Improved dementia prediction in cerebral small vessel disease using deep learning-derived diffusion scalar maps from T1

Yutong Chen BA[1], Daniel Tozer PhD[1], Rui Li MEng[1], Hao Li PhD[2], Anil Tuladhar PhD[2], Frank Erik De Leeuw PhD[2] and Hugh S Markus DM, FMed Sc[1]

Author affiliations:

1 Stroke Research Group, University of Cambridge, Cambridge, UK

2 Radboud University Medical Center, Department of Neurology; Donders Center for Medical Neurosciences, 6500 HB Nijmegen, The Netherlands

Correspondence to: Professor Hugh Markus

Stroke Research Group, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK

hsm32@medschl.cam.ac.uk

Short title: Synthetic diffusion scalar maps predict dementia

Keywords: deep learning, cerebral small vessel disease, image synthesis, dementia prediction, diffusion tensor imaging

Word count: 5980

Abstract

Background: Cerebral small vessel disease (SVD) is the most common pathology underlying vascular dementia. In SVD, diffusion tensor imaging (DTI) is more sensitive to white matter damage and better predicts dementia risk than conventional MRI sequences such as T1 and FLAIR, but DTI takes longer to acquire and is not routinely available in clinical practice. As DTI-derived scalar maps—fractional anisotropy (FA) and mean diffusivity (MD)—are frequently used in clinical settings, one solution is to synthesise FA/MD from T1 images.

Methods: We developed a deep learning model to synthesise FA/MD from T1. The training dataset consisted of 4,998 participants with the highest white matter hyperintensity (WMH) volumes in UK biobank. Four external validations datasets with SVD were included: SCANS (n=120), RUNDMC (n=502), PRESERVE (n=105), NETWORKS (n=26), along with 1,000 normal controls from UK biobank.

Results: The synthetic maps resembled ground truth maps (structural similarity index of >0.89 for MD maps and >0.80 for FA maps across all external validation datasets except for SCANS). The prediction accuracy of dementia using whole-brain median MD from the synthetic maps is comparable to the ground truth (SCANS ground truth c-index 0.822; synthetic 0.821; RUNDMC ground truth 0.816; synthetic 0.812) and better than WMH volume (SCANS 0.534, RUNDMC 0.710).

Conclusions: We have developed a fast and generalisable method to synthesise FA/MD maps from T1 to improve the prediction accuracy of dementia in SVD when DTI data has not been acquired.

Non-standard Abbreviations and Acronyms

CI: confidence interval

DS-GAN: diffusion scalar generative adversarial network

DTI: diffusion tensor imaging

FA: fractional anisotropy

FLAIR: Fluid Attenuation Inversion Recovery

GAN: generative adversarial network

IQR: interquartile range

MD: mean diffusivity

MNI: Montreal Neurological Institute

NAWM: normal-appearing white matter

PSMD: peak width of skeletonized mean diffusivity

PSNR: peak signal to noise ratio

RMSE: root mean squared error

SSIM: structural similarity index measure

SVD: Cerebral small vessel disease

UKB: UK biobank

Introduction

Cerebral small vessel disease (SVD) accounts for a quarter of all strokes, and is the most common pathology underlying vascular cognitive impairment and dementia.[1] Characteristic appearances on magnetic resonance imaging (MRI) include lacunar infarcts, white matter hyperintensities (WMH), and cerebral microbleeds. The extent of WMH correlates with the degree of cognitive impairment, and predicts future stroke and dementia.[2] However, stronger correlations with cognition have been found using diffusion tensor imaging (DTI), which is more sensitive to ultrastructural damages in the white matter. DTI can reveal abnormalities not only in WMH regions, but also in "normal-appearing white matter", which strongly correlates with cognition and predicts future dementia.[3,4] DTI-derived metrics have been proposed as surrogate endpoints to monitor therapeutic intervention in clinical trials in SVD.[3]

However, DTI takes longer to acquire and is not routinely available in clinical practice. As DTI-derived scalar maps—fractional anisotropy (FA) and mean diffusivity (MD)—are widely used in clinical settings,[5] one approach is to synthesise FA/MD from conventional MRI such as T1 and Fluid Attenuation Inversion Recovery (FLAIR) sequences. This has now become possible with deep learning, as T1 and FLAIR share structural similarities with FA/MD maps.[6,7] Deep learning models can detect subtle and intricate features too nuanced for human perception[8] and learn the complex mapping from T1 and/or FLAIR to FA/MD maps. Several models have been proposed to synthesise FA/MD maps from T1 images.[6,7] However, these models have been developed on healthy participants or Alzheimer disease patients and to date have only been evaluated in moderate sample sizes, while lacking external validation to show generalisability. It is uncertain whether these models can capture FA/MD changes in SVD and whether the synthetic maps predict cognitive outcomes as accurately as ground truth maps.

To synthesise FA/MD maps from T1 images in SVD, we developed the diffusion scalar generative adversarial network (DS-GAN). DS-GAN is based on generative adversarial networks (GAN),[9] which is well suited for synthesising images of one type of contrast such as FA/MD, from another such as T1.[10,11] We evaluated how the synthetic maps correlated with the ground truth, how they correlated with cognition and predicted dementia. To evaluate generalisability, we developed the model on a cohort with SVD, and then tested its performance on four independent SVD cohorts.

Methods

Cohort selection

Data availability

The datasets may be shared with researchers upon reasonable request to the corresponding author and after permission by the regulatory authorities. The UK biobank (UKB) data are available via www.ukbiobank.ac.uk/register-apply. Study protocols were not prepared.

