



BrainLossNet: a fast, accurate and robust method to estimate brain volume loss from longitudinal MRI

Roland Opfer¹ · Julia Krüger¹ · Thomas Buddenkotte² · Lothar Spies¹ · Finn Behrendt³ · Sven Schippling^{4,5} · Ralph Buchert²

Received: 20 December 2023 / Accepted: 27 May 2024 / Published online: 16 June 2024
© The Author(s) 2024

Abstract

Purpose MRI-derived brain volume loss (BVL) is widely used as neurodegeneration marker. SIENA is state-of-the-art for BVL measurement, but limited by long computation time. Here we propose “BrainLossNet”, a convolutional neural network (CNN)-based method for BVL-estimation.

Methods BrainLossNet uses CNN-based non-linear registration of baseline(BL)/follow-up(FU) 3D-T1w-MRI pairs. BVL is computed by non-linear registration of brain parenchyma masks segmented in the BL/FU scans. The BVL estimate is corrected for image distortions using the apparent volume change of the total intracranial volume. BrainLossNet was trained on 1525 BL/FU pairs from 83 scanners. Agreement between BrainLossNet and SIENA was assessed in 225 BL/FU pairs from 94 MS patients acquired with a single scanner and 268 BL/FU pairs from 52 scanners acquired for various indications. Robustness to short-term variability of 3D-T1w-MRI was compared in 354 BL/FU pairs from a single healthy men acquired in the same session without repositioning with 116 scanners (Frequently-Traveling-Human-Phantom dataset, FTHP).

Results Processing time of BrainLossNet was 2–3 min. The median [interquartile range] of the SIENA-BrainLossNet BVL difference was 0.10% [– 0.18%, 0.35%] in the MS dataset, 0.08% [– 0.14%, 0.28%] in the various indications dataset. The distribution of apparent BVL in the FTHP dataset was narrower with BrainLossNet ($p = 0.036$; 95th percentile: 0.20% vs 0.32%).

Conclusion BrainLossNet on average provides the same BVL estimates as SIENA, but it is significantly more robust, probably due to its built-in distortion correction. Processing time of 2–3 min makes BrainLossNet suitable for clinical routine. This can pave the way for widespread clinical use of BVL estimation from intra-scanner BL/FU pairs.

Keywords Magnetic resonance imaging · Brain volume loss · Convolutional neural network · Multiple sclerosis

Introduction

Whole brain volume loss (BVL) estimated from two MRI scans of the same individual, a baseline (BL) and a follow-up (FU) scan, is an useful neurodegeneration marker in chronic neurological and psychiatric conditions [1, 2]. In multiple sclerosis (MS), MRI-based BVL is associated with disability and the risk of conversion from clinically isolated syndrome to definite MS [3]. It has been proposed as a complementary marker to assess treatment response in MS [4].

SIENA (Structural Image Evaluation using Normalization of Atrophy) [5] is the state-of-the-art tool for MRI-based BVL estimation [3, 6]. It has been very well validated, including detailed characterization regarding measurement error and physiological short-term test–retest variability [7, 8]. Age-dependent thresholds on SIENA BVL estimates have

✉ Ralph Buchert
r.buchert@uke.de

¹ Jung Diagnostics GmbH, Hamburg, Germany
² Department of Diagnostic and Interventional Radiology and Nuclear Medicine, University Medical Center Hamburg-Eppendorf, Martinistr. 52, 20246 Hamburg, Germany
³ Institute of Medical Technology and Intelligent Systems, Hamburg University of Technology, Hamburg, Germany
⁴ Multimodal Imaging in Neuroimmunological Diseases (MINDS), University of Zurich, Zurich, Switzerland
⁵ Neuroscience and Rare Diseases (NRD), Roche Pharma Research and Early Development (pRED), Basel, Switzerland

been established to discriminate disease-related BVL from healthy aging [9, 10]. The in-depth validation of SIENA is the basis for its use to measure BVL as primary or secondary endpoint in phase II/III MS drug trials [11, 12].

SIENA primarily relies on brain surface changes, which limits its ability to measure regional BVL. Jacobian determinant integration leverage (JI) methods overcome this limitation [13, 14]. They are based on high-dimensional non-linear image registration of BL/FU pairs. Regional volume changes are obtained by integrating the Jacobian determinant of the deformation field across the region-of-interest. JI gained widespread acceptance for measuring regional BVL [3, 14].

SIENA and JI are computationally demanding: typical computation time per BL/FU pair is about 1 h on a standard workstation. While this is not a limitation in research, it restricts their use in clinical routine. For seamless integration into the radiologist's workflow, the volumetric analyzes must not delay report-writing [15].

SIENA and JI are built on conventional image analysis techniques. Thus, each case has to go through the entire processing pipeline. In contrast, deep learning-based approaches can be very expensive in terms of computation time for training and validation, but once this is completed, processing of new cases is fast. Thus, deep learning can drastically reduce computation time for complex image analyzes and, therefore, support their integration into busy clinical workflows. Furthermore, deep learning-based methods provide superior performance compared to conventional techniques in many image analysis tasks [16, 17]. For example, CNN-based non-linear image registration [18, 19] outperforms conventional methods including ANTS [20] and DARTEL [21].

Against this background, the current study designed, trained and tested a novel CNN-based method, "BrainLossNet", for estimating BVL from intra-scanner T1-weighted 3D-MRI BL/FU pairs with the same accuracy as SIENA, but improved robustness and ≤ 5 min processing time.

