRM-ESMRMB*.