Training cohort

The training cohort for DS-GAN was obtained from the UKB, a longitudinal cohort study of 100,000 predominantly healthy individuals under the application number 36509.[12] 4,998 participants with the highest WMH volumes were selected (Figure S1). Participants without T1, FLAIR or DTI were excluded. This cohort is named the UKB_WMH cohort, and was split into training and internal validation sets in a 9:1 ratio.

Validation cohort

The model was validated on four independent cohorts of patients with symptomatic SVD.

1. **S**t George's **C**ognition and Neuroimaging in Stroke (SCANS): 121 participants with severe symptomatic SVD defined as a symptomatic lacunar infarct with confluent WMH (Fazekas grade ≥2).[13]

2. **R**adboud **U**niversity **N**ijmegen **D**iffusion tensor and **M**agnetic **R**esonance Imaging **C**ohort (RUN DMC) study: 503 participants with predominantly mild symptomatic SVD defined as the presence of lacunes and/or any WMH on neuroimaging and accompanying stroke, subacute cognitive or motor symptoms.[14]

3. **H**ow intensively should we treat blood **PRES**sure in established c**ER**ebral small **VE**ssel disease (PRESERVE): multicenter clinical trial including 111 participants with severe symptomatic SVD defined as a symptomatic lacunar infarct with confluent WMH (Fazekas grade ≥2).[15]

4. NETWORKS: 26 participants with severe symptomatic SVD.[16] The inclusion criteria were: symptomatic lacunar infarct with confluent WMH (Fazekas grade ≥2). Fourteen of the participants underwent repeated imaging approximately two weeks after the first imaging. Due to the small sample size, all results pertaining to the NETWORKS study are shown in the supplemental tables.

To test the generalisability of DS-GAN in healthy participants, 1,000 participants with the smallest WMH volumes were selected from the UKB. This is named the UKB_normal dataset.

To investigate the cross-sectional association between baseline MRI metrics and cognition, we examined three cognitive measures (global cognition, executive functioning and processing speed). Cognitive scores were determined as z-scores using published normative data (Table S1), except in the NETWORKS cohort where an associated control cohort was used to normalise the data.

Two SVD cohorts (SCANS and RUNDMC) provided both cross-sectional and prospective longitudinal data with follow-up (5 years in SCANS and 14 years in RUNDMC) which allowed us to not only examine correlations between baseline MRI and cognition, but also to determine whether baseline MRI parameters predicted future dementia.

MRI acquisition parameters are described in Tables S2-S4. Briefly, all T1 images were acquired as 3D at a resolution of 1x1x1 mm. FLAIR images were acquired as 3D in UKB and NETWORKS, and as 2D in SCANS, PRESERVE and RUNDMC.

MRI preprocessing

For UKB, image preprocessing pipeline has been described.[12] For the external validation dataset, image preprocessing followed a similar pipeline described in Figure S2 and detailed in Supplemental Methods section 1.1. Briefly, for T1 images, Gibbs artefacts were removed[17], magnetic field inhomogeneity was corrected[18], and skullstripping[19] and tissue segmentation were performed. For DTI images, eddy correction[20] and diffusion tensor fitting were performed to yield FA/MD maps. FLAIR, MD and FA were rigidly registered to T1.[21] All images were rigidly registered to the Montreal Neurological Institute (MNI) space.

T1 and FLAIR images were used for WMH segmentation using the hypermapper package,[22] except for UKB, where WMH was segmented using BIANCA on FLAIR images.[23]

In all cohorts, the total brain volume (TBV) was calculated using SIENAX[24] by summing the volumes of the grey matter, white matter and ventricles and normalising the total volume by the skull size. The number of lacunes was manually counted by experienced radiologists.

Deep learning

DS-GAN consists of two deep learning models: a generator and a discriminator. The generator synthesises a FA or MD map using T1 and/or FLAIR images. The goal of the discriminator is to classify the map from the generator as "fake" and the ground truth map as "real". The goal of the generator is the opposite: to generate realistic maps so that the discriminator cannot tell the difference of "fake" from "real" maps. The competing goals between generator and discriminator allow both models to improve over time. In this study, the generator was built upon a 3D U-Net architecture (Figure 1) and the discriminator was the 3D extension of PatchGAN.[10]

The following hyperparameters were used: epoch number 60, learning rate 0.0001, batch size 1, and Adam optimizer with $\beta 0$ of 0.5 and $\beta 1$ of 0.999. To augment the training dataset, in

each training epoch, each image was randomly translated, downsampled and random brightness and blurring were applied. The details of the image augmentation, network architecture and hyperparameter selections were in the Supplemental Methods sections 1.2-1.4.

Statistics

Cohort characteristics

The distribution of continuous variables was evaluated using the Shapiro-Wilk test. Descriptive statistics, including mean and standard deviation for normally distributed variables or median and interquartile range (IQR) for non-normally distributed variables, were reported. Demographic variables among three or more datasets were compared using one-way ANOVA for normally distributed data, Kruskal-Wallis test for non-normally distributed data, or Chi-square test with Yates correction for categorical data. For comparisons between two datasets, t-test was used for normally distributed data, Wilcoxon rank-sum test for non-normally distributed data, and Chi-square test with Yates correction for categorical data.