Materials and methods

Datasets

All datasets comprised 3D gradient echo brain T1w-MRI scans acquired with sequences recommended by the scanner manufacturer. FU scans were acquired with the same scanner and the same sequence as the corresponding BL scan. The datasets were not curated manually. A summary of the datasets is given in Table 1.

Development dataset

The development dataset was included retrospectively from a large dataset of MRI acquired in clinical routine for various indications. The only inclusion criterion was a minimum time interval of 6 months between BL and FU scan. This resulted in the inclusion of 1793 BL/FU pairs from 1793 different patients acquired with 83 different MRI scanners (32 different scanner models, Supplementary section "MR scanner models in the training dataset").

The data of the development dataset had been provided to jung-diagnostics GmbH under the terms of the European Data Protection Regulation for remote image analysis. Subsequently, the data had been anonymized. The need for written informed consent for the retrospective use of the anonymized data was waived by the ethics review board of the general medical council of the state of Hamburg, Germany (reference number 2021-300047-WF).

The data was randomly split into 1525 (85%) BL/FU pairs for training and 268 (15%) BL/FU pairs for testing.

Multiple sclerosis (MS) dataset

The MS dataset retrospectively included 94 patients with confirmed MS diagnosis from an observational study at the University Hospital of Zurich, Switzerland [22]. On average, 3.5 scans were available per patient. BL/FU pairs were built from consecutive scans (for example, 3 BL/FU pairs were built for a patient with 4 scans: first/second, second/third, third/fourth). This resulted in a total of 225 BL/FU pairs. Mean BL-to-FU time interval was 1.1 ± 0.5 y.

The retrospective use of the data was approved by the ethics committee of the Canton of Zurich, Switzerland (reference numbers 2013-0001, 2020-01187).

Frequently traveling human phantom (FTHP) dataset

The FTHP dataset was recently introduced by our group [23]. It comprises 557 3D-T1w-MRI scans of a single healthy male who completed 157 imaging sessions on 116 different MRI scanners within about 2.5y. The current study included all imaging sessions that comprised at least two back-to-back repeat scans that had been acquired consecutively (delay < 1 min) with the same sequence in the same session without repositioning (123 sessions on 116 different MRI scanners).

Most imaging sessions comprised three to five back-to-back scans. BL/FU pairs were built from consecutive scans within a session. This resulted in a total of 354 BL/FU pairs. BL/FU pairs combining scans from different imaging sessions were not considered. The rationale for this was that the FTHP dataset was used to assess the stability of BVL estimates with respect to short-term test-retest variability of 3D-T1w-MRI. For this purpose, the BL/FU pairs were

Table 1 Datasets

	Training	Testing		
	Development dataset	Development dataset	MS patients	FTHP dataset
No. BL/FU pairs	1525	268	225	354
No. different subjects	1525	268	94	1
Age at BL in years, mean [range]	46.9 [18.1, 88.8]	46.2 [19.4, 83.1]	34.2 [20.1, 54.7]	49.6 [48.6, 51.2]
Delay between BL/FU pairs in years, mean (std)	1.9 (1.2)	1.9 (1.4)	1.1 (0.5)	< 1 min
No. different scanners	83	52	1	116
No. different scanner models	32	26	1	30

intended to combine scans as similar as possible, that is, no volume change from BL to FU, neither true volume change associated with aging or physiological status nor apparent volume change due to scanner-/sequence-dependent variability.

The healthy individual had given written informed consent for the retrospective use of the dataset. This was approved by the ethics review board of the general medical council of the state of Hamburg, Germany (reference number PV5930). The FTHP dataset was made freely available for research purposes by our group at <https://www.kaggle.com/datasets/ukeependorf/frequently-traveling-human-phantom-fthp-dataset>.

SIENA

For benchmarking, BVL estimates were computed with SIENA (version 2.6) [5], which is part of the FMRIB Software Library (FSL; <http://www.fmrib.ox.ac.uk/fsl>). Parameter settings for SIENA were as described previously [7, 9]. SIENA BVL estimates were computed for all BL/FU pairs in all datasets.

BrainLossNet

The BrainLossNet pipeline was specifically designed to estimate BVL from BL/FU pairs acquired with the same scanner and the same acquisition sequence. A detailed description is given in the following subsections. A schematic presentation is given in Fig. 1.

Preprocessing

Preprocessing (Fig. 1A) included reslicing to 1mm cubic voxels and (if necessary) rearranging to transversal orientation. Next, a $176 \times 208 \times 208$ matrix covering the entire skull and brainstem down to the foramen magnum was cropped

from the image. These steps were performed independently for BL and FU. Finally, BL and FU crop were rigidly registered in halfway space using ANTS [20].

