Article:

Stewart, N.J. orcid.org/0000-0001-8358-394X, de Arcos, J., Biancardi, A.M. orcid.org/0009-0000-2765-0773 et al. (7 more authors) (2024) Improving xenon-129 lung ventilation image SNR with deep-learning based image reconstruction. Magnetic Resonance in Medicine, 92 (6). pp. 2546-2559. ISSN 0740-3194

https://doi.org/10.1002/mrm.30250

Received: 23 February 2024 | Revised: 26 June 2024 | Accepted: 26 July 2024

DOI: 10.1002/mrm.30250

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Improving Xenon-129 lung ventilation image SNR with deep-learning based image reconstruction

Neil J. Stewart[1,2] | **Jose de Arcos**[3] | **Alberto M. Biancardi**[1,2] | **Guilhem J. Collier**[1,2] | **Laurie J. Smith**[1] | **Graham Norquay**[1,2] | **Helen Marshall**[1,2] | **Anja C. S. Brau**[4] | **R. Marc Lebel**[5] | **Jim M. Wild**[1,2]

[1]POLARIS, Division of Clinical Medicine, School of Medicine & Population Health, Faculty of Health, The University of Sheffield, Sheffield, UK

[2]Insigneo Institute, The University of Sheffield, Sheffield, UK

[3]GE HealthCare, Little Chalfont, Amersham, UK

[4]GE HealthCare, Menlo Park, California USA

[5]GE HealthCare, Calgary, Alberta Canada

Correspondence
Neil J. Stewart, Division of Clinical Medicine, School of Medicine & Population Health, Faculty of Health, The University of Sheffield. POLARIS Building, 18 Claremont Crescent, Sheffield, S10 2TA, UK.
Email: neil.stewart@sheffield.ac.uk

Funding information
UK Research and Innovation, Grant/Award Number: MR/W008556/1; Engineering and Physical Sciences Research Council, Grant/Award Number: EP/X025187/1

Abstract

Purpose: To evaluate the feasibility and utility of a deep learning (DL)-based reconstruction for improving the SNR of hyperpolarized ^{129}Xe lung ventilation MRI.

Methods: ^{129}Xe lung ventilation MRI data acquired from patients with asthma and/or chronic obstructive pulmonary disease (COPD) were retrospectively reconstructed with a commercial DL reconstruction pipeline at five different denoising levels. Quantitative imaging metrics of lung ventilation including ventilation defect percentage (VDP) and ventilation heterogeneity index (VH_I) were compared between each set of DL-reconstructed images and alternative denoising strategies including: filtering, total variation denoising and higher-order singular value decomposition. Structural similarity between the denoised and original images was assessed. In a prospective study, the feasibility of using SNR gains from DL reconstruction to allow natural-abundance xenon MRI was evaluated in healthy volunteers.

Results: ^{129}Xe ventilation image SNR was improved with DL reconstruction when compared with conventionally reconstructed images. In patients with asthma and/or COPD, DL-reconstructed images exhibited a slight positive bias in ventilation defect percentage (1.3% at 75% denoising) and ventilation heterogeneity index (~1.4) when compared with conventionally reconstructed images. Additionally, DL-reconstructed images preserved structural similarity more effectively than data denoised using alternative approaches. DL reconstruction greatly improved image SNR (greater than threefold), to a level that ^{129}Xe ventilation imaging using natural-abundance xenon appears feasible.

Conclusion: DL-based image reconstruction significantly improves ^{129}Xe ventilation image SNR, preserves structural similarity, and leads to a minor bias in ventilation metrics that can be attributed to differences in the image sharpness. This tool should help facilitate cost-effective ^{129}Xe ventilation imaging with natural-abundance xenon in the future.

KEYWORDS

deep learning, hyperpolarized 129Xe, image reconstruction, lung

1 | INTRODUCTION

SNR is a significant limiting factor in hyperpolarized ^{129}Xe imaging of lung ventilation, impacting both visual interpretation and quantitative analysis of functional metrics such as the ventilation defect percentage (VDP).[1] Low SNR in hyperpolarized ^{129}Xe imaging can result from several factors: low ^{129}Xe polarization; inability of patients (especially those with severe lung disease) to completely inhale the gas dose and/or maintain breath-hold; unpredictable logistical delays such as patient arrival at the MR unit, and set-up time at the scanner, which can lead to T_1 decay of pre-prepared doses of polarized ^{129}Xe. To maximize SNR, most human ^{129}Xe imaging studies use 129-enriched xenon (>85% ^{129}Xe), which provides approximately threefold SNR benefits over natural-abundance xenon (26% ^{129}Xe)[2]; however, the enrichment process is expensive, resulting in an approximate five- to tenfold increase in cost per liter of xenon. Moreover, doses that do not satisfy a certain polarization (or dose equivalence)[3] requirement before administration, as defined by ethical constraints or FDA guidelines, may have to be discarded.[4]

As such, methods to improve ^{129}Xe image SNR are highly desirable, both economically and for quantitative clinical interpretation. To date, efforts to improve the SNR of hyperpolarized ^{129}Xe images have largely been focused on: optimization of polarizer hardware,[5–7] development of high-sensitivity radiofrequency coils,[8] and implementation of efficient acquisition trajectories that require fewer RF pulses to encode, and therefore, allow higher flip angles to be used.[9–11] With the exception of compressed sensing—which has been used with a primary motivation of image acceleration rather than denoising[11]—image reconstruction/postprocessing techniques have not been fully explored for improving SNR in hyperpolarized ^{129}Xe imaging. However, a recent report of the denoising of ^{19}F images with a low-rank matrix recovery with optimal shrinkage of singular value approach[12] may also hold promise for ^{129}Xe imaging.

In recent years, artificial intelligence (AI) methods have been used in hyperpolarized gas image segmentation[13,14] and analysis,[15] but the potential for denoising images has been underexplored.[16] A deep convolutional neural network (CNN)-based image reconstruction tool that acts on raw MR data and produces images with increased sharpness and reduced noise[17] has been recently commercialized and applied to ^{1}H imaging of various organs. In particular, deep learning (DL) reconstruction led to higher SNR, CNR, and image quality scores[18,19] in the prostate, higher sharpness and contrast in the liver,[20] and improved diagnostic performance in the pituitary gland[21] and lumbar spine.[22]

The purpose of this work was to assess the feasibility and utility of applying this commercially available DL reconstruction method developed for ^{1}H imaging to hyperpolarized ^{129}Xe lung ventilation images, and to evaluate whether the fidelity of quantitative metrics including VDP and ventilation heterogeneity index (VH_I) are preserved.

2 | METHODS

2.1 | Datasets—retrospective

^{129}Xe lung ventilation MRI data acquired from patients with asthma and/or chronic obstructive pulmonary disease (COPD) under a National Research Ethics Committee approved protocol (16/EM/0439) were retrospectively re-processed. Datasets (n = 34) with a range of SNR values (median [range]; 29.3 [8.66, 38.0]) were randomly selected from a database of >100 MRI examinations with SNR <40 performed between 2020 and 2022. These data were acquired from patients with median (range) age of 57.5 (30.7, 82.2) years; sex: 17 female (F), 17 male (M); diagnosis: asthma, n = 17, asthma + COPD, n = 13, COPD, n = 4; physician-assigned severity: mild, n = 8, moderate, n = 17, severe n, = 9; and forced expiratory volume in 1 s (FEV_1) z-score of −0.83 (−3.77, 2.66).

2.2 | Datasets—prospective

To explore the effect of DL reconstruction on low SNR images, and the feasibility for low-dose natural-abundance xenon MRI, n = 3 healthy volunteers (age: 23, 29, 36 years; sex: 1 F, 2 M) were scanned with a 50:50 mix of natural-abundance xenon and N_2. One of these healthy volunteers was also scanned with the usual dose of 129-enriched xenon described below, but with increased spatial resolution. A simplified diagram summarizing the datasets included in the two sub-studies is included in Figure S1.

2.3 | MRI acquisition

In all cases, a gas dose comprising a 50:50 mix of xenon (^{129}Xe polarization ~25%[5]) and N_2 of total volume 1 L (or less depending on patient height, see Chan et al.[23]) was inhaled from functional residual capacity and breath-hold was maintained while images were acquired as described below. For the retrospective study, 129-enriched xenon was used (~86% enrichment), and for the prospective study, either 129-enriched or natural-abundance (26% ^{129}Xe) xenon was used.

A 1.5 T GE Healthcare MR scanner (HDx for retrospective study, HDx or 450w for prospective study) was used in combination with a flexible transmit-receive vest coil (Clinical MR Solutions). ^{129}Xe ventilation images were acquired using a 3D balanced steady state sequence[2] with in-plane FOV between 36 and 48 cm, matrix 100×100 (in-plane resolution 3.6–4.8 cm^2) with partial phase FOV, slice thickness 10 mm, flip angle −10°, bandwidth 16 kHz or 32 kHz. For the prospective 129-enriched xenon acquisition, the slice thickness was halved (5 mm) and flip angle decreased (∼7°), as optimum for the phase encoding resolution.[2]

To enable calculation of the VDP, ^{1}H anatomical images of the lungs were acquired during a separate breath-hold after inhalation of a bag of air of equivalent volume. Imaging parameters were: 3D spoiled gradient recalled acquisition in steady state (SPGR) sequence with FOV matching that of ^{129}Xe imaging, slice thickness 5 mm, flip angle ∼5°, and bandwidth ±83.3 kHz.