Comparison between synthetic and ground truth maps

Three metrics were used to evaluate the similarity between the synthetic FA/MD maps and the ground truth: peak signal to noise ratio (PSNR),[25] root mean squared error (RMSE), structural similarity index measure (SSIM).[26]

For each validation dataset, 10% of synthetic MD maps were randomly selected for visual evaluation against the ground truth by an independent researcher in terms of presence of artefacts, contrast between the normal-appearing white matter (NAWM) and WMH region, presence of new structures, absence of existing structures and sharpness of the structures seen within the maps (evaluation criteria in Table S5). The researcher was not blinded to the

source or the dataset. All maps were evaluated under a fixed intensity range between 0 and 0.003 mm^2/s. Due to the small sample size of the NETWORKS cohort, six maps were selected.

To investigate the cause for the errors in the synthetic maps, Pearson correlation was computed between the SSIM of the synthetic MD maps with the ground truth, and three different variables: age, WMH volume and T1-to-DTI registration error, which was defined as voxel-wise intensity correlation between T1 and FA.

Correlation of metrics derived from synthetic maps with ground truth

Metrics derived from the ground truth and synthetic FA/MD maps include peak width of skeletonized mean diffusivity (PSMD),[27] and median FA/MD in the whole-brain (grey matter, white matter, sulcal CSF and ventricles), all white matter, NAWM and WMH regions. NAWM area was defined as the white matter mask excluding the WMH lesion area. All calculations were done in the MNI space. Pearson correlation was computed for each metric between the synthetic and ground truth maps.

Reproducibility of synthetic FA/MD-derived metrics

Reproducibility of synthetic FA/MD-derived metrics was evaluated using the baseline and two-week follow-up scans which were available in 14 out of the 26 participants from the NETWORKS study. Reproducibility was defined by the Pearson correlation of the metrics obtained between the two time points.

Correlation with cognition

Pearson correlation was computed between the three cognitive domains and the metrics derived from synthetic and ground truth FA/MD maps. Patients without cognitive data were excluded (0, 0, 1 and 3 patients in SCANS, RUNDMC, PRESERVE and NETWORKS cohorts respectively). For comparisons, correlations were performed with WMH volume and

TBV, a marker of brain atrophy. To test the significance of the correlation, linear regression was performed between cognition and different metrics while adjusting for age and sex. P-values of the slopes were obtained and were adjusted by the Benjamini-Hochberg method separately for each cognitive domain per dataset.

Causal mediation analysis was performed in the RUNDMC cohort to investigate how imaging markers related to SVD—WMH volume, lacune count and TBV—mediated the association between FA/MD-derived metrics and cognitive performance (Supplemental Methods section 1.5).

Prediction of dementia

Univariate Cox proportional hazard models were constructed to predict the onset of dementia using each metric derived from synthetic and ground truth FA/MD maps, WMH volumes and TBV. P-values for the hazard ratio of each metric were adjusted by Benjamini-Hochberg method per dataset.

Software

All image analysis and deep learning models were implemented in python 3.8. All statistical analyses were carried out in R 4.2.0. All CNN models were run on a Nvidia A100 16GiB GPU using pytorch version 1.9.0, cuda version 11.2 and cudnn version 8.1. The computation time to synthesise one MD or FA volume from preprocessed T1 and FLAIR images was 8.5 ± 0.1s on 3 Intel Xeon CPUs and 32 ± 1 ms on a Nvidia A100 GPU. The source code is published (https://github.com/Yutong441/DS-GAN). This manuscript follows the TRIPOD+AI reporting guideline.[28]

Ethical statement

The SCANS study received ethical approval from the London–Wandsworth ethics committee (ukctg.nihr.ac.uk; study ID: 4577). The RUNDMC study received ethical approval from the

Medical Review Ethics Committee region Arnhem-Nijmegen. The PRESERVE study received ethical approval from the Harrow National Research Ethics Service committee (REC number: 11/LO/0458) and is registered with the UK Clinical Research Network (CRN number: 10962). The NETWORKS study received ethical approval from East of England - Cambridge East research ethics committee (reference: 14/EE/0014). All participants provided written informed consent according to the Declaration of Helsinki.

Results

Cohort characteristics

After excluding participants without DTI, T1 or FLAIR images, 4998, 1000, 120, 502, 105 and 26 participants remained in the UKB_WMH, UKB_normal, SCANS, RUNDMC, PRESERVE and NETWORKS datasets respectively (Figure S1). Compared with UKB_normal, participants in the UKB_WMH cohort are older (71 [IQR 67.0- 74.0] vs 55 years old [IQR 52.0- 60.0], P<0.001), less likely to be female (44% vs 60%, P<0.001), and have higher WMH volumes (15.5 [IQR 12.3- 21.9] vs 0.3mL [IQR 0.3- 0.4], P<0.001) (Table 1). Across the external validation datasets, participants in the RUNDMC cohort are younger (64.8 years [IQR 58.1-73.0] vs 69.8 [IQR 63.7-75.6] P< 0.001) and have smaller WMH volumes (2.7 mL [IQR 1.0-7.7] vs 24.1 [IQR 13.2-39.8] P< 0.001) than the other validation datasets.

Development of DS-GAN

We tested the performance of DS-GAN in synthesising FA/MD from different input MRI sequences: T1 and FLAIR, T1 only and FLAIR only. In all external validation sets except UKB_normal, the model using T1 as the sole input achieved the highest performance in most performance metrics (Table S6). The SSIM of the synthetic maps were >0.89 for MD maps and >0.80 for FA maps across all external validation datasets except for SCANS (SSIM=0.818 for MD, 0.766 for FA). The model using FLAIR as the sole input achieved the lowest performance. Therefore, in subsequent analyses, we used the model using T1 as the sole input. Performance of the model using T1 as the sole input is shown in Table 2.

