



## Deep learning from MRI-derived labels enables automatic brain tissue classification on human brain CT

Meera Srikrishna<sup>a,b</sup>, Joana B. Pereira<sup>c,d</sup>, Rolf A. Heckemann<sup>e</sup>, Giovanni Volpe<sup>f</sup>, Danielle van Westen<sup>g,h</sup>, Anna Zettergren<sup>i</sup>, Silke Kern<sup>j</sup>, Lars-Olof Wahlund<sup>c</sup>, Eric Westman<sup>c</sup>, Ingmar Skoog<sup>i,1</sup>, Michael Schöll<sup>a,b,k,l,1,\*</sup>

<sup>a</sup> Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden

<sup>b</sup> Department of Psychiatry and Neurochemistry, Institute of Physiology and Neuroscience, University of Gothenburg, Gothenburg, Sweden

<sup>c</sup> Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden

<sup>d</sup> Clinical Memory Research Unit, Department of Clinical Sciences, Lund University, Malmö, Sweden

<sup>e</sup> Department of Medical Radiation Sciences, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg, Sweden

<sup>f</sup> Department of Physics, University of Gothenburg, Gothenburg, Sweden

<sup>g</sup> Department of Clinical Sciences, Diagnostic Radiology, Lund University Sweden

<sup>h</sup> Department of Imaging and Function, Skånes University Hospital, Lund, Sweden

<sup>i</sup> Neuropsychiatric Epidemiology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, Centre for Ageing and Health (AgeCap), University of Gothenburg, Gothenburg, Sweden

<sup>j</sup> Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Mölndal, Sweden

<sup>k</sup> Dementia Research Centre, Institute of Neurology, University College London, London, UK

<sup>l</sup> Department of Clinical Physiology, Sahlgrenska University Hospital, Gothenburg, Sweden

### ARTICLE INFO

#### Keywords:

Brain image segmentation  
computed tomography (CT)  
Deep learning  
Convolutional neural networks (CNN)

### ABSTRACT

Automatic methods for feature extraction, volumetry, and morphometric analysis in clinical neuroscience typically operate on images obtained with magnetic resonance (MR) imaging equipment. Although CT scans are less expensive to acquire and more widely available than MR scans, their application is currently limited to the visual assessment of brain integrity and the exclusion of co-pathologies. CT has rarely been used for tissue classification because the contrast between grey matter and white matter was considered insufficient. In this study, we propose an automatic method for segmenting grey matter (GM), white matter (WM), cerebrospinal fluid (CSF), and intracranial volume (ICV) from head CT images. A U-Net deep learning model was trained and validated on CT images with MRI-derived segmentation labels. We used data from 744 participants of the Gothenburg H70 Birth Cohort Studies for whom CT and T1-weighted MR images had been acquired on the same day. Our proposed model predicted brain tissue classes accurately from unseen CT images (Dice coefficients of 0.79, 0.82, 0.75, 0.93 and 0.98 for GM, WM, CSF, brain volume and ICV, respectively). To contextualize these results, we generated benchmarks based on established MR-based methods and intentional image degradation. Our findings demonstrate that CT-derived segmentations can be used to delineate and quantify brain tissues, opening new possibilities for the use of CT in clinical practice and research.

### 1. Introduction

Image-based tissue classification is an integral part of many analysis procedures in neuroimaging. It involves distinguishing tissue classes in a brain image based on, for example, signal intensity and prior probability maps. Specifically, mapping grey matter (GM), white matter (WM), cerebrospinal fluid (CSF), brain volume (BV) and intracranial

volume (ICV) has widespread applications in quantitative brain analysis and morphological research (Driscoll et al., 2009; Erickson et al., 2014; Fotenos et al., 2008; Gautam et al., 2014; Grieve et al., 2013). Importantly, it enables measurement of brain atrophy in cortical and subcortical brain regions (Pini et al., 2016) and their relationship with cognitive decline during ageing and in neurodegenerative diseases. In

**Abbreviations:** CT, X-ray computed tomography; MRI, Magnetic resonance imaging; GM, Grey matter; WM, White matter; CSF, Cerebrospinal fluid; ICV, Intracranial volume; CNN, Convolutional neural networks; AD, Alzheimer's disease; HD, Hausdorff distances; VE, Volumetric error.

\* Corresponding author: Michael Schöll, Sahlgrenska University Hospital, MedTech West, Röda stråket 10B, 41324 Gothenburg, Sweden.

E-mail address: [michael.scholl@neuro.gu.se](mailto:michael.scholl@neuro.gu.se) (M. Schöll).

<sup>1</sup> These authors contributed equally.