2.4 | Image reconstruction

The DL reconstruction pipeline—commercially available as AIR Recon DL (GE Healthcare)—was trained using a supervised learning approach with pairs of low-noise, high resolution images, and typical noisy, lower resolution counterparts across a broad range of image content to enable generalizability of the model across anatomies.[17] This training was performed previously as part of the product optimization, a process that is not related to the present study. The training data had not included any ^{129}Xe lung images, and no additional training was performed for the purpose of this study. In addition to reducing image noise, the model also improves sharpness and suppresses truncation artifacts, having an integrated "de-ringing" step that is independent from the denoising step. The 3D Cartesian pipeline of this DL model was applied as is without any adjustment or retraining to our 3D ^{129}Xe raw k-space data. Images were generated using an offline version of the product reconstruction pipeline at denoising levels of 0 (no denoising), 25%, 50%, 75%, and 100% (a.k.a. 0, 0.25, 0.5, 0.75, 1.0), with de-ringing active in all cases. In addition, images were generated without de-ringing for all denoising levels (including 0) in attempt to isolate the effects of the two independent processing steps. (Unless explicitly specified, DL denoised images reported in the following include de-ringing).

The resulting images were compared to several alternative de-noising pathways: (1) k-space filtering using a 2D Hamming window with periodicity defined by the image size (this acts in addition to the GE Healthcare product reconstruction pipeline's Fermi filter); (2) an image-based total variation (TV) denoising approach using FASTA (fast adaptive shrinkage/thresholding algorithm) (total variation regularization parameter, $\lambda = 0.03$)[24]; (3) higher-order singular value decomposition (HOSVD) denoising[25] applied to the 3D images post-reconstruction (using a 3D tensor with rank [40 40 13] on the DICOM images of size [256 256 number of slices], with forward and inverse variance-stabilizing transformation[26] applied pre- and post-HOSVD, respectively).

2.5 | Image analysis

For all sets of reconstructed ^{129}Xe ventilation images, quantitative metrics of lung ventilation, namely; VDP,[27] coefficient of variation (CV),[28] and its interquartile range, the ventilation heterogeneity index (VH_I),[29] were calculated for comparison with conventional images. In brief, ^{1}H structural images were co-registered to the ^{129}Xe images; then, the image pairs (^{129}Xe and ^{1}H) were used as inputs to an in-house CNN-based segmentation algorithm, developed from Bertin et al.,[30] to segment the lung cavity and main airways, followed by a manual editing stage. After N4 bias correction and normalization, the ^{129}Xe signal was binned into four bins (defect, low, normal and hyper ventilation; see Figure 1),[31,32] and VDP was calculated as the ratio of the volume of pixels with signal in the "defect" bin to the volume of the lung cavity from the ^{1}H image. Local CV was calculated slice-by-slice, by first subsampling the images by 50% and then sliding a (nearest-neighbor) 3×3 voxel kernel across the images, calculating the signal variation within the kernel at each step.[28,33] The ventilation heterogeneity index is the interquartile range of the resulting CV distribution.[29] Unless otherwise specified, the lung cavity masks derived from the original ^{129}Xe images (conventional reconstruction) were used to process all other ^{129}Xe images (i.e., those reconstructed using the DL model and alternative denoising pathways). However, in a sub-analysis, lung cavity masks were also derived from the DL 0.75 images, and VDP and VH_I were re-calculated.

Bland–Altman analyses were performed to identify differences in VDP and VH_I across different reconstruction approaches. In addition, SNR was computed from the DICOM images as the ratio between the estimated signal and the estimated noise. The signal was estimated as the average over a region identified by thresholding at a level selected as proposed by Ridler and Calvard.[34] The noise was estimated as the most frequent occurrence when fitting a Rayleigh function (a special case of a Rician estimation[35] with an expected zero signal) on a sliding window spanning the whole image. The images included in the retrospective study represent a range of SNR values,

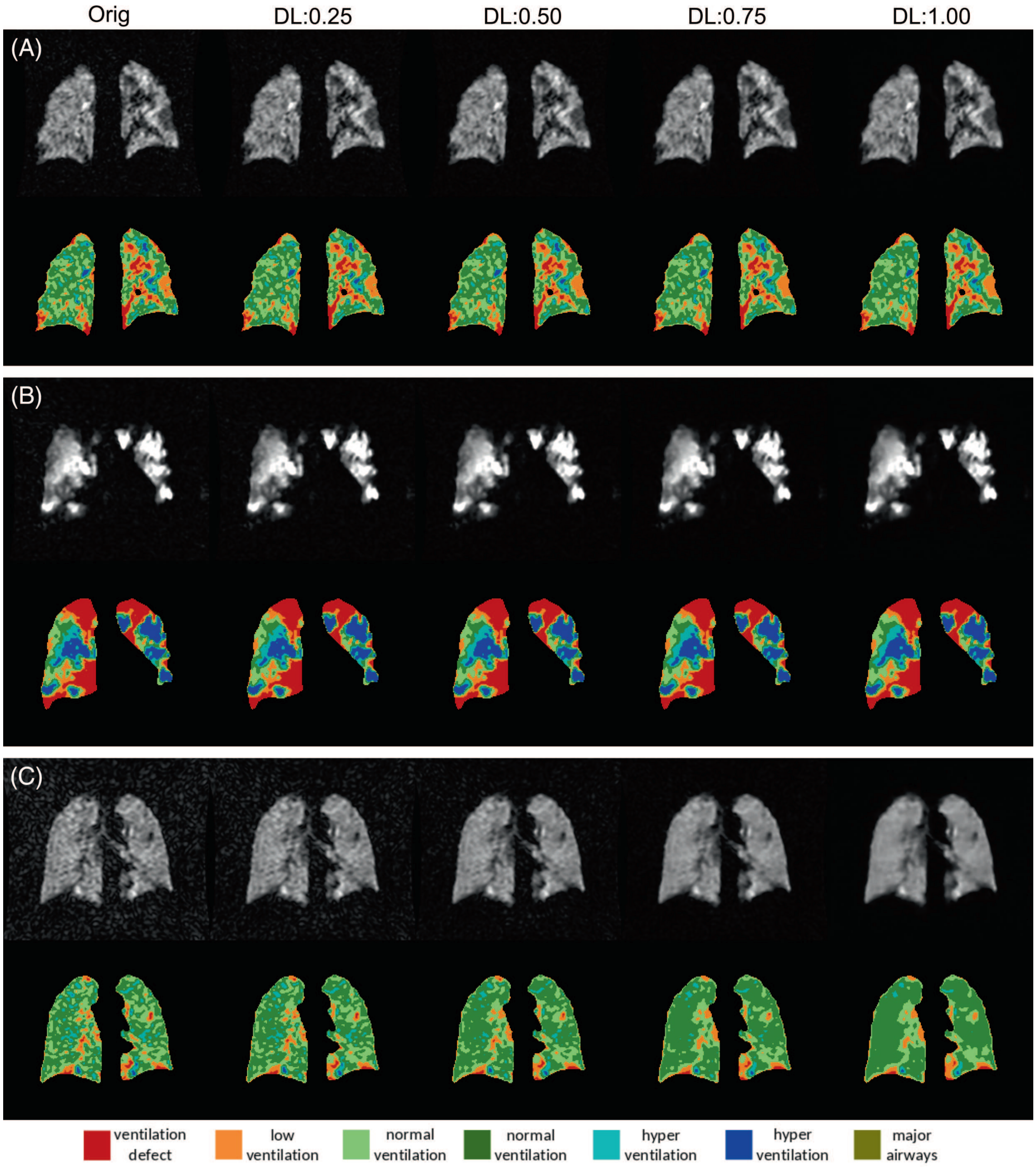

FIGURE 1 (Retrospective study) From left to right: Conventional (Orig) reconstruction, deep learning (DL) reconstruction at incremental denoising levels from 0.25 to 1.0 (de-ringing active in all cases) for ventilation images and their corresponding binning maps. (A) Patient with severe asthma. (B) Patient with severe asthma + chronic obstructive pulmonary disease. (C) Patient with moderate asthma.

however, to explore the effect of SNR on the DL recon and derived metrics within one dataset, complex Gaussian noise of increasing SD was added to the raw k-space data before reconstruction for one dataset with high baseline SNR. Finally, the structural similarity index measure (SSIM)[36] was evaluated for each set of denoised images to quantify similarity to the images reconstructed using the conventional pipeline. SSIM was calculated for three regions; (1) the entire image; (2) the region containing the lungs and airways only (i.e., lung cavity mask + airway mask); and (2) the background (i.e., the difference between regions [1] and [2]). Metrics are presented as mean (±SD) or median (range) depending on whether data followed a normal distribution (Shapiro–Wilk test).