Comparison between synthetic and ground truth maps

DS-GAN successfully synthesised FA/MD maps from T1 in the validation cohorts. The synthetic FA/MD-derived metrics were highly correlated with ground truth (for whole-brain

median MD, R=0.927 [SCANS], 0.907 [RUNDMC], 0.818 [PRESERVE]; 0.757 [UKB normal]). From the synthetic maps, Whole-brain median FA/MD demonstrated higher correlations with the ground truth than median FA/MD in WMH regions and NAWM (Table S7).

Performance of DS-GAN, as assessed by PSNR, SSIM and RMSE, was similar in PRESERVE and RUNDMC (Table 2), but the accuracy of synthesis was lower in SCANS. Performance was highest in normal control (UKB_normal). Variation in the synthesis performance could not be explained by age (R^2<0.1 across all datasets) or WMH volume (R^2<0.1 across all datasets). The synthesis performance was correlated with registration errors (R^2>0.3 across all datasets) (Figure S4).

Synthetic FA/MD maps resembled the ground truth (examples shown in Figure 2). The white matter tracts were clearly visualised on the synthetic FA maps (Figure 2), although in the magnified view of the internal capsule, the synthetic maps appeared smoothened with fewer fine structural details (Figure S3). In visual evaluation, synthetic maps neither created new structures nor missed existing structures in any maps (Table S8). However, they exhibited moderately higher levels of artefacts (in 8%, 8% and 45% of a selection of the SCANS, RUNDMC and PRESERVE cohorts respectively), moderately lower levels of contrast between WMH and NAWM (8%, 0%, and 36%) and moderately lower sharpness in SCANS, PRESERVE (58% and 27%). In RUNDMC, 83% of the evaluated maps have moderately higher levels of sharpness compared with the ground truth.

Reproducibility of synthetic FA/MD-derived metrics

Comparing repeated MRI scans in the NETWORKS study showed a high level of reproducibility in synthetic FA/MD-derived metrics (correlations ranging between 0.927 to 0.996). (Table S9). For the synthetic FA/MD maps, most metrics displayed higher

reproducibility than the ground truth, whereas median MD within NAWM showed lower reproducibility.

Correlation with cognition

Synthetic FA/MD-derived metrics correlated with cognition at baseline almost as well as the ground truth, and better than WMH lesion volume and TBV. For example, for whole-brain median MD, correlations with global cognition were: in SCANS, ground truth -0.450 (P<0.001), synthetic -0.410 (P<0.001), WMH volume -0.138 (P=0.069), TBV 0.240 (P=0.029); in RUNDMC, ground truth -0.518 (P<0.001), synthetic -0.463 (P<0.001), WMH volume -0.278 (P=0.005), TBV 0.330 (P=0.002). Results of correlations for all metrics are shown for global cognition in Table 3, and for executive function and processing speed in Tables S10 and S11 and separately for the NETWORKS cohort in Table S12.

We investigated how SVD imaging markers (WMH volume, lacunes, TBV) mediated the association between cognition and the whole-brain median FA/MD. Between ground truth versus the synthetic maps, SVD imaging markers did not have significantly different effects on the association between cognition and whole-brain median FA/MD. (Table S13, Supplemental Results section 1.2).

Prediction of dementia

The synthetic FA/MD-derived metrics predicted dementia to a similar level to that found with the ground truth. The c-index for prediction by whole-brain median MD were similar in SCANS (ground truth 0.822; synthetic 0.821) and in RUNDMC (ground truth 0.816; synthetic 0.812) (Table 4). The performance of synthetic whole-brain MD in predicting dementia was higher than that for WMH volume (SCANS 0.534, RUNDMC 0.710) and TBV (SCANS 0.709, RUNDMC 0.739). The accuracy of predicting dementia was minimally improved by incorporating three demographic factors (SCANS 0.828, RUNDMC 0.845,

Table S14) and was more highly improved by incorporating three cognitive scores (executive function, processing speed, global cognition) (SCANS 0.903, RUNDMC 0.858, Table S15).

Discussion

We have demonstrated that, using deep learning, it is possible to synthesise FA/MD maps from T1 images in patients with SVD, and that synthetic maps predict future dementia almost as accurately as the ground truth maps, and better than WMH lesion volume. Although the synthetic maps themselves appear to be less sharp and over-smoothened and have not reached the point of replacing DTI in clinical settings, the metrics obtained from these maps correlated well with metrics from the ground truth FA/MD maps and were shown to be reproducible in a cohort with repeat scans. Future studies could evaluate the use of these metrics in clinical trials.

Our study has a number of strengths. This is the first application of deep learning to synthesise FA/MD maps in SVD patients. We used a large training sample (n=4,998). Performance was consistent across 4 independent validation cohorts encompassing a wide range of SVD severity. The model generalisability was corroborated by the high performance in normal controls without SVD (SSIM=0.971±0.007 for MD and 0.903±0.014 for FA respectively). This matched the values reported by previous studies in Alzheimer disease patients (SSIM=0.963±0.009 for MD, 0.959±0.007 for FA)[7] and healthy participants (0.937 for MD, 0.861 for FA).[7]

However, the model also has limitations. The median MD in NAWM is not as accurate compared with whole-brain values. This could be attributed to the lack of sensitivity of T1 images in capturing the subtle changes in NAWM. Contrastingly, the WMH regions display higher contrast in MD, which could be more easily captured by DS-GAN. This inaccuracy in MD calculation in NAWM could explain the low correlation of the associated synthetic

metrics with cognitive performance. Thus, the ability to investigate MD changes in NAWM are reduced compared with DTI-derived FA/MD maps.