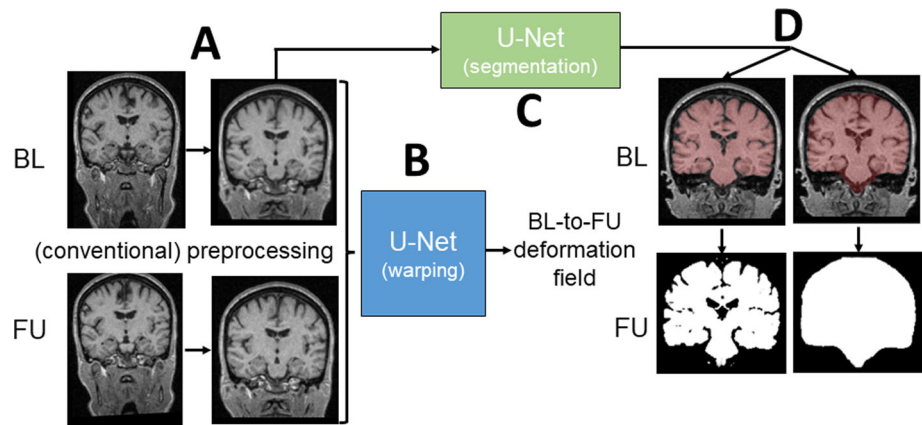
Non-linear registration

A convolutional U-Net [24] following the architecture described in [18] was trained specifically for non-linear registration of preprocessed BL/FU pairs (Fig. 1B). The training subset of the development dataset was used for this purpose.

The unsupervised learning approach proposed by the VoxelMorph learning framework for deformable medical image registration was used [18]. In brief, the two crops from the BL/FU pair were concatenated and served as input to the U-Net. The output of the U-Net was the 3D tensor field characterizing the local deformations required to warp the BL crop to the FU crop (matrix size of the deformation field = $176 \times 208 \times 208 \times 3$). The weighted sum of the cross-correlation of the warped BL crop with the FU crop plus the spatial gradient of the deformation field (for regularization) was used as loss function. Further details are provided in the Supplementary material (“3D-CNN for non-linear registration: architecture and training”, Supplementary Figs. 1, 2) and in [18].

Data augmentation was performed on the fly during the training including left–right flipping and rotation (same for both crops) and adding Gaussian noise (separately for both crops). The direction of the non-linear registration was randomly interchanged (e.g., BL crop warped to FU crop or reversed). The training was performed for 350 epochs with a batch size 2 using the Adam optimizer with learning rate 0.0001. The training took 5 days (python 3.10, pytorch 2.0, GPU NVIDIA RTX A5000 24 GB memory).

Fig. 1 BrainLossNet pipeline



Computation of brain volume loss

First, a custom 3D-CNN [23] was used to segment the brain parenchyma in the BL crop (Fig. 1C, supplementary section “3D-CNN for BPV and TIV segmentation: architecture and training”, Supplementary Fig. 3). This resulted in a binary BL parenchyma mask (1/0 inside/outside). The brain parenchyma volume (BPV) at BL (BPV_{BL}) was obtained by counting the voxels in the BL mask.

Then, the U-Net for non-linear registration was used to compute the BL-to-FU deformation field. This was used to warp the BL parenchyma mask from the BL crop to the FU crop (Fig. 1D). The BPV at FU ($BPV_{FU \leftarrow BL}$) was estimated from the warped mask by summing the voxel intensities of the warped mask across all voxels (including boundary voxels with intensity < 1 due to interpolation).

Then, the processing was reversed. More precisely, the FU-to-BL deformation field for the same BL/FU pair was estimated by the U-Net for non-linear registration and used to warp the parenchyma mask segmented in the FU crop to the BL crop. In this (“backward”) processing, the BPV at FU (BPV_{FU}) was computed from the original mask, the BPV at BL ($BPV_{BL \leftarrow FU}$) was estimated from the warped mask.

Then, the percentage brain volume change from BL to FU was estimated according to the following “forward” (fw) and “backward” (bw) formulas

$$\Delta BPV_{fw} = 100 * \frac{BPV_{FU \leftarrow BL} - BPV_{BL}}{BPV_{BL}} \quad (1)$$

$$\Delta BPV_{bw} = 100 * \frac{BPV_{FU} - BPV_{BL \leftarrow FU}}{BPV_{BL \leftarrow FU}} \quad (2)$$

In principle, both formulas should yield the same result. However, in order to increase robustness, particularly by reducing asymmetry-induced bias [25], the two estimates were averaged

$$\Delta BPV = 0.5 * (\Delta BPV_{fw} + \Delta BPV_{bw}) \quad (3)$$

Correction for distortion effects

Pilot experiments had shown an association between the apparent BL-to-FU change of the total intracranial volume (TIV) and the severity of the distortion of BL and FU image relative to each other. This was the rationale for using the apparent TIV difference between BL and FU to correct ΔBPV for distortion effects.

The TIV was segmented using the same custom 3D-CNN that was used for segmentation of the brain parenchyma [23], separately for BL and FU. The BL TIV mask was warped to the FU crop and the FU TIV mask was warped to the BL crop using the same BL-to-FU and FU-to-BL deformation fields as for the brain parenchyma. The percentage TIV change from BL to FU was computed analogous to Eqs. (1)–(3) as

$$\Delta TIV = 0.5 * (\Delta TIV_{fw} + \Delta TIV_{bw}) \quad (4)$$

Then, regression analysis was performed in the training dataset with ΔBPV as independent variable and ΔTIV as predictor:

$$\Delta BPV = m_{TIV} * \Delta TIV + c_{TIV} \quad (5)$$

This resulted in $m_{TIV}=0.95$ (95%-confidence interval [0.93, 0.96]), $c_{TIV}=-0.23$ [-0.24,-0.22] (Fig. 2). In order to eliminate distortion effects from ΔBPV , residuals were computed with respect to the regression line:

$$res \Delta BPV = \Delta BPV - m_{TIV} * \Delta TIV = \Delta BPV - 0.95 * \Delta TIV \quad (6)$$

The offset parameter c_{TIV} in the regression model (5) represents the mean BVL in the training dataset in the absence of distortion effects ($\Delta TIV = 0$) and, therefore, does not need to be subtracted from ΔBPV to correct for distortion effects (if c_{TIV} would be included in formula (6), it would be removed by rescaling described below).