3 | RESULTS

3.1 | Retrospective study

Example ^{129}Xe ventilation images obtained from two subjects with severe disease reconstructed using the scanner's

TABLE 1 (Retrospective study) Quantitative metrics of lung ventilation (ventilation defect percentage, VDP and ventilation heterogeneity index, VH_I) for different reconstruction methods ~~(retrospective dataset)~~.

Denoising method	VDP (%)	VDP diff: method—original	VH_I	VH_I diff: method—original	Apparent SNR[a]
Original (none)	3.8 (6.3)	–	10.5 (3.2)	–	29.3 (8.7, 38.0)
DL DN (0%) + DR	5.2 (6.4)	1.22 (0.01, 2.43)	12.5 (4.5)	1.44 (0.18, 2.70)	31.7 (9.1, 52.3)
DL DN (25%) + DR	5.2 (6.4)	1.25 (−0.04, 2.54)	12.5 (4.4)	1.41 (0.05, 2.77)	42.1 (11.8, 67.6)
DL DN (50%) + DR	5.2 (6.5)	1.28 (−0.09, 2.66)	12.5 (4.3)	1.40 (0.00, 2.80)	61.5 (17.4, 101)
DL DN (75%) + DR	5.2 (6.5)	1.31 (−0.15, 2.77)	12.6 (4.2)	1.41 (0.00, 2.83)	123 (34.8, 192)
DL DN (100%) + DR	5.2 (6.5)	1.34 (−0.19, 2.88)	12.6 (4.1)	1.43 (0.02, 2.83)	933 (284, 1580)
DL DN (0%) (no DR)	4.2 (5.7)	0.06 (−1.11, 1.23)	11.4 (4.0)	0.35 (−0.11, 0.81)	31.2 (8.9, 51.7)
DL DN (25%) (no DR)	4.2 (5.8)	0.11 (−0.96, 1.18)	11.3 (4.0)	0.33 (−0.19, 0.86)	40.6 (11.4, 67.5)
DL DN (50%) (no DR)	4.2 (5.8)	0.14 (−0.89, 1.18)	11.4 (3.9)	0.34 (−0.25, 0.86)	60.6 (17.0, 98.7)
DL DN (75%) (no DR)	4.2 (5.8)	0.18 (−0.83, 1.20)	11.4 (3.8)	0.37 (−0.25, 0.98)	120 (33.9, 197)
DL DN (100%) (no DR)	4.3 (5.8)	0.22 (−0.83, 1.26)	11.4 (3.5)	0.41 (−0.22, 1.04)	874 (277, 1484)
Filtering	3.3 (4.6)	−0.95 (−3.46, 1.56)	9.2 (3.0)	−2.02 (−4.16, 0.12)	43.2 (12.4, 71.0)
TV	3.9 (4.2)	−0.95 (−4.78, 2.87)	9.9 (3.8)	−1.64 (−4.11, 0.83)	33.5 (9.4, 55.8)
HOSVD	3.7 (5.2)	−0.46 (−2.67, 1.75)	10.5 (3.5)	−0.94 (−2.25, 0.38)	37.1 (9.3, 438)

Note: All metrics are reported as median (interquartile range), alongside Bland–Altman metrics—mean difference (i.e., bias [± 1.96 SDs])—of the difference between values for each method and the original (conventional manufacturer pipeline processed) images.
Abbreviations: diff, difference; DL DN, deep learning based denoising; DR, de-ringing ("on" by default); HOSVD, higher-order singular value decomposition; Orig, original (conventional scanner manufacturer pipeline); TV, total variation.
[a]Because of the alterations to the signal and noise introduced by DL processing and other denoising methods, the SNR values quoted are only considered "true" SNR for the original reconstruction.

conventional pipeline ("Orig") and at all four DL denoising levels (0.25, 0.5, 0.75, and 1.0) are shown in Figure 1. Qualitatively, across all datasets, the image SNR and sharpness were improved by the DL reconstruction, whereas the physiological ventilation distribution remained visually unchanged. Alongside the raw images, binning maps of the ventilation distribution are shown, which qualitatively illustrate that in addition to ventilation defect regions, areas of low, normal and hyper ventilation are generally preserved by the DL reconstruction. Quantitatively, the median VDP was increased for DL-reconstructed images when compared with conventional images, as summarized in Table 1. However, when the de-ringing step was removed, the VDP values were considerably closer to the original values. A denoising level of 0.75 was chosen empirically to provide a good balance between optimal SNR improvement and relatively low bias in VDP. (Moreover, the highest denoising level of 1.0 is not available prospectively). In Figure 2, images reconstructed using the "Orig" are shown alongside images reconstructed at the empirically optimal DL denoising level of 0.75 (75%) and difference images to highlight the regions of the image that are most significantly affected by the denoising (and de-ringing) pipeline.

Figure 3 shows a Bland–Altman plot for the DL reconstructed images at a denoising level of 0.75 (including de-ringing), with a positive bias toward increased VDP for the DL images of 1.31%; this was reduced to 0.18% when de-ringing was removed (Figure S2). The difference in VDP was generally lower at lower VDP; the bias calculated for datasets with mean VDP <10% was reduced to 1.15%. When re-calculated using the masks generated from the DL (0.75) images, the mean bias in VDP was reduced to 0.88% (Figure S3). The VDP values derived from filtering, TV denoising and HOSVD were all lower than those of the original images, with a smaller mean difference than that of the DL reconstructed images (Table 1).

The VH_I was also increased for DL-reconstructed images when compared with original images (Table 1), with a positive bias of ~1.4 for all denoising levels. Removing the de-ringing pipeline reduced the positive bias to 0.3 to 0.4 for all denoising levels. When re-calculated using the masks generated from the DL (0.75) images, the mean bias in VH_I was reduced to 0.98. In contrast, VH_I values were lower for all other denoising methods compared with those of the original images, with HOSVD providing the closest agreement with the original values. Figure 3 shows a Bland–Altman plot for the DL reconstructed images at

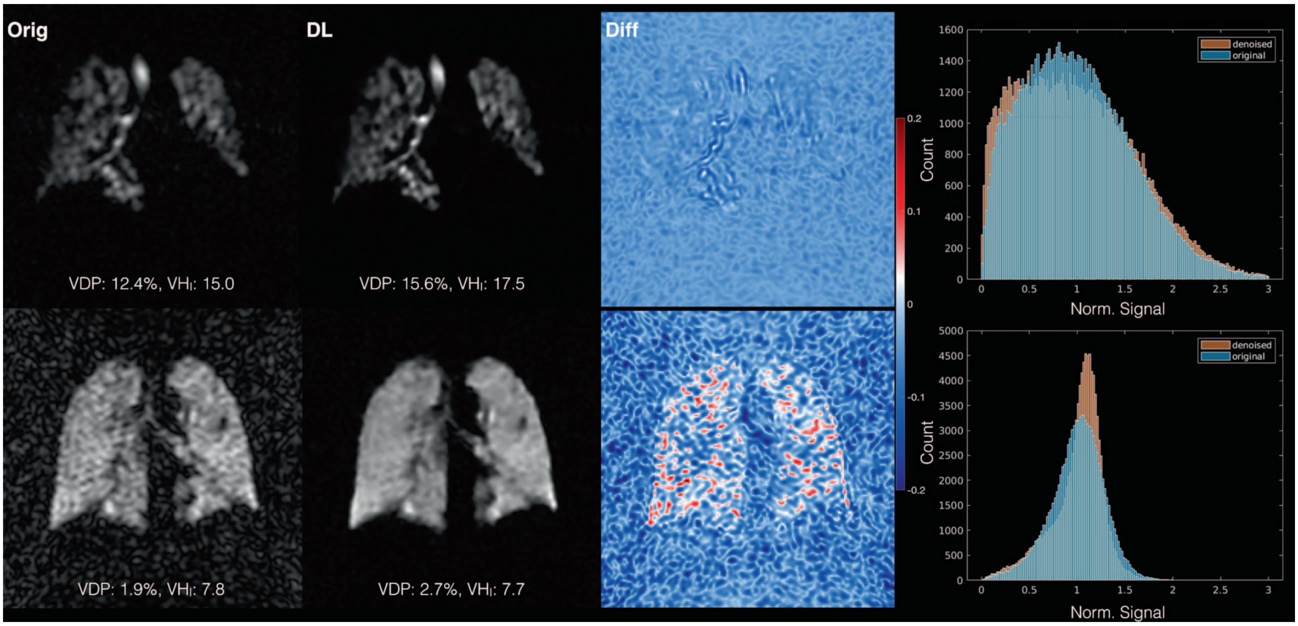

FIGURE 2 (Retrospective study) From left to right: Conventional (Orig) reconstruction, deep learning (DL) reconstruction (denoising level:0.75 + de-ringing), "difference image" calculated as the subtraction of the former from the latter after normalization, and normalized histograms of the pixel intensities for conventional and DL reconstructed images, for a patient with asthma + chronic obstructive pulmonary disease (top) and a patient with asthma (bottom; same raw data as in Figure 1C); in both cases physician-assigned disease severity was moderate. Ventilation defect percentage (VDP) and ventilation heterogeneity index (VH_I) are noted for each image. The color bar has been chosen to accentuate the signal/noise differences. The SSIM in the lung and airway region for the case in the top row was 0.931 and for the case in the bottom row was 0.865.