Secondly, it is unclear whether FA/MD synthesis could be generalised to other neurological conditions. As the testing samples of DS-GAN mainly consisted of participants with SVD, DS-GAN was not assessed in patients with larger non-lacunar infarcts, tumours or demyelinating lesions such as multiple sclerosis. It is unclear how these other pathologies would be reflected in the synthetic FA/MD maps.

Thirdly, the performance of DS-GAN on SCANS was lower, possibly because of the poor registration between the T1 and the FA/MD maps. This misalignment confounds the evaluation of the voxel-wise similarity between the ground truth and synthetic maps.

Fourthly, combining T1 and FLAIR images into DS-GAN did not improve model performance compared with only using T1. This could be because FLAIR did not convey additional structural information beyond that in T1. Minor structural misalignments between T1 and the co-registered FLAIR images could also limit the accuracy of synthesis. The accuracy of synthesising FA/MD was lower in the model using only FLAIR images, compared with that using only T1 images. This could be because the FLAIR images in the validation datasets were of lower resolution along the axial slices compared to T1. This could limit the application of DS-GAN to high resolution T1 images.

Fifthly, dementia prediction using synthetic FA/MD maps was validated in two out of the five external validation datasets that contain follow-up information on the dementia status. Future studies should evaluate the dementia prediction value of the synthetic maps in more datasets.

Lastly, from the standpoint of rapid acquisition of diffusion scalar maps that correlate with cognition, diffusion-weighted imaging (DWI), which is faster to acquire than DTI, can generate apparent diffusion coefficient (ADC) maps that correlate with cognitive outcomes.[29] However, compared with DWI, the advantage of DS-GAN in synthesising FA/MD from T1 is

retrospective FA/MD synthesis in old datasets where DWI or DTI is unavailable, such as DNA Lacunar.[30] Also, future studies could extend DS-GAN to synthesise other diffusion MRI scalar maps such as orientation dispersion index and isotropic volume fraction.

In conclusion, DS-GAN is a fast, reproducible, and generalisable deep learning model that can synthesise FA/MD maps from T1 images. These synthetic metrics correlate with ground truth and predict dementia in SVD patients almost as well as ground truth FA/MD maps and better than WMH lesion volume, the most widely used SVD-related clinical imaging marker. The model offers a quick and cost effective way to estimate FA/MD-based metrics from conventional MRI sequences.

Acknowledgements

Y.C., D.T. and H.M. conceived the study. H.M. provided the SCANS, PRESERVE and NETWORKS datasets. FE.DL provided the RUNDMC dataset. Y.C., D.T., R.L. and H.L. were involved in preprocessing the T1, FLAIR and DTI images. Y.C. constructed the deep learning model and performed statistical analysis. Y.C. and H.M. drafted the manuscript. All authors were involved in manuscript revision. No patients were involved.

Sources of funding

This research was funded by a British Heart Foundation programme grant (RG/F/22/110052). Infrastructural support was provided by the Cambridge British Heart Foundation Centre of Research Excellence (RE/18/1/34212) and by the Cambridge University Hospitals National Institute for Health and Care Research (NIHR) Biomedical Research Centre (NIHR203312). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Disclosure

The authors report no conflicts of interests.

Supplemental Material

Supplemental Methods

Supplemental Results

Figure S1-S6

Tables S1–S16

References

1. Markus HS, de Leeuw FE. Cerebral small vessel disease: Recent advances and future directions. *Int J Stroke*. 2023;18:4–14.
2. Debette S, Schilling S, Duperron M-G, Larsson SC, Markus HS. Clinical Significance of Magnetic Resonance Imaging Markers of Vascular Brain Injury: A Systematic Review and Meta-analysis. *JAMA Neurology*. 2019;76:81–94.
3. van den Brink H, Doubal FN, Duering M. Advanced MRI in cerebral small vessel disease. *Int J Stroke*. 2023;18:28–35.
4. Zeestraten EA, Lawrence AJ, Lambert C, Benjamin P, Brookes RL, Mackinnon AD, Morris RG, Barrick TR, Markus HS. Change in multimodal MRI markers predicts dementia risk in cerebral small vessel disease. *Neurology*. 2017;89:1869–1876.
5. Ranzenberger LR, Das JM, Snyder T. Diffusion Tensor Imaging. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing; 2024.
6. Gu X, Knutsson H, Nilsson M, Eklund A. Generating Diffusion MRI Scalar Maps from T1 Weighted Images Using Generative Adversarial Networks. In: Felsberg M, Forssén P-E, Sintorn I-M, Unger J, eds. *Image Analysis*. Cham: Springer International Publishing; 2019:489–498.
7. Lan H, Initiative the ADN, Toga AW, Sepehrband F. Three-dimensional self-attention conditional GAN with spectral normalization for multimodal neuroimaging synthesis. *Magnetic Resonance in Medicine*. 2021;86:1718–1733.
8. Feng R, Badgeley M, Mocco J, Oermann EK. Deep learning guided stroke management: a review of clinical applications. *Journal of NeuroInterventional Surgery*. 2018;10:358–362.
9. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks. *arXiv:1406.2661*. 2014. https://arxiv.org/abs/1406.2661. Accessed July 4, 2023.
10. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI: IEEE; 2017.
11. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, Wu X, Zhou J, Shen D, Zhou L. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *NeuroImage*. 2018;174:550–562.
12. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, Bartsch AJ, Jbabdi S, Sotiropoulos SN, Andersson JLR, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19:1523–1536.
13. Zeestraten EA, Benjamin P, Lambert C, Lawrence AJ, Williams OA, Morris RG, Barrick TR, Markus HS. Application of Diffusion Tensor Imaging Parameters to Detect Change in Longitudinal Studies in Cerebral Small Vessel Disease. *PLOS ONE*. 2016;11:e0147836.
14. van Norden AG, de Laat KF, Gons RA, van Uden IW, van Dijk EJ, van Oudheusden LJ, Esselink RA, Bloem BR, van Engelen BG, Zwarts MJ, et al. Causes and consequences of cerebral small vessel disease. The RUN DMC study: a prospective cohort study. Study rationale and protocol. *BMC Neurology*. 2011;11:29.
15. Croall ID, Lohner V, Moynihan B, Khan U, Hassan A, O'Brien JT, Morris RG, Tozer DJ, Cambridge VC, Harkness K, et al. Using DTI to assess white matter microstructure in cerebral small vessel disease (SVD) in multicentre studies. *Clinical Science*. 2017;131:1361–1373.
16. Lawrence AJ, Tozer DJ, Stamatakis EA, Markus HS. A comparison of functional and tractography based networks in cerebral small vessel disease. *Neuroimage Clin*.