<https://doi.org/10.1016/j.neuroimage.2021.118606>.

Received 11 March 2021; Received in revised form 15 September 2021; Accepted 20 September 2021

Available online 25 September 2021.

1053-8119/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

line with this, GM atrophy associated with neurodegeneration is a common biomarker used to characterize neurodegenerative disorders.

The most frequently used structural neuroimaging modalities are X-ray computed tomography (CT) and magnetic resonance (MR) imaging. CT uses dedicated X-ray equipment to generate 3D image volumes of the body, whereas MR uses electromagnetic fields to record and map the spatial variation of images according to the properties of the tissues (Jacobs et al., 2007; Westbrook and Talbot, 2018). In clinical settings, CT and MR imaging are the main examination tools for the structural assessment of brain abnormalities in dementia disorders (Ashburner et al., 1997; Pasi et al., 2011; Wattjes et al., 2009).

Due to its high spatial resolution and tissue contrast, MRI is currently the most common imaging modality used to measure brain atrophy (Despotović et al., 2015). However, CT equipment is substantially more common than MR equipment (Supplementary Fig. 1). In Europe, there were 1.24 CT scanners for every MR scanner in 2017 (Eurostat 2021)source: European statistics, ("Eurostat, Medical technology,"), and in China, the ratio was 2.56 in 2013 (He et al., 2018). Moreover, CT scans can be acquired more rapidly, suiting the needs of people with dementia, as restlessness and reduced ability to cooperate is common in this patient group (Hort et al., 2010; Musicco et al., 2004; Stewart, 2001). Most primary healthcare centres and hospitals refer patients with cognitive symptoms for CT scanning for visual assessment of brain integrity and exclusion of co-pathologies (Dane Rayment et al., 2016; Musicco et al., 2004). Evidence suggests that visual ratings of brain volume changes derived from CT are usable predictors for dementia diagnostics with comparable diagnostic properties to visual assessment of MR scans (Sacuiu et al., 2018; Thiagarajan et al., 2018). Yet, quantitative analysis of CT images for cortical atrophy and white matter changes remains largely unexplored.

Brain image segmentation on CT is a challenging task due to lower soft-tissue contrast in comparison to MR imaging. Few studies have proposed the extraction of brain volumes from CT using typical MR image analysis software such as FSL (Cauley et al., 2020, 2018) or Freesurfer (Manniesing et al., 2017). In other studies, methods have been developed to identify brain tissue classes directly from head CT based on techniques such as region growing (Sandor et al., 1991), adaptive intensity thresholding (Gupta et al., 2010), mixed modelling (Aguilar et al., 2015) and probabilistic classification using Hounsfield units (Kemmling et al., 2012). However, these studies produced coarse segmentations and lacked thorough evaluation.

Recently, deep learning methods have shown promise for extracting valuable information from CT scans. Convolutional neural networks (CNN) are a type of deep learning network predominantly used for image analysis. CNNs trained on MR datasets have been used experimentally for skull stripping (Kleesiek et al., 2016), detection of atrophy in Alzheimer's disease (AD) (Suzuki, 2017), segmentation of fine brain structures (Gibson et al., 2018), multimodal imaging-based predictions (Lu et al., 2018), and creation of pseudo-CT images from other modalities for attenuation correction (Han, 2017; Liu et al., 2018). In CT, deep learning methods have a wide range of applications such as segmentation and detection of tumours (Mlynarski et al., 2019), ischemic lesions (Clèrigues et al., 2019), and intracranial haemorrhages (Lee et al., 2019); prediction of AD by classification (Gao et al., 2017); and reconstruction of low dose CT (Chen et al., 2017). In particular, U-Net, a CNN developed in 2015 for biomedical image segmentation, allows image information classification at a pixel level. Until now, few studies have explored U-Net based semantic segmentation of tissue classes (Van De Leemput et al., 2019) and brain extraction from head CT scans (Akkus et al., 2019). Evidence from these studies suggests that using deep learning methods trained with suitable labels can extract valuable information from head CT data reducing the necessity to use MR imaging.

The aim of this study was to develop a method for automatically segmenting GM, WM, CSF, and ICV from any given brain CT scan. We trained this model using MRI-derived segmentation labels and assessed

the model's accuracy by comparing CT segmentation results with MR segmentation results. Our findings show that robust, accurate and reproducible segmentation of brain tissue classes in CT can be obtained using our model, which is of great interest for the first-line clinical assessment of neurodegenerative diseases.

## 2. Materials and methods

### 2.1. CT and MR image datasets

We obtained paired CT and MR datasets from the Gothenburg H70 