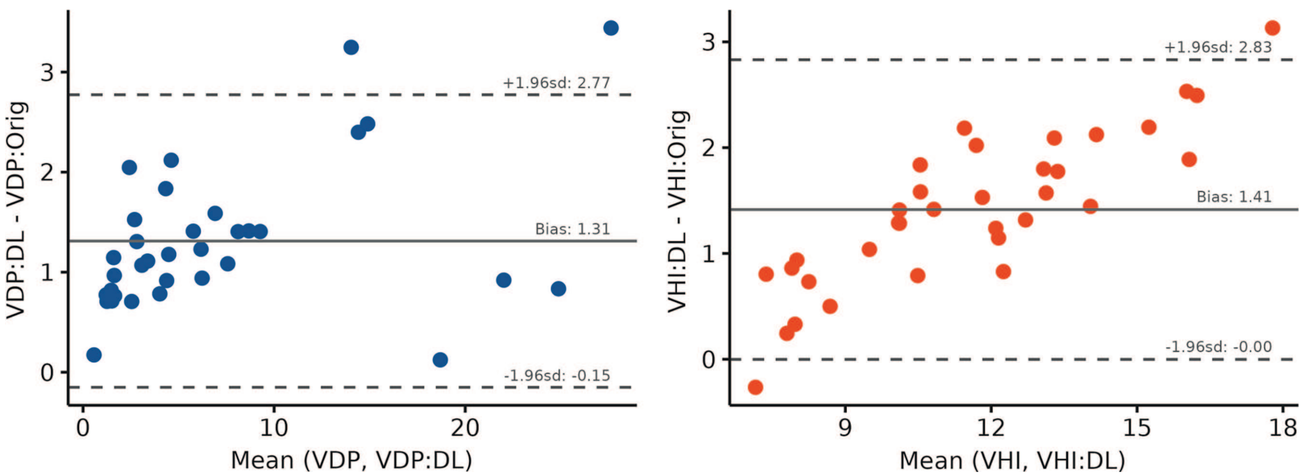

FIGURE 3 (Retrospective study) Bland–Altman plots of the difference in ventilation defect percentage (VDP) and ventilation heterogeneity index (VH_I) as a function of their mean values, calculated from conventionally reconstructed images versus deep learning (DL)-reconstructed (denoising level:0.75 + de-ringing) images for n = 34 patients with asthma and/or chronic obstructive pulmonary disease.

a denoising level of 0.75, indicating a positive bias toward increased VH_I for the DL images of 1.41.

Median SSIM values are reported in Table 2. For DL reconstructed images, the SSIM decreased in all three regions of assessment (entire image, lung and airway region only, and background only) with increasing denoising level. Evaluating SSIM across the whole 3D image set, the SSIM was lower for DL reconstructed images with high denoising levels (≥0.5) than all alternative denoising techniques. When evaluating SSIM for the lung and airway region and background region separately, it was found that the SSIM was well preserved in the lung and airway region for all denoising levels (SSIM ≥0.95), whereas a sharp decrease in SSIM of the background was observed

TABLE 2 (Retrospective study) Structural similarity of images reconstructed using the DL model and alternative pipelines, evaluated in comparison with the original images.

	SSIM		
Denoising method	**Entire image**	**Lungs + airways**	**Background**
DL DN (0%) + DR	0.992 (0.007)	0.972 (0.012)	0.988 (0.006)
DL DN (25%) + DR	0.962 (0.027)	0.972 (0.014)	0.957 (0.025)
DL DN (50%) + DR	0.881 (0.060)	0.968 (0.016)	0.874 (0.055)
DL DN (75%) + DR	0.753 (0.099)	0.963 (0.018)	0.739 (0.102)
DL DN (100%) + DR	0.620 (0.145)	0.954 (0.023)	0.598 (0.151)
DL DN (0%)	0.997 (0.002)	0.989 (0.008)	0.995 (0.002)
DL DN (25%)	0.973 (0.015)	0.989 (0.007)	0.970 (0.014)
DL DN (50%)	0.900 (0.051)	0.986 (0.010)	0.895 (0.046)
DL DN (75%)	0.778 (0.102)	0.981 (0.013)	0.766 (0.100)
DL DN (100%)	0.637 (0.141)	0.971 (0.017)	0.616 (0.153)
Filtering	0.953 (0.028)	0.921 (0.041)	0.935 (0.029)
TV	0.936 (0.039)	0.931 (0.015)	0.921 (0.031)
HOSVD	0.927 (0.058)	0.900 (0.075)	0.889 (0.046)

Note: Metrics are quoted as a median (interquartile range) across all subjects. Denoising level of 0% refers to data processed through the DL pipeline with a denoising level of 0% (i.e., not necessarily identical to "original" data i.e., that processed through the manufacturer stock [non-DL] pipeline).
Abbreviations: DL DN, deep learning based denoising; DR, de-ringing ("on" by default); HOSVD, higher-order singular value decomposition; TV, total variation.

with increasing denoising level. For DL 0.75, the SSIM in the lung and airway region was 0.963. For all denoising levels, a marginal, but significant increase in SSIM was observed when the de-ringing step was removed; this increase was exhibited in all three regions of assessment, indicating increased similarity to the original images as a whole. The SSIM values for DL images at all denoising levels in the lung and airway region were higher than those of any of the alternative denoising techniques in the same region.

Example images reconstructed using each alternative denoising technique is shown in Figure 4 alongside histograms of the corresponding signal intensity in the lung region. The DL (0.75) image histogram is the most similar in shape to that of the TV denoising technique; both exhibit a sharper fall-off compared with the histograms obtained from other techniques. On the other hand, the histograms generated from the filtering and HOSVD denoising methods are most similar in shape to that of the original images.

The apparent SNR of the reconstructed images was observed to increase in a non-linear manner with increasing denoising level, with a trend toward slightly higher SNR with de-ringing compared with no de-ringing (Table 1). (Although the SNR is deemed to be true SNR for the original images, the DL reconstruction alters the noise profile such that SNR values calculated after DL reconstruction are considered as "apparent" SNR values in this manuscript). Figure 5 shows the effect of adding Gaussian noise to reduce the baseline SNR before running the DL recon. At a denoising level of 0.75, the original SNR (32.0) is boosted approximately fourfold by DL recon, and the original SNR is recoverable from a baseline SNR of ~8 (Figure 5A). The SSIM in the lung and airway region was >0.8 for all DL denoising levels down to a baseline SNR ~10 (Figure 5B). VDP and VH_I biases were fairly constant for SNR >10, but SNR <10 caused the values to vary considerably. Concurrent with other observations, the VDP and VH_I were closer to that of the original images for the images processed without de-ringing.

3.2 | Prospective Study

^{129}Xe ventilation images acquired from two healthy subjects using a 50% natural abundance xenon inhaled gas mixture and reconstructed using the conventional pipeline exhibited borderline SNR acceptability criteria for clinical use, however, application of the DL reconstruction led to a greater than threefold increase in SNR (see examples in Figure 6). Images acquired with 129-enriched xenon with conventional sequence parameters, natural abundance xenon with the same parameters, and 129-enriched xenon