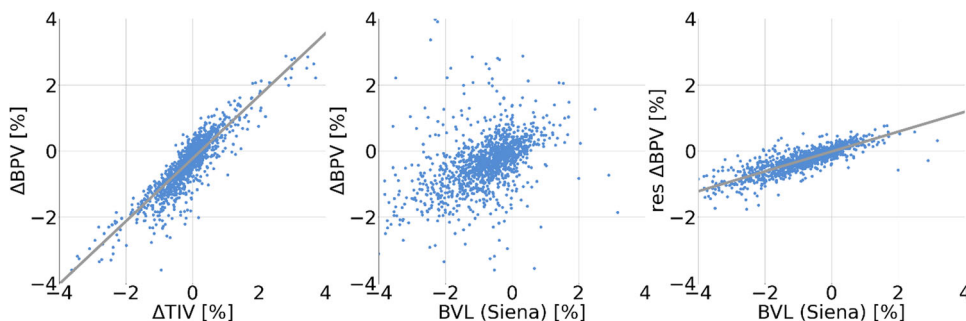


Fig. 2 Distortion correction and rescaling. Linear regression of $\Delta B P V$ versus $\Delta T I V$ (left) in the training dataset was used for distortion correction, linear regression of the distortion-corrected residual of $\Delta B P V$

versus SIENA BVL (right) in the training dataset was used for rescaling. The scatter plot in the middle shows SIENA BVL versus $\Delta B P V$ (without distortion correction)

$\Delta T I V$ is computed for each new BL/FU pair and then used for “individual” distortion correction of $\Delta B P V$ according to formula (6).

The residuals can be assumed to be proportional to BVL, but they do not represent BVL in %. For rescaling, regression analysis was performed in the training dataset with the residual as dependent variable and the SIENA BVL estimate as predictor:

$$\text{res}\Delta B P V = m_{\text{Siena}} * \text{BVL}_{\text{Siena}} + c_{\text{Siena}} \tag{7}$$

This revealed $m_{\text{Siena}} = 0.30$ (95% confidence interval [0.29, 0.31]), $c_{\text{Siena}} = -0.01$ [-0.02, -0.01] (Fig. 2). The slope was used to rescale the residuals for quantitative estimation of the percentage BVL as

$$\text{BVL} = \text{res}\Delta B P V / m_{\text{Siena}} = \text{res}\Delta B P V / 0.30 \tag{8}$$

The trained BrainLossNet is available from the authors upon request under a non-disclosure agreement for non-commercial use.

Testing of BrainLossNet

Internal consistency

BrainLossNet computes the BVL as the mean of a forward and a backward estimate. More precisely, from Eqs. (8), (6), (4) and (3) one easily derives.

$$\text{BVL} = 0.5 * (\text{BVL}_{\text{fw}} + \text{BVL}_{\text{bw}}) / 0.30 \tag{9}$$

with

$$\text{BVL}_{\text{fw/bw}} = (\Delta B P V_{\text{fw/bw}} - 0.95 * \Delta T I V_{\text{fw/bw}}) \tag{10}$$

Forward and backward estimates of BVL are independent in the sense that they are based on two different deformation

fields. The consistency of forward and backward estimates was tested in the test sample of the development dataset using Bland–Altman analysis.

Agreement between BrainLossNet and SIENA

BVL estimates from BrainLossNet were compared with BVL estimates from SIENA using Bland–Altman analysis in the test datasets. Median and interquartile range (IQR) of the differences were computed as well as the intra-class correlation coefficient (ICC) of the BVL estimates.

Robustness to short-term test–retest variability of 3D-T1w-MRI

For the BL/FU pairs of consecutive back-to-back scans from the FTHP dataset, the true BVL is zero. Thus, the distribution of BVL estimates for these BL/FU pairs was used to characterize the robustness with respect to short-term test–retest variability of 3D-T1w-MRI, separately for BrainLossNet and SIENA. The width of the distribution of the BVL estimates was compared between BrainLossNet and SIENA using the Pitman test for comparison of variances of paired samples [26]. The Pitman test consists of testing for correlation between the sum and the difference between paired observations, a significant correlation indicating a significant difference regarding the variances.