2018;18:425–432.
17. Tournier J-D, Smith R, Raffelt D, Tabbara R, Dhollander T, Pietsch M, Christiaens D, Jeurissen B, Yeh C-H, Connelly A. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*. 2019;202:116137.
18. Tustison NJ, Cook PA, Holbrook AJ, Johnson HJ, Muschelli J, Devenyi GA, Duda JT, Das SR, Cullen NC, Gillen DL, et al. The ANTsX ecosystem for quantitative biological and medical imaging. *Sci Rep*. 2021;11:9068.
19. Isensee F, Schell M, Pflueger I, Brugnara G, Bonekamp D, Neuberger U, Wick A, Schlemmer H-P, Heiland S, Wick W, et al. Automated brain extraction of multisequence MRI using artificial neural networks. *Human Brain Mapping*. 2019;40:4952–4964.
20. Schilling KG, Blaber J, Huo Y, Newton A, Hansen C, Nath V, Shafer AT, Williams O, Resnick SM, Rogers B, et al. Synthesized b0 for diffusion distortion correction (Synb0-DisCo). *Magn Reson Imaging*. 2019;64:62–70.
21. Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, Smith SM. Bayesian analysis of neuroimaging data in FSL. *Neuroimage*. 2009;45:S173-186.
22. Mojiri Forooshani P, Biparva M, Ntiri EE, Ramirez J, Boone L, Holmes MF, Adamo S, Gao F, Ozzoude M, Scott CJM, et al. Deep Bayesian networks for uncertainty estimation and adversarial resistance of white matter hyperintensity segmentation. *Human Brain Mapping*. 2022;43:2089–2108.
23. Griffanti L, Zamboni G, Khan A, Li L, Bonifacio G, Sundaresan V, Schulz UG, Kuker W, Battaglini M, Rothwell PM, et al. BIANCA (Brain Intensity AbNormality Classification Algorithm): A new tool for automated segmentation of white matter hyperintensities. *NeuroImage*. 2016;141:191–205.
24. Smith SM, Zhang Y, Jenkinson M, Chen J, Matthews PM, Federico A, De Stefano N. Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*. 2002;17:479–489.
25. Korhonen J, You J. Peak signal-to-noise ratio revisited: Is simple beautiful? In: *2012 Fourth International Workshop on Quality of Multimedia Experience*. 2012:37–38.
26. Renieblas GP, Nogués AT, González AM, Gómez-Leon N, del Castillo EG. Structural similarity index family for image quality assessment in radiological images. *J Med Imaging (Bellingham)*. 2017;4:035501.
27. Baykara E, Gesierich B, Adam R, Tuladhar AM, Biesbroek JM, Koek HL, Ropele S, Jouvent E, Initiative ADN, Chabriat H, et al. A Novel Imaging Marker for Small Vessel Disease Based on Skeletonization of White Matter Tracts and Diffusion Histograms. *Annals of Neurology*. 2016;80:581–592.
28. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Calster BV, Ghassemi M, Liu X, Reitsma JB, Smeden M van, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*. 2024;385:e078378.
29. Jokinen H, Schmidt R, Ropele S, Fazekas F, Gouw AA, Barkhof F, Scheltens P, Madureira S, Verdelho A, Ferro JM, et al. Diffusion changes predict cognitive and functional outcome: The LADIS study. *Annals of Neurology*. 2013;73:576–583.
30. Traylor M, Persyn E, Tomppo L, Klasson S, Abedi V, Bakker MK, Torres N, Li L, Bell S, Rutten-Jacobs L, et al. Genetic basis of lacunar stroke: a pooled analysis of individual patient data and genome-wide association studies. *Lancet Neurol*. 2021;20:351–361.
31. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980*. 2017. http://arxiv.org/abs/1412.6980. Accessed November 3, 2022.
32. Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *arXiv:1711.11585*. 2018. https://arxiv.org/abs/1711.11585. Accessed April 7, 2023.

33. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral Normalization for Generative Adversarial Networks. *arXiv:1802.05957*. 2018. https://arxiv.org/abs/1802.05957. Accessed April 7, 2023.
34. Ramachandran P, Parmar N, Vaswani A, Bello I, Levskaya A, Shlens J. Stand-Alone Self-Attention in Vision Models. *arXiv:1906.05909*. 2019. http://arxiv.org/abs/1906.05909. Accessed April 7, 2023.
35. Paternoster R, Brame R, Mazerolle P, Piquero A. Using the Correct Statistical Test for the Equality of Regression Coefficients. *Criminology*. 1998;36:859–866.
36. Reitan RM. The relation of the trail making test to organic brain damage. *J Consult Psychol*. 1955;19:393–394.
37. Delis DC, Kaplan E, Kramer JH. Delis-Kaplan Executive Function System. *APA PsycTests*. 2001. https://doi.org/10.1037/t15082-000. Accessed July 3, 2023.
38. Nagahama Y, Okina T, Suzuki N, Matsuzaki S, Yamauchi H, Nabatame H, Matsuda M. Factor structure of a modified version of the wisconsin card sorting test: an analysis of executive deficit in Alzheimer's disease and mild cognitive impairment. *Dement Geriatr Cogn Disord*. 2003;16:103–112.
39. Lawrence AJ, Brookes RL, Zeestraten EA, Barrick TR, Morris RG, Markus HS. Pattern and Rate of Cognitive Decline in Cerebral Small Vessel Disease: A Prospective Study. *PLOS ONE*. 2015;10:e0135523.
40. Dumont R, Willis JO, Veizel K, Zibulsky J. Wechsler Adult Intelligence Scale-Fourth Edition. In: Reynolds CR, Vannest KJ, Fletcher-Janzen E, eds. *Encyclopedia of Special Education*. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2014:2520.
41. Ryan JJ, Lopez SJ. Wechsler Adult Intelligence Scale-III. In: Dorfman WI, Hersen M, eds. *Understanding Psychological Assessment.* Boston, MA: Springer US; 2001:19–42.
42. Stroop JR. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*. 1935;18:643–662.
43. Jensen AR, Rohwer WD. The Stroop color-word test: a review. *Acta Psychol (Amst)*. 1966;25:36–93.
44. Mahurin RK, Cooke N. Verbal series attention test: Clinical utility in the assessment of dementia. *The Clinical Neuropsychologist*. 1996;10:43–53.
45. Brand N, Jolles J. Information processing in depression and anxiety. *Psychol Med*. 1987;17:145–153.
46. van der Elst W, van Boxtel MPJ, van Breukelen GJP, Jolles J. The Letter Digit Substitution Test: normative data for 1,858 healthy participants aged 24-81 from the Maastricht Aging Study (MAAS): influence of age, education, and sex. *J Clin Exp Neuropsychol*. 2006;28:998–1009.
47. Rey A. L'examen psychologique dans les cas d'encéphalopathie traumatique. (Les problems.). *APA PsycTests*. 1941. https://psycnet.apa.org/record/1943-03814-001. Accessed July 3, 2023.
48. Kolakowsky-Hayner SA. Wisconsin Card Sorting Test. In: Kreutzer JS, DeLuca J, Caplan B, eds. *Encyclopedia of Clinical Neuropsychology*. New York, NY: Springer; 2011:2719–2720.
49. Weshcler D. Wechsler Memory Scale. *APA PsycTests*. 1945. https://doi.org/10.1037/t27207-000. Accessed August 17, 2023.

Tables

Table 1 Baseline demographics.

	UKB			External validation				
Feature	**UKB_WMH (n=4998)**	**UKB_ normal (n=1000)**	***P***	**SCANS (n=120)**	**RUNDMC (n=502)**	**PRESERVE (n=105)**	**NETWORKS (n=26)**	***P***
Selection criteria	highest WMH volume	lowest WMH volume		SVD (Fazekas ≥2)	SVD (lacunes or WMH)	SVD (Fazekas ≥2)	SVD (Fazekas ≥2)	
Sex (Female), n (%)	2201 (44.0)	600 (60.0)	<0.001	42 (35.0)	219 (43.6)	40 (38.5)	11 (42.3)	0.150
Age (years), median (IQR)	71.0 (67.0-74.0)	55.0 (52.0-60.0)	<0.001	71.4 (63.7-76.9)	64.8 (58.1-73.0)	69.6 (63.1-75.2)	68.9 (58.8-74.2)	<0.001
Race								
Caucasian, n (%)	4640 (92.8)	913 (91.3)	0.104	83 (69.2)	NA	88 (84.6)	24 (92.3)	NA
African, n (%)	13 (0.3)	4 (0.4)	0.664	7 (5.8)	NA	8 (7.7)	2 (7.7)	NA
Asian, n (%)	32 (0.6)	16 (1.6)	0.004	3 (2.5)	NA	8 (7.7)	0 (0)	NA
WMH volume (mL), median (IQR)	15.5 (12.3-21.9)	0.3 (0.3- 0.4)	<0.001	24.2 (14.2-41.6)	2.7 (1.0- 7.7)	26.8 ((15.2-38.2)	27.1 (17.5-41.3)	<0.001
PSMD (10^{-3} mm^2/s), median (IQR)	0.3 (0.2-0.3)	0.2 (0.2-0.2)	<0.001	0.3 (0.3-0.4)	0.3 (0.3-0.4)	0.4 (0.3-0.4)	0.5 (0.4-0.6)	<0.001
Dementia cases, n(%)	NA	NA		23 (19.2)	81 (16.1)	NA	NA	

Abbreviations: PSMD: peak width of skeletonized mean diffusivity, WMH: white matter hyperintensity, UKB_WMH: subset of patients within the UKB that contains the highest WMH lesions, UKB_normal: subset of patients within the UKB that contains the lowest WMH lesions, NA: data unavailable.