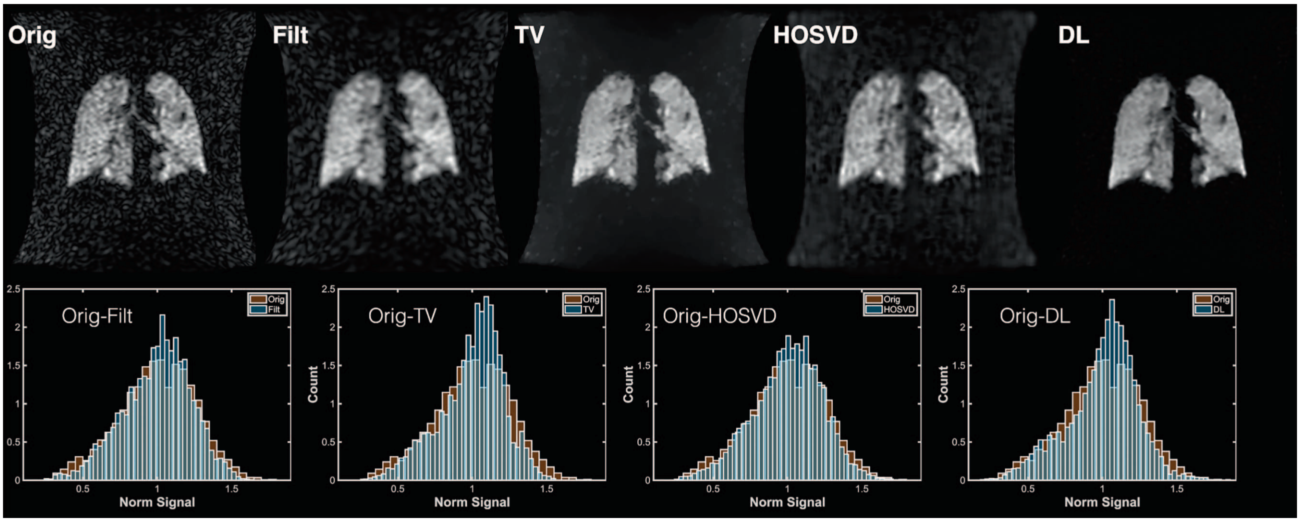

FIGURE 4 (Retrospective study) Images reconstructed using different denoising pipelines for a patient with moderate asthma (same raw data as Figure 2). Below each image, histograms of the pixel intensities within the lung mask region are shown overlaid on those of the original images. The dark bands on the left and right of the images arise from the "gradwarp" gradient non-linearity correction to the image. This is applied after denoising and creates 0 value pixels in these bands that are distinct from the true noise (>0). For the DL image, the gradwarp correction is incorporated into the DL pipeline and the noise level is even closer to zero in value; as such these bands are less noticeable. Orig, original (conventional scanner manufacturer pipeline); Filt, additional filtering; TV, Total variation denoising; HOSVD: Higher-order singular value decomposition; DL, deep learning based denoising level:0.75 + de-ringing.

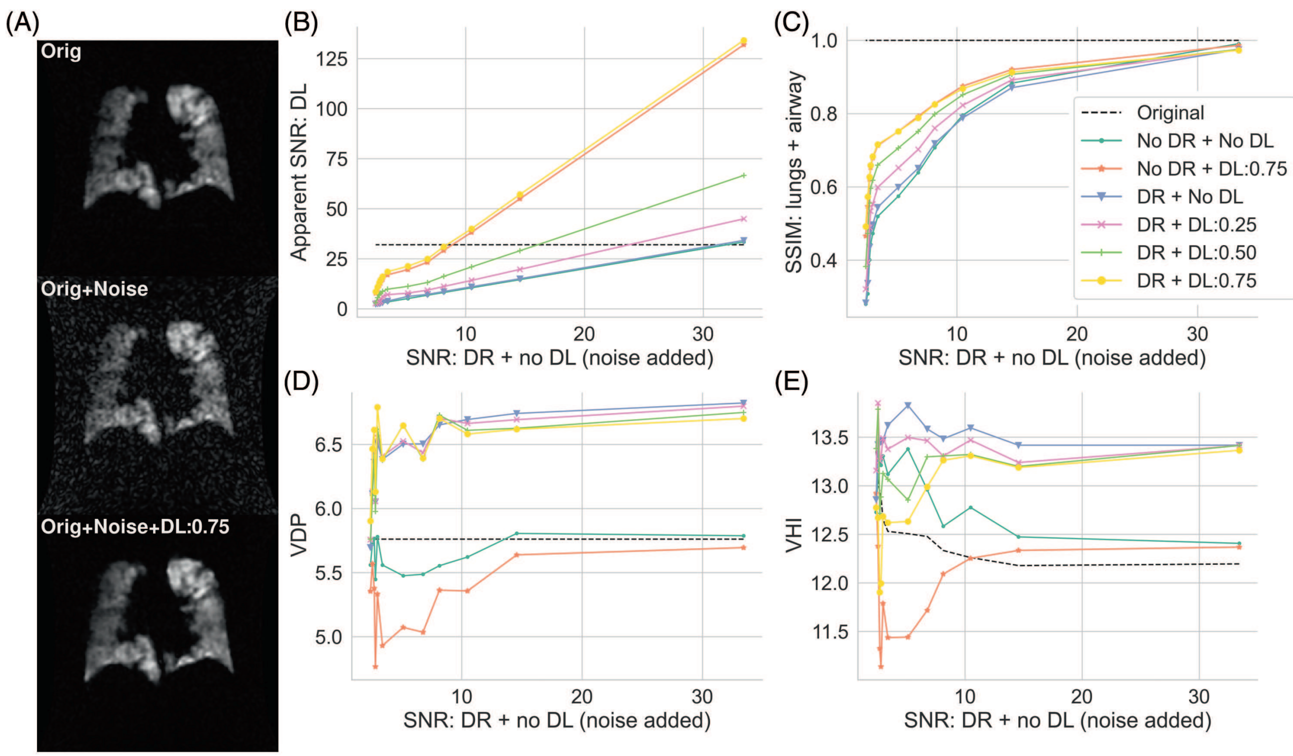

FIGURE 5 (Retrospective study) Deep learning (DL) reconstruction ventilation analysis with Gaussian noise added retrospectively to the raw k-space data of a high baseline SNR dataset acquired from a patient with moderate asthma (different to the data in Figures 1C, 2 and 5). (A) Initial image (SNR = 32), image with noise added (SNR = 10.5), and image with noise added and subsequent DL:0.75 reconstruction (SNR 40). (B–E) Apparent SNR, SSIM, ventilation defect percentage (VDP) and ventilation heterogeneity index (VHI) as a function of baseline SNR after adding different levels of Gaussian noise, and reconstructing at different denoising levels with de-ringing, and for no de-ringing at the 0.75 denoising level.

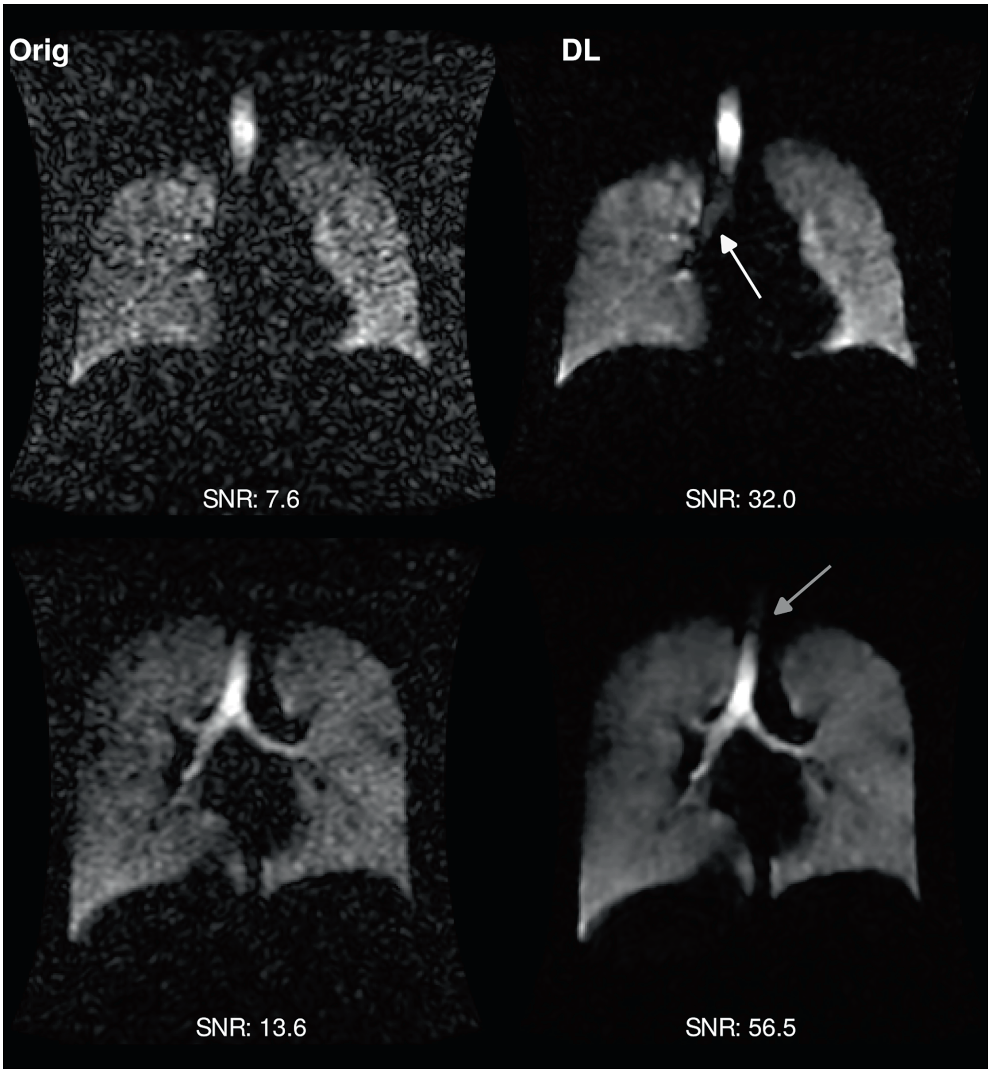

FIGURE 6 (Prospective study) Original reconstruction and deep learning (DL) reconstruction (denoising level:0.75 + de-ringing) of data acquired in two healthy subjects after inhalation of a dose of natural-abundance xenon. Quoted SNR is the mean across all slices. In the top row, the right bronchus becomes more visible after the DL reconstruction (white arrow), whereas in the bottom row, an upper region of the bronchus becomes apparent on DL reconstruction (gray arrow).

with increased spatial resolution (half the slice thickness) are shown for one subject in Figure 7. In all cases, the SNR of the DL reconstructed images was increased greater than fourfold compared with the original reconstructions. The difference maps indicated a negative difference at the borders of the lungs, particularly basally, indicating increased sharpness, with relatively little signal distribution change in most of the lung region.