Results

The computation time for the complete BrainLossNet pipeline (including preprocessing) was approximately 150 s per BL/FU pair on a Linux workstation (CPU: Intel, Xeon, Silver, 4110 CPU 2.10 GHz 32 kernels, GPU Quadro P5000 16 GB). Computation time for SIENA depends on the resolution of the images. For an image pair with $260 \times 256 \times$

Fig. 3 Consistency of forward and backward estimates. Scatter plot (left) and Bland–Altman plot (right) of forward versus backward BrainLossNet BVL estimates in the test sample from the development dataset. (IQR: interquartile range)

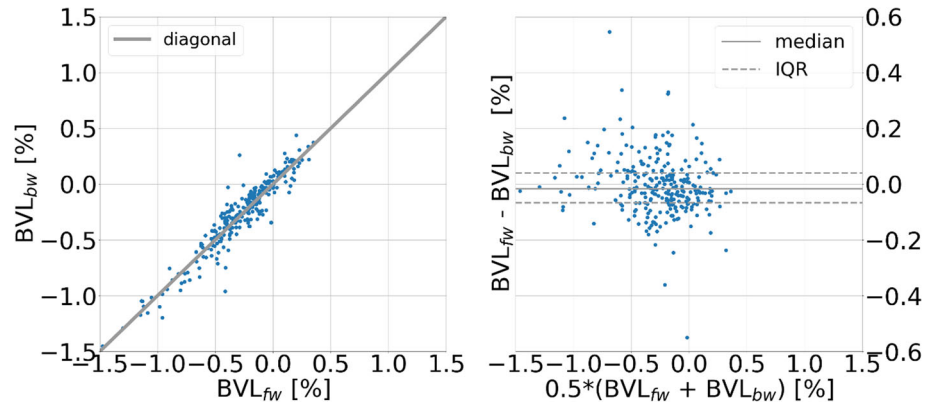
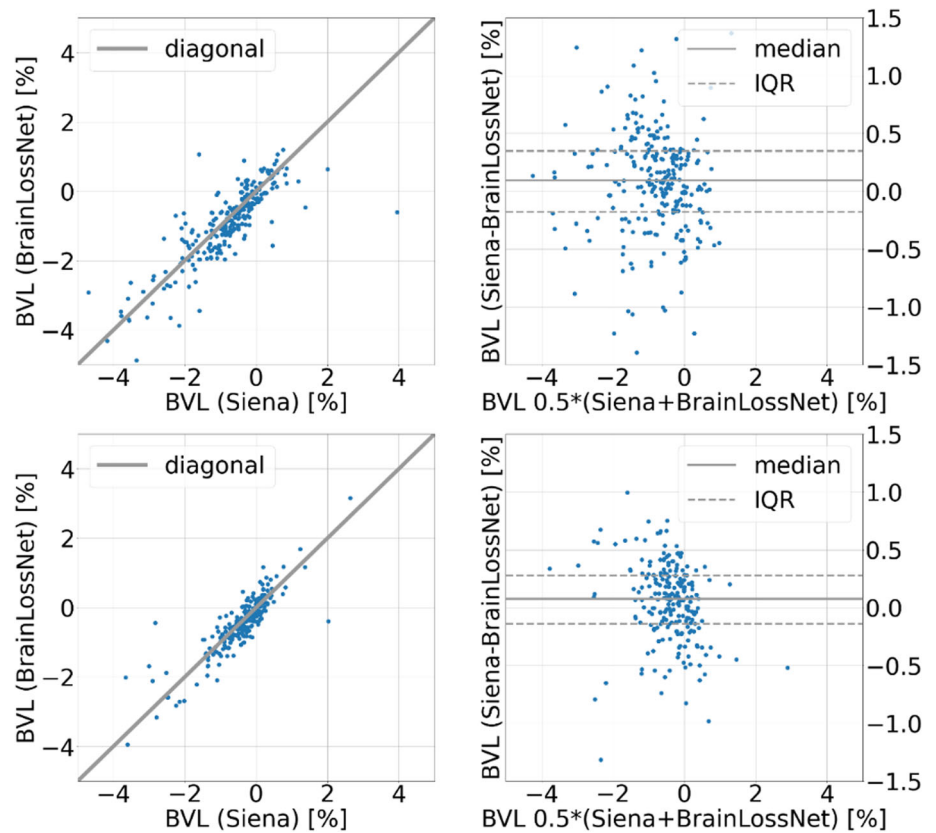


Fig. 4 BrainLossNet versus SIENA. Scatter plot (left) and Bland–Altman plot (right) of BVL estimates from BrainLossNet versus SIENA in the test sample from the development dataset (first row) and in the MS dataset (second row). (IQR: interquartile range)



256 voxels, the computation time was 63 min on the same workstation.

Concerning internal consistency of the BrainLossNet pipeline, Fig. 3 shows the scatter and the Bland–Altman plot of forward versus backward BVL estimates in the test sample from the development dataset. The median [IQR] forward–backward difference was -0.02% [-0.07% , 0.04%]. 95% of all forward–backward differences (absolute value) were smaller than 0.19%.

Concerning agreement between BrainLossNet and SIENA, Fig. 4 shows scatter and Bland–Altman plots of BrainLossNet versus SIENA BVL estimates. For the test sample from the development dataset, the median [IQR]

SIENA–BrainLossNet BVL difference was 0.10% [-0.18% , 0.35%]. The 95th percentile of the absolute differences was 1.23%, the ICC was 0.80. For the MS dataset, the corresponding values were 0.08% [-0.14% , 0.28%], 0.67% (95th percentile), and 0.85, respectively.

Concerning robustness with respect to short-term test–retest variability of 3D-T1w-MRI, Fig. 5 shows the histograms of the apparent BVL between consecutive back-to-back scans from the FTHP dataset. The mean apparent BVL was $0.02\% \pm 0.17\%$ for SIENA and $0.04\% \pm 0.11\%$ for BrainLossNet. The variance of the apparent BVLs was significantly larger for SIENA than for BrainLossNet (Pitman test: Spearman correlation coefficient 0.112, $p = 0.036$).