Table 2 Comparison of FA/MD maps between the ground truth and those synthesised by DS-GAN, using peak signal to noise ratio (PSNR), root mean square error (RMSE) and structural similarity index measure (SSIM).

Dataset	PSNR		RMSE		SSIM	
	FA	MD	FA	MD	FA	MD
UKB_normal (n=1000)	29.044 (± 0.975)	31.617 (± 0.996)	0.204 (± 0.024)	0.171 (± 0.024)	0.903 (± 0.014)	0.971 (±0.007)
SCANS (n=120)	23.246 (±1.578)	22.517 (± 1.984)	0.513 (±0.093)	0.374 (± 0.081)	0.766 (±0.056)	0.818 (±0.062)
RUNDMC (n=502)	25.364 (±0.921)	26.459 (±0.923)	0.354 (±0.040)	0.271 (±0.028)	0.821 (±0.025)	0.908 (±0.017)
PRESERVE (n=105)	24.195 (±1.042)	24.937 (±1.356)	0.425 (±0.047)	0.288 (±0.038)	0.810 (±0.029)	0.912 (±0.019)

Mean (standard deviation) was displayed in each cell. Higher similarity is reflected by higher PSNR, lower RMSE and higher SSIM. This table displayed the RMSE of synthesising the MD images that had been multiplied by 100. See Table S6 for results pertaining to the NETWORKS cohort.

Table 3 Correlation of the metrics derived from synthetic FA/MD and the ground truth metrics with global cognition.

Metric	SCANS (n=120)		RUNDMC (n=502)		PRESERVE (n=104)	
	Ground truth	Synthetic	Ground truth	Synthetic	Ground truth	Synthetic
WMH volume (mL)	-0.138 .		-0.278 **		-0.225 **	
TBV (mL)	0.240 *		0.332 **		0.069 .	
median FA whole-brain	0.361 **	0.306 **	0.386 ***	0.428 ***	0.212 **	0.150 **
median FA All WM	0.274 **	0.382 ***	0.412 ***	0.365 ***	**0.360** ***	**0.297** **
median FA WMH	0.244 **	0.248 **	0.107 .	0.096	0.339 **	0.149 .
median FA NAWM	0.259 **	0.374 ***	0.408 ***	0.357 ***	0.352 ***	0.294 **
median MD whole-brain	**-0.450** ***	**-0.410** ***	**-0.518** ***	**-0.459** ***	-0.189 **	-0.247 ***
median MD AllWM	-0.255 **	-0.104	-0.458 ***	-0.061	-0.292 **	-0.244 **
median MD WMH	-0.220 *	-0.112	-0.218	-0.214	-0.210 *	-0.051
median MD NAWM	-0.244 *	-0.056	-0.454 ***	-0.034 .	-0.273 **	-0.214 *
PSMD	-0.308 **	-0.303 **	-0.410 ***	-0.385 ***	-0.303 ***	-0.193 **

P values were labelled as: ***: <0.001, **: 0.001-0.01, *: 0.01-0.05, .: 0.05-0.1. Bold text highlights the metrics achieving the highest correlation in each dataset. Abbreviations: FA: fractional anisotropy, MD: mean diffusivity, All WM: all white matter, AWM: abnormal white matter, NAWM: normal-appearing white matter, PSMD:.peak width of skeletonized mean diffusivity. See Table S12 for results pertaining to the NETWORKS cohort.

Table 4 C-index in predicting dementia onset by different metrics in univariate Cox Proportional Hazard models.

Metric	SCANS (n=120)		RUNDMC (n=502)	
	Ground truth	Synthetic	Ground truth	Synthetic
WMH volume (mL)	0.534		0.712 ***	
TBV (mL)	0.709 ***		0.739 ***	
median FA whole-brain	0.776 ***	0.722 **	0.745 ***	0.756 ***
median FA All WM	0.629 .	0.713 **	0.749 ***	0.758 ***
median FA WMH	0.608 .	0.519	0.515	0.535
median FA NAWM	0.626 .	0.718 **	0.748 ***	0.756 ***
median MD whole-brain	**0.822** ***	**0.821** ***	**0.816** ***	**0.812** ***
median MD All WM	0.650 *	0.618 .	0.761 ***	0.614 ***
median MD WMH	0.630 *	0.497	0.692 ***	0.687 ***
median MD NAWM	0.651 *	0.608 .	0.756 ***	0.592 **
PSMD	0.689 ***	0.679 **	0.787 ***	0.768 ***

P values were labelled as: ***: <0.001, **: 0.001-0.01, *: 0.01-0.05, .: 0.05-0.1. Bold text highlights the metrics achieving the highest c-index in each dataset. Abbreviations: TBV: total brain volume, FA: fractional anisotropy, MD: mean diffusivity, All WM: all white matter, AWM: abnormal white matter, NAWM: normal-appearing white matter, PSMD:.peak width of skeletonized mean diffusivity.

Figures

Figure 1 DS-GAN generator structure.

The dimensions of the intermediary outputs are shown as number of channels × image height × image width × image depth. The diagram shows DS-GAN synthesising MD maps, which can also be used to synthesise FA maps (not shown).

Figure 2 Examples of ground truth and synthetic FA/MD maps from five validation datasets.