4 | DISCUSSION

VDP is the principal metric used in ^{129}Xe ventilation image interpretation, and the significant increase in VDP in the DL-based reconstructed images compared with the original images should not be overlooked. However, this increase is explainable in part because of the sharpness of the lung edges that is introduced by the DL reconstruction process, particularly the de-ringing algorithm. This increased sharpness is highlighted by the high relative signal differences at the borders of the lungs in Figure 7, and furthermore in Figure 8, the effects of the denoising and de-ringing layers are separated visually for an example case, illustrating that the de-ringing has the most significant impact on the edge definition.

As the same lung cavity mask (generated from the original images) was used to calculate VDP for all sets of images, there are pixels at the borders of the lungs that were classified as signal in the original images, but have been set to noise in the DL images; these will, therefore, contribute to increased VDP. To investigate this further, we re-generated the lung cavity masks using the DL (0.75) images and found that the mean bias in VDP compared with the original images decreased to 0.88% (compared with 1.31%) (see Figure S3). This is still a non-negligible bias and is likely because of the increased sharpness at the border of low and high signal regions inside the lung mask (i.e., the borders of defects internally rather than peripherally). Again, most of this sharpness can be attributed to

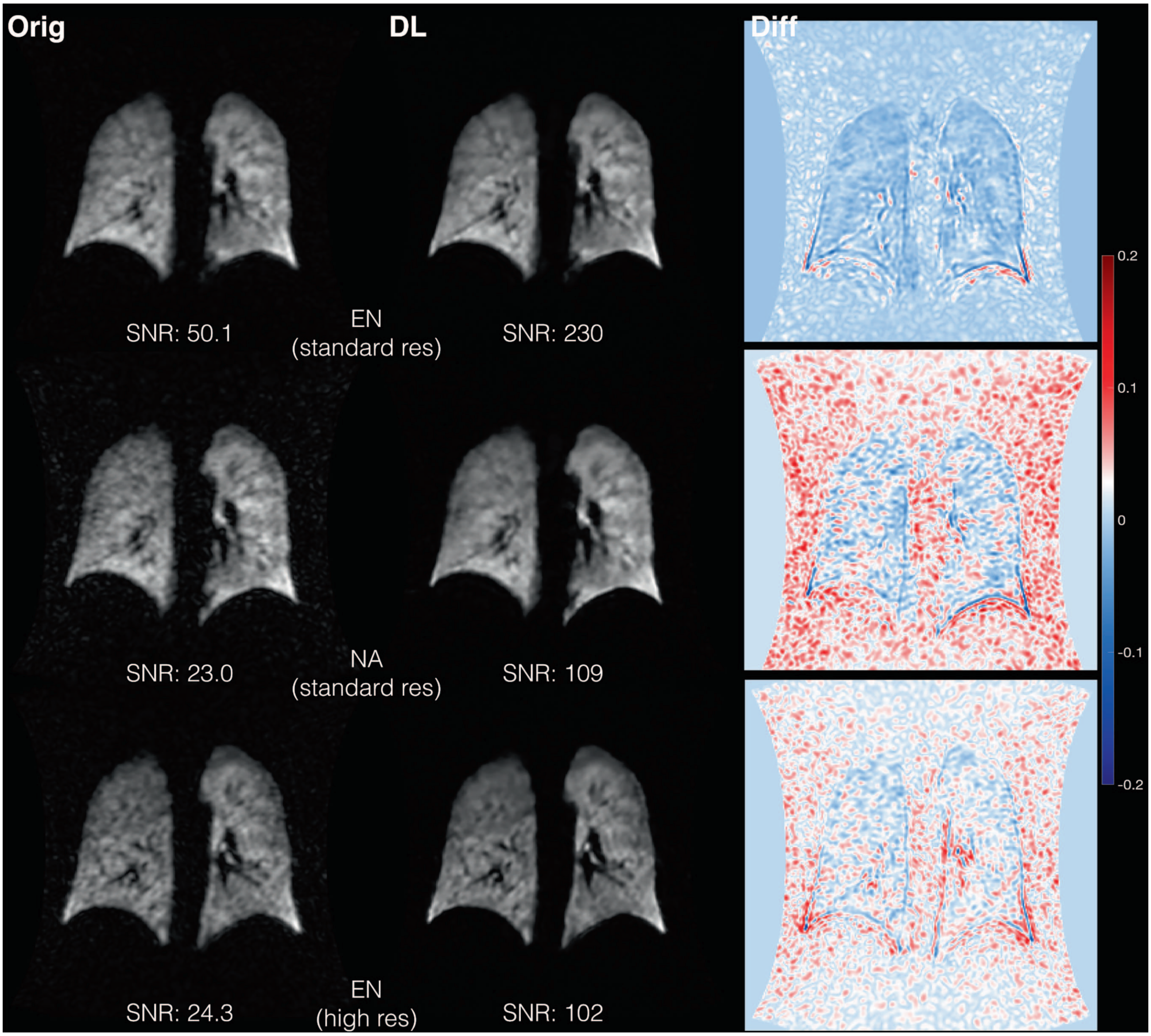

FIGURE 7 (Prospective study) From top to bottom: conventional 129-enriched xenon acquisition with 10 mm slice thickness; natural-abundance xenon acquisition at the same resolution; 129-enriched xenon acquisition with 5 mm slice thickness, all acquired in the same healthy volunteer. Conventionally reconstructed (Orig) images are shown on the left, deep learning (DL)-reconstructed images (denoising level:0.75 + de-ringing) in the middle, and "difference image" (subtraction after normalization) on the right. The color bar has been chosen to accentuate the signal/noise differences. SNR is noted for each image (mean across all slices).

the de-ringing process, as disabling de-ringing reduced the bias of VDP from 1.31% to 0.18% (using the original masks) (see Figure S2).

In DL denoised images, the separation between signal and noise is greater than in the original lower SNR images. In very poor SNR images, the ventilation binning algorithm may erroneously classify noise within ventilation defects as signal, artifactually reducing VDP. Therefore, although VDP is not the same when calculated from the original and DL images, the VDP calculated from the DL images is less affected by noise and, therefore, we believe it to be a more robust and accurate value; therefore, in future studies we recommend that DL recon is used prospectively and that the lung cavity masks are generated from the DL images. It is worth noting that the difference in VDP between original and DL images correlates significantly with the mean absolute VDP value (Spearman's $r = 0.810$, $p < 0.001$), as seen in Figure 3. Interestingly, the difference in VDP did not correlate with SNR for any of the DL levels when de-ringing was present ($p > 0.05$), however, difference in VDP weakly correlated with SNR for the DL (0.75) images when de-ringing was disabled (Pearson's $r = -0.377$, $p = 0.028$); perhaps explained by the reasons noted above. None of the other de-noising

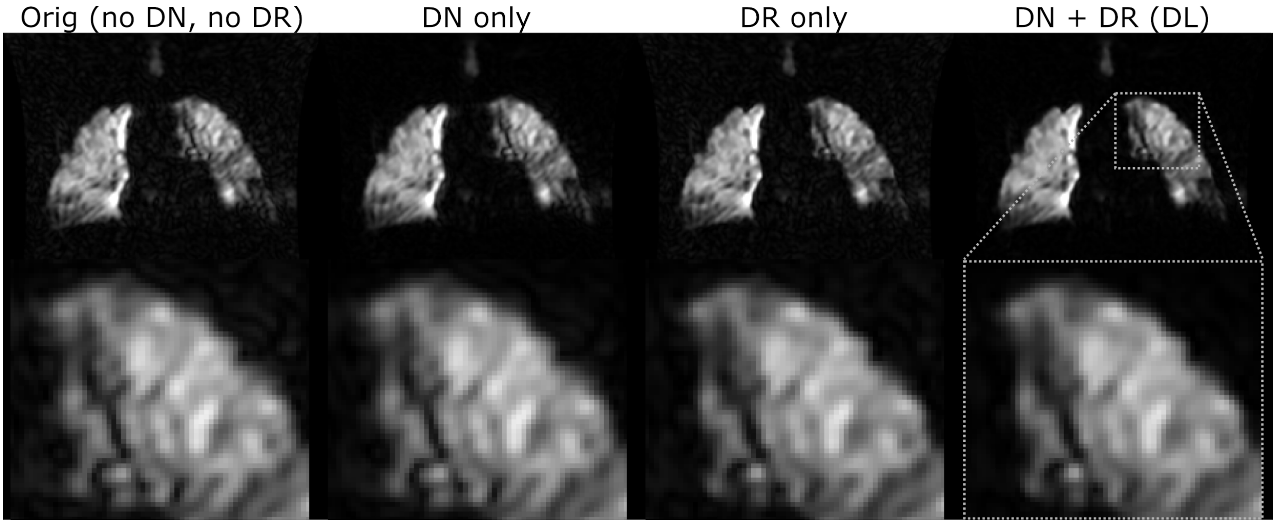

FIGURE 8 Visualization of the effects of de-ringing and de-noising on the sharpness of the image and consequently the delineation of the borders of the lungs. From left to right: images reconstructed from the same data—acquired from a patient with mild chronic obstructive pulmonary disease—using the original (conventional) reconstruction pipeline, the deep learning (DL) denoising (DN) pipeline only, the DL de-ringing (DR) pipeline only, and the DL denoising + de-ringing pipelines. A zoomed in view of the top left lung is provided in the bottom row.

methods exhibited a correlation between VDP difference and SNR ($p > 0.05$).