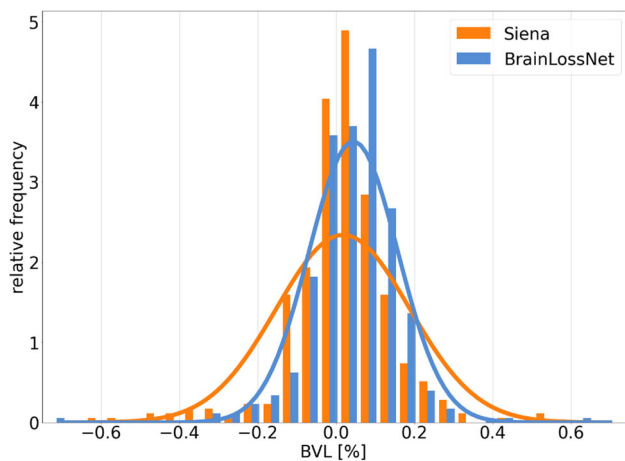


Fig. 5 Robustness with respect to short-term test–retest variability of 3D-T1w-MRI. Histograms of the apparent BVL between consecutive back-to-back scans from the FTHP dataset. The continuous lines represent Gaussian fits

The 95th percentile of the apparent BVL (absolute value) was 0.32% for SIENA and 0.20% for BrainLossNet.

Discussion

BrainLossNet BVL estimation from a BL/FU pair took 2–3 min including all preprocessing steps. Most time was spent for the rigid BL-FU registration with the conventional ANTS method during preprocessing. Thus, further acceleration might be achieved by employing an additional CNN for this step. However, 2–3 min computation time of the current implementation makes BrainLossNet suitable for integration into routine workflows in clinical practice also without further acceleration.

The difference of BVL estimates between BrainLossNet and SIENA was not larger than 1.23% in 95% of the cases in the test sample from the development dataset and not larger than 0.67% in 95% of the cases in the MS dataset. This is in the same range as the 95th percentile of the within-patient variability of BVL estimates from short-term repeated MRI measurements with SIENA: 0.43%, 0.63%, and 1.15% in three different datasets [7]. Thus, BrainLossNet provided essentially the same BVL estimates as SIENA in the vast majority of cases, indicating that BrainLossNet and SIENA BVL estimates are largely equivalent.

The somewhat larger difference between BrainLossNet and SIENA in about 5% of the cases might be explained by better robustness of BrainLossNet to some extent. While the median difference of the BVL estimates from two consecutive scans in the FTHP dataset was similar for BrainLossNet and SIENA, the distribution of the differences was somewhat wider for SIENA (Fig. 5): the 95th percentile of the differences was 0.32% for SIENA compared to 0.20% for

BrainLossNet. This suggests that among the about 5% cases with larger difference between BrainLossNet and SIENA, the BrainLossNet estimate might be more reliable.

In a pilot experiment, a strong association between ΔTIV and image distortions had been observed, in line with previous reports [15]. Based on the assumption that the strong correlation between ‘raw’ values of ΔBPV with ΔTIV observed here (Fig. 2) is mainly due to the fact that the impact of image distortion is similar for ΔBPV and ΔTIV , the correlation was used to correct ‘raw’ ΔBPV estimates for image distortions by computing residuals with respect to the regression line of ΔBPV versus ΔTIV in the training dataset. The proposed distortion-correction might have made BrainLossNet less sensitive to image distortions than SIENA. This might have contributed to its improved robustness.

For rescaling of the residuals to properly represent BVL in %, regression of the residuals versus SIENA BVL estimates was used. Importantly, the regression analyzes for distortion correction and rescaling were performed in the training dataset. The resulting formulas for distortion correction and rescaling were applied in all test datasets. No attempts were made to optimize distortion correction and/or rescaling for any of the test datasets. SIENA also involves a calibration step [5].

BrainLossNet relies on non-linear registration of BL/FU pairs, as do JI methods. However, in contrast to JI methods, which estimate brain volume change by integration of the Jacobian determinant of the deformation field, BrainLossNet applies the deformation field to binary brain masks obtained by delineation of the brain parenchyma with another CNN (Fig. 1). The rationale for this was that the computation of partial derivatives of the deformation field can be numerically challenging and, therefore, can result in noisy Jacobian determinant images (e.g., Fig. 7 in [22]). In pilot experiments comparing JI of the CNN-based deformation field with the warping of the binary parenchyma mask, the mask warping approach outperformed JI regarding agreement with SIENA (Supplementary Fig. 4). We hypothesize that this is explained by local ‘non-smoothness’ of the CNN-based deformation field. Conventional registration approaches usually ensure that the deformation field has certain smoothness properties, either by restricting the deformation field to a pre-defined space of functions (e.g., splines of a given order) or by the optimization process. As a consequence, the derivatives of the deformation field are well defined. In contrast, CNN-based non-linear registration does not employ strict smoothness constraints (beyond favoring smooth solutions by the loss function used for the CNN training).

Zhan and co-workers [27] recently proposed a different CNN-based method for BVL estimation. They used supervised training of an U-Net to reproduce SIENA change maps. Potential advantages of this approach compared to BrainLossNet include that it might be more robust when BL and FU

scans were acquired with different MR scanners or different acquisition sequences. Potential advantages of BrainLossNet compared to the approach by Zhan and co-workers include that it might be more robust with respect to image distortions. Given that Zhan and co-workers used SIENA change maps for the U-Net training, the sensitivity to image distortions might be similar for their U-Net as for SIENA. Furthermore, BrainLossNet is easily adapted to estimate regional volume loss, which might be less straightforward for the U-Net method. These hypotheses should be tested in a head-to-head comparison of both methods.