The increase in ventilation heterogeneity index may be similarly explained by the increase in image sharpness because of the DL recon pipeline, which would cause more low valued pixels to be included in the mask and, therefore, increase the local CV at the borders of defects and the lung periphery, and concordantly increase the interquartile range of the CV (i.e., VH_I). Disabling de-ringing significantly reduced the bias in VH_I, again highlighting the significant influence of this processing step. The median CV was found to have a mean positive bias of 0.17 toward DL (0.75) images; this bias is less significant than that of the VH_I and implies that the local distribution of ventilation is similar between the original images and DL reconstructed images across most of the lung regions. Moreover, a reduction in heterogeneity of ventilation—as was observed for the alternative denoising techniques—could be interpreted as a smoothing out of physiological ventilation heterogeneity, and therefore, a loss of functional information. However, it is worth noting that the nearest-neighbor CV calculation is dependent on the image SNR. In low SNR images, signal variations may arise from noise-related granularity as well as local physiology; smoothing out of the former would lead to a difference in the local CV as compared with the original images, although not associated with a loss in functional information.

Importantly, the similarity between DL reconstructed and original images (as assessed by the SSIM) was close to 1 (worst case SSIM of 0.954) and better preserved in DL images relative to the other denoising methods, indicating that the main shapes/structural features of the image are relatively unchanged by the DL reconstruction. Disabling the de-ringing pipeline led to a minor, but significant improvement in similarity, particularly in the lung and airway region. As with the comparison of quantitative ventilation metrics, it is worth reiterating that the SSIM calculation uses the original images as the reference/ground-truth, but given the improved SNR and sharpness of the DL images, a decrease in SSIM should not necessarily be viewed negatively. Indeed, as the de-ringing improves image sharpness and, therefore, aids VDP analysis, we recommend using DL denoising in combination with de-ringing in future studies.

There are several potential uses for this reconstruction model in ^{129}Xe imaging of lung ventilation in a clinical setting. For example, reducing the need to repeat scans when the SNR is insufficient, whether caused by incomplete inhalation of the gas dose, or suboptimal performance of the ^{129}Xe polarizer, or otherwise. As shown in Figure 5B, DL recon with a denoising level of 0.75 is able to recover a baseline SNR of ~10 to apparent SNR >30, and as shown in Figure 5C, a high similarity (SSIM >0.8) is preserved for all DL denoising levels for a baseline SNR of 10. Based on the data in Figure 5, we propose that a minimum image SNR of ~8 can be recovered using DL reconstruction (i.e., the image with SNR of 7.6 in Figure 6 is a good example of this lower limit). In addition, DL reconstruction should facilitate routine use of natural-abundance Xe in place of 129-enriched Xe for ^{129}Xe ventilation imaging, which would reduce the cost per scan approximately fivefold based on current gas mixture prices, therefore, increasing the economic viability of the technique. However, the small number of healthy volunteer datasets acquired with natural abundance Xe (and one dataset with 129-enriched Xe at higher spatial resolution) is a proof of concept only, and further studies in patients with ventilation heterogeneity are still needed to demonstrate clinical utility. In this work we have confined our investigations to 3D Cartesian data acquired using a balanced SSFP (bSSFP) sequence at 1.5 T. The reasons for this are twofold; (1) 3D bSSFP is used routinely for ventilation imaging at our center, therefore we have the largest database available for testing data acquired with this sequence; and (2) the DL model used here is specific to 3D acquisitions and a different model is needed for denoising 2D gradient echo data. To demonstrate the applicability of this 3D model beyond bSSFP, we have confirmed that it can effectively denoise 3D SPGR data using six datasets acquired at 3 T. Figure S4 shows the denoising performance in representative example datasets acquired from one healthy volunteer and one patient with sarcoidosis. Similar to in 3D bSSFP datasets, DL reconstruction at the 0.75 denoising level led to a three- to four-fold apparent SNR increase. It is important to note that the DL model used here cannot be directly applied for denoising k-space data acquired with a non-Cartesian acquisition trajectory, for example, the 3D radial trajectory used for dissolved-phase ^{129}Xe imaging.[37,38] Training and application of a non-Cartesian model will be the subject of future work. We also note that the current model holds substantial potential for application to other non-proton nuclei (e.g., ^{13}C, ^{23}Na, and ^{31}P) where the images have no background signal and SNR is typically a limiting factor. However, we acknowledge that a major limitation of the DL method presented here is that it is only available on GE Healthcare scanners. Alternative models are likely under-development, and ultimately, cross-vendor comparison and validation of the robustness of any proposed DL method will be needed to ensure maximum benefit to the ^{129}Xe MRI community.

5 | CONCLUSION

DL-based image reconstruction of ^{129}Xe images was found to be feasible using a commercially available reconstruction pipeline and was found to enhance ^{129}Xe ventilation image sharpness and greatly suppress image noise, while incurring a minor, explainable bias in key physiological metrics. Further application of this tool on images acquired from patients with a range of lung pathologies is required to fully evaluate the physiological interpretation of the resulting images and determine the optimal parameters for future studies. This approach holds potential for routine low-cost ^{129}Xe ventilation imaging using natural-abundance xenon, and/or improved spatial resolution imaging with 129-enriched xenon.

ACKNOWLEDGMENTS

This is independent research funded by Engineering and Physical Sciences Research Council and Medical Research Council Prosperity Partnership grant (EP/X025187/1) and a United Kingdom Research and Innovation Future Leaders Fellowship (MR/W008556/1) awarded to N.J.S., and was carried out at the National Institute for Health and Care Research (NIHR) Sheffield Biomedical Research Centre (BRC).

CONFLICT OF INTEREST STATEMENT

The views expressed are those of the author(s) and not necessarily those of the EPSRC, MRC, UKRI, the NIHR or the Department of Health and Social Care. J.d.A., A.C.S.B., and R.M.L. are employees of GE HealthCare.

ORCID

Neil J. Stewart https://orcid.org/0000-0001-8358-394X
Guilhem J. Collier https://orcid.org/0000-0002-1874-4775
Graham Norquay https://orcid.org/0000-0002-4108-9035

REFERENCES

1. He M, Zha W, Tan F, Rankine L, Fain S, Driehuys B. A comparison of two hyperpolarized 129Xe MRI ventilation quantification pipelines: the effect of signal to noise ratio. *Acad Radiol*. 2019;26:949-959. doi:10.1016/j.acra.2018.08.015
2. Stewart NJ, Norquay G, Griffiths PD, Wild JM. Feasibility of human lung ventilation imaging using highly polarized naturally abundant xenon and optimized three-dimensional steady-state free precession. *Magn Reson Med*. 2015;74:346-352. doi:10.1002/mrm.25732
3. He M, Robertson SH, Kaushik SS, et al. Dose and pulse sequence considerations for hyperpolarized 129Xe ventilation MRI. *Magn Reson Imaging*. 2015;33:877-885. doi:10.1016/J.MRI.2015.04.005
4. Food and Drug Administration – Dose Prescription Information. XENOVIEW (xenon Xe 129 hyperpolarized), for oral inhalation. 2022; Reference ID: 5100335 https://www.accessdata.fda.gov/drugsatfda_docs/label/2022/214375s000lbl.pdf
5. Norquay G, Collier GJ, Rao M, Stewart NJ, Wild JM. Xe 129 -Rb spin-exchange optical pumping with high photon efficiency. *Phys Rev Lett*. 2018;121:153201. doi:10.1103/PhysRevLett.121.153201
6. Nikolaou P, Coffey AM, Walkup LL, et al. XeNA: an automated 'open-source' 129Xe hyperpolarizer for clinical use. *Magn Reson Imaging*. 2014;32:541-550. doi:10.1016/J.MRI.2014.02.002
7. Hersman FW, Ruset IC, Ketel S, et al. Large production system for hyperpolarized 129Xe for human lung imaging studies. *Acad Radiol*. 2008;15:683-692. doi:10.1016/J.ACRA.2007.09.020
8. Dregely I, Ruset IC, Wiggins G, et al. 32-channel phased-array receive with asymmetric birdcage transmit coil for

hyperpolarized xenon-129 lung imaging. *Magn Reson Med*. 2013;70:576-583. doi:10.1002/mrm.24482
9. Zanette B, Munidasa S, Friedlander Y, Ratjen F, Santyr G. A 3D stack-of-spirals approach for rapid hyperpolarized 129Xe ventilation mapping in pediatric cystic fibrosis lung disease. *Magn Reson Med*. 2023;89:1083-1091. doi:10.1002/mrm.29505
10. Willmering MM, Niedbalski PJ, Wang H, et al. Improved pulmonary 129Xe ventilation imaging via 3D-spiral UTE MRI. *Magn Reson Med*. 2020;84:312-320. doi:10.1002/mrm.28114
11. Collier GJ, Hughes PJC, Horn FC, et al. Single breath-held acquisition of coregistered 3D 129Xe lung ventilation and anatomical proton images of the human lung with compressed sensing. *Magn Reson Med*. 2019;82:342-347. doi:10.1002/mrm.27713
12. Chung SH, Huynh KM, Goralski JL, et al. Feasibility of free-breathing 19F MRI image acquisition to characterize ventilation defects in CF and healthy volunteers at wash-in. *Magn Reson Med*. 2023;90:79-89. doi:10.1002/mrm.29630
13. Tustison NJ, Avants BB, Lin Z, et al. Convolutional neural networks with template-based data augmentation for functional lung image quantification. *Acad Radiol*. 2019;26:412-423. doi:10.1016/j.acra.2018.08.003
14. Astley JR, Biancardi AM, Hughes PJC, et al. Large-scale investigation of deep learning approaches for ventilated lung segmentation using multi-nuclear hyperpolarized gas MRI. *Sci Rep*. 2022;12:10566. doi:10.1038/s41598-022-14672-2
15. Astley JR, Wild JM, Tahir BA. Deep learning in structural and functional lung image analysis. *Br J Radiol*. 2022;95:20201107. doi:10.1259/bjr.20201107
16. Duan C, Deng H, Xiao S, et al. Fast and accurate reconstruction of human lung gas MRI with deep learning. *Magn Reson Med*. 2019;82:2273-2285. doi:10.1002/mrm.27889
17. Lebel RM. Performance characterization of a novel deep learning-based MR image reconstruction pipeline. *arXiv*. 2020;arXiv:2008.06559. doi:10.48550/arXiv.2008.06559
18. Park JC, Park KJ, Park MY, Kim MH, Kim JK. Fast T_2-weighted imaging with deep learning-based reconstruction: evaluation of image quality and diagnostic performance in patients undergoing radical prostatectomy. *J Magn Reson Imaging*. 2022;55:1735-1744. doi:10.1002/jmri.27992
19. Wang X, Ma J, Bhosale P, et al. Novel deep learning-based noise reduction technique for prostate magnetic resonance imaging. *Abdom Radiol (NY)*. 2021;46:3378-3386. doi:10.1007/s00261-021-02964-6
20. Zerunian M, Pucciarelli F, Caruso D, et al. Artificial intelligence based image quality enhancement in liver MRI: a quantitative and qualitative evaluation. *Radiol Med*. 2022;127:1098-1105. doi:10.1007/s11547-022-01539-9
21. Kim M, Kim HS, Kim HJ, et al. Thin-slice pituitary MRI with deep learning–based reconstruction: diagnostic performance in a postoperative setting. *Radiology*. 2021;298:114-122. doi:10.1148/radiol.2020200723
22. Sun S, Tan ET, Mintz DN, et al. Evaluation of deep learning reconstructed high-resolution 3D lumbar spine MRI. *Eur Radiol*. 2022;32:6167-6177. doi:10.1007/s00330-022-08708-4
23. Chan HF, Smith LJ, Biancardi AM, et al. Image phenotyping of preterm-born children using hyperpolarised 129Xe lung MRI and multiple-breath washout. *Am J Respir Crit Care Med*. 2023;207:89-100. doi:10.1164/rccm.202203-0606OC
24. Goldstein T, Studer C, Baraniuk R. A field guide to forward-backward splitting with a FASTA implementation. *arXiv preprint arXiv:1411.3406*. 2016:arXiv:1411.3406. doi:10.48550/arXiv.1411.3406
25. Zhang X, Xu Z, Jia N, et al. Denoising of 3D magnetic resonance images by using higher-order singular value decomposition. *Med Image Anal*. 2015;19:75-86. doi:10.1016/j.media.2014.08.004
26. Maggioni M, Katkovnik V, Egiazarian K, Foi A. Nonlocal transform-domain filter for volumetric data Denoising and reconstruction. *IEEE Trans Image Process*. 2013;22:119-133. doi:10.1109/TIP.2012.2210725
27. Eddy RL, Svenningsen S, McCormack DG, Parraga G. What is the minimal clinically important difference for helium-3 magnetic resonance imaging ventilation defects? *Eur Respir J*. 2018;51:1800324. doi:10.1183/13993003.00324-2018
28. Tzeng YS, Lutchen K, Albert M. The difference in ventilation heterogeneity between asthmatic and healthy subjects quantified using hyperpolarized 3He MRI. *J Appl Physiol*. 2009;106:813-822. doi:10.1152/japplphysiol.01133.2007
29. Smith LJ, Collier GJ, Marshall H, et al. Patterns of regional lung physiology in cystic fibrosis using ventilation magnetic resonance imaging and multiple-breath washout. *Eur Respir J*. 2018;52:1800821. doi:10.1183/13993003.00821-2018
30. Bertin FJ, Collier GJ, Hughes PJ, et al. *Simultaneous Segmentation of Airways and Ventilated Lung in Hyperpolarised-Gas MR Images by Deep Learning*. International Society for Magnetic Resonance in Medicine; 2021:3226.
31. He M, Driehuys B, Que LG, Huang YCT. Using hyperpolarized 129Xe MRI to quantify the pulmonary ventilation distribution. *Acad Radiol*. 2016;23:1521-1531. doi:10.1016/j.acra.2016.07.014
32. Collier GJ, Acunzo L, Smith LJ, et al. Linear binning maps for image analysis of pulmonary ventilation with hyperpolarized gas MRI: transferability and clinical applications. *Annual Meeting*; International Society for Magnetic Resonance in Medicine; 2018:4482.
33. Hughes PJC, Smith L, Chan HF, et al. Assessment of the influence of lung inflation state on the quantitative parameters derived from hyperpolarized gas lung ventilation MRI in healthy volunteers. *J Appl Physiol*. 2019;126:183-192. doi:10.1152/japplphysiol.00464.2018
34. Ridler TW, Calvard S. Picture Thresholding using an iterative selection method. *IEEE Trans Syst Man Cybern*. 1978;8:630-632.
35. Gudbjartsson H, Patz S. The rician distribution of noisy mri data. *Magn Reson Med*. 1995;34:910-914. doi:10.1002/mrm.1910340618
36. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600-612. doi:10.1109/TIP.2003.819861
37. Kaushik SS, Robertson SH, Freeman MS, et al. Single-breath clinical imaging of hyperpolarized (129)Xe in the airspaces, barrier, and red blood cells using an interleaved 3D radial 1-point Dixon acquisition. *Magn Reson Med*. 2016;75:1434-1443. doi:10.1002/mrm.25675
38. Collier GJ, Eaden JA, Hughes PJC, et al. Dissolved 129Xe lung MRI with four-echo 3D radial spectroscopic imaging: quantification of regional gas transfer in idiopathic pulmonary fibrosis. *Magn Reson Med*. 2021;85:2622-2633. doi:10.1002/mrm.28609

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Figure S1. Simplified diagram of the cohorts included in the retrospective and prospective parts of this study.

Figure S2. Bland–Altman plots of the difference between VDP and VH_I as a function of their mean values, calculated from conventionally reconstructed images vs. DL-reconstructed images *without de-ringing* (denoising level:0.75) for $N = 34$ patients with asthma and/or COPD. These plots are analogous to those in Figure 3 of the main manuscript, with the de-ringing pipeline disabled.

Figure S3. Bland–Altman plots of the difference between VDP and VH_I as a function of their mean values, calculated from conventionally reconstructed images vs. DL-reconstructed (denoising level:0.75 + de-ringing) images for $N = 34$ patients with asthma and/or COPD. These plots are analogous to those in Figure 3 of the main manuscript, however, whilst the DL-reconstructed data in Figure 3 were processed using a lung cavity mask derived from the original images, here the data were processed using a lung cavity mask derived from the DL-reconstructed images.

Figure S4. DL reconstruction denoising performance when applied to 3D spoiled gradient echo (SPGR) data. (A) Example original images and DL:0.75 reconstructed images for a patient with sarcoidosis (top row) and a healthy volunteer (bottom row). (B) Apparent SNR in six datasets for different denoising levels.

How to cite this article: Stewart NJ, de Arcos J, Biancardi AM, et al. Improving Xenon-129 lung ventilation image SNR with deep-learning based image reconstruction. *Magn Reson Med*. 2024;1-14. doi: 10.1002/mrm.30250