BrainLossNet was trained and tested specifically to estimate BVL from BL/FU pairs acquired with the same scanner and the same acquisition sequence. Thus, its use should be restricted to this scenario (which is the most frequent scenario at many sites). We intentionally did not train BrainLossNet to estimate BVL also in scenarios with BL and FU from different scanners, because we hypothesized that this would have compromised BrainLossNet's performance in BL/FU pairs acquired with the same MR scanner and the same acquisition sequence. There are no further restrictions on the use of BrainLossNet, in particular, there are no specific restrictions regarding the MR scanner or the 3D-T1w acquisition sequence. This is due to the fact that the development dataset used for training and testing of BrainLossNet was quite heterogeneous, essentially covering the whole range of image characteristics typically encountered in clinical practice.

In conclusion, BrainLossNet BVL estimates from BL/FU pairs acquired with the same MR scanner using the same acquisition sequence are essentially the same as those from SIENA in the vast majority of cases. In the remaining “outliers”, BrainLossNet estimates might be more reliable, given the better robustness of BrainLossNet. The BrainLossNet processing time of 2–3 min per case on a standard workstation makes it suitable for integration into routine workflows in clinical practice.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11548-024-03201-3>.

Author contributions Roland Opfer: conceptualization, methodology, software, writing-original draft, visualization, formal analysis. Julia Krüger: methodology, software, visualization, formal analysis. Thomas Buddenkotte: methodology, software, visualization, formal analysis. Lothar Spies: writing-reviewing and editing. Finn Behrend: software, methodology. Sven Schippling: writing-reviewing and editing, data curation. Ralph Buchert: conceptualization, methodology, writing-original draft, supervision.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest Roland Opfer, Julia Krüger and Lothar Spies are employees of jung diagnostics GmbH, Germany (www.jung-diagnostics.de). Sven Schippling is employee of Roche Pharma Research and Early Development. There is no actual or potential conflict of interest for the other authors. The nonemployee authors had control of the data and information that might present a conflict of interest for the employee authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- O'Brien JT, Firbank MJ, Ritchie K, Wells K, Williams GB, Ritchie CW, Su L (2020) Association between midlife dementia risk factors and longitudinal brain atrophy: the PREVENT-Dementia study. *J Neurol Neurosurg Psychiatry* 91:158–161. <https://doi.org/10.1136/jnnp-2019-321652>
- Zhang B, Lin L, Wu S (2021) A review of brain atrophy subtypes definition and analysis for alzheimer's disease heterogeneity studies. *J Alzheimers Dis* 80:1339–1352. <https://doi.org/10.3233/jad-201274>
- Matthews PM, Gupta D, Mittal D, Bai W, Scalfari A, Pollock KG, Sharma V, Hill N (2023) The association between brain volume loss and disability in multiple sclerosis: a systematic review. *Mult Scler Relat Disord* 74:104714. <https://doi.org/10.1016/j.msard.2023.104714>
- Giovannoni G, Turner B, Gnanapavan S, Offiah C, Schmierer K, Marta M (2015) Is it time to target no evident disease activity (NEDA) in multiple sclerosis? *Mult Scler Relat Disord* 4:329–333. <https://doi.org/10.1016/j.msard.2015.04.006>
- Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N (2002) Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage* 17:479–489. <https://doi.org/10.1006/nimg.2002.1040>
- Rocca MA, Battaglini M, Benedict RH, De Stefano N, Geurts JJ, Henry RG, Horsfield MA, Jenkinson M, Pagani E, Filippi M (2017) Brain MRI atrophy quantification in MS: from methods to clinical application. *Neurology* 88:403–413. <https://doi.org/10.1212/wnl.0000000000003542>
- Opfer R, Ostwaldt AC, Walker-Egger C, Manogaran P, Sormani MP, De Stefano N, Schippling S (2018) Within-patient fluctuation of brain volume estimates from short-term repeated MRI measurements using SIENA/FSL. *J Neurol*. <https://doi.org/10.1007/s00415-018-8825-8>
- Narayanan S, Nakamura K, Fonov VS, Maranzano J, Caramanos Z, Giacomini PS, Collins DL, Arnold DL (2020) Brain volume loss in individuals over time: Source of variance and limits of detectability. *Neuroimage* 214:116737. <https://doi.org/10.1016/j.neuroimage.2020.116737>

9. Opfer R, Ostwaldt AC, Sormani MP, Gocke C, Walker-Egger C, Manogaran P, De Stefano N, Schippling S (2017) Estimates of age-dependent cutoffs for pathological brain volume loss using SIENA/FSL—a longitudinal brain volumetry study in healthy adults. *Neurobiol Aging* 65:1–6. <https://doi.org/10.1016/j.neurobiolaging.2017.12.024>
10. Battaglini M, Gentile G, Luchetti L, Giorgio A, Vrenken H, Barkhof F, Cover KS, Bakshi R, Chu R, Sormani MP, Enzinger C, Ropele S, Ciccarelli O, Wheeler-Kingshott C, Yiannakas M, Filippi M, Rocca MA, Preziosa P, Gallo A, Biseco A, Palace J, Kong Y, Horakova D, Vaneckova M, Gasperini C, Ruggieri S, De Stefano N (2019) Lifespan normative data on rates of brain volume changes. *Neurobiol Aging* 81:30–37. <https://doi.org/10.1016/j.neurobiolaging.2019.05.010>
11. Kappos L, Fox RJ, Burcklen M, Freedman MS, Havrdová EK, Hennessy B, Hohlfeld R, Lublin F, Montalban X, Pozzilli C, Scherz T, D'Ambrosio D, Linscheid P, Vaclavkova A, Pirozek-Lawniczek M, Kracker H, Sprenger T (2021) Ponesimod compared with teriflunomide in patients with relapsing multiple sclerosis in the active-comparator phase 3 OPTIMUM study: a randomized clinical trial. *JAMA Neurol* 78:558–567. <https://doi.org/10.1001/jama.neurol.2021.0405>
12. Chataway J, De Angelis F, Connick P, Parker RA, Plantone D, Doshi A, John N, Stutters J, MacManus D, Prados Carrasco F, Barkhof F, Ourselin S, Braishner M, Ross M, Cranswick G, Pavitt SH, Giovannoni G, Gandini Wheeler-Kingshott CA, Hawkins C, Sharrack B, Bastow R, Weir CJ, Stallard N, Chandran S (2020) Efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis (MS-SMART): a phase 2b, multiarm, double-blind, randomised placebo-controlled trial. *Lancet Neurol* 19:214–225. [https://doi.org/10.1016/s1474-4422\(19\)30485-5](https://doi.org/10.1016/s1474-4422(19)30485-5)
13. Ashburner J (2007) A fast diffeomorphic image registration algorithm. *Neuroimage* 38:95–113. <https://doi.org/10.1016/j.neuroimage.2007.07.007>
14. Nakamura K, Guizard N, Fonov VS, Narayanan S, Collins DL, Arnold DL (2014) Jacobian integration method increases the statistical power to measure gray matter atrophy in multiple sclerosis. *Neuroimage Clin* 4:10–17. <https://doi.org/10.1016/j.nicl.2013.10.015>
15. Omoumi P, Ducarouge A, Tournier A, Harvey H, Kahn CE Jr, Louvet-de Verchère F, Pinto Dos Santos D, Kober T, Richiardi J (2021) To buy or not to buy—evaluating commercial AI solutions in radiology (the ECLAIR guidelines). *Eur Radiol* 31:3786–3796. <https://doi.org/10.1007/s00330-020-07684-x>
16. Krüger J, Opfer R, Gessert N, Ostwaldt AC, Manogaran P, Kitzler HH, Schlaefer A, Schippling S (2020) Fully automated longitudinal segmentation of new or enlarged multiple sclerosis lesions using 3D convolutional neural networks. *Neuroimage Clin* 28:102445. <https://doi.org/10.1016/j.nicl.2020.102445>
17. Xu M, Ouyang Y, Yuan Z (2023) Deep learning aided neuroimaging and brain regulation. *Sensors (Basel)*. <https://doi.org/10.3390/s23114993>
18. Guha Balakrishnan AZ, Mert R, Sabuncu, John Guttag, Adrian V. Dalca (2019) VoxelMorph: A Learning Framework for Deformable Medical Image Registration.
19. Chen J, Frey EC, He Y, Segars WP, Li Y, Du Y (2022) TransMorph: transformer for unsupervised medical image registration. *Med Image Anal* 82:102615. <https://doi.org/10.1016/j.media.2022.102615>
20. Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC (2011) A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage* 54:2033–2044. <https://doi.org/10.1016/j.neuroimage.2010.09.025>
21. Ashburner J, Ridgway GR (2012) Symmetric diffeomorphic modeling of longitudinal structural MRI. *Front Neurosci* 6:197. <https://doi.org/10.3389/fnins.2012.00197>
22. Opfer R, Krüger J, Spies L, Hamann M, Wicki CA, Kitzler HH, Gocke C, Silva D, Schippling S (2020) Age-dependent cut-offs for pathological deep gray matter and thalamic volume loss using Jacobian integration. *Neuroimage Clin* 28:102478. <https://doi.org/10.1016/j.nicl.2020.102478>
23. Opfer R, Krüger J, Spies L, Ostwaldt AC, Kitzler HH, Schippling S, Buchert R (2022) Automatic segmentation of the thalamus using a massively trained 3D convolutional neural network: higher sensitivity for the detection of reduced thalamus volume by improved inter-scanner stability. *Eur Radiol*. <https://doi.org/10.1007/s00330-022-09170-y>
24. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp 234–241
25. Reuter M, Fischl B (2011) Avoiding asymmetry-induced bias in longitudinal image processing. *Neuroimage* 57:19–21. <https://doi.org/10.1016/j.neuroimage.2011.02.076>
26. Armitage P, Berry G (1994) *Statistical methods in medical research*, 3rd edn. Blackwell Science, Oxford
27. Zhan G, Wang D, Cabezas M, Bai L, Kyle K, Ouyang W, Barnett M, Wang C (2023) Learning from pseudo-labels: deep networks improve consistency in longitudinal brain volume estimation. *Front Neurosci* 17:1196087. <https://doi.org/10.3389/fnins.2023.1196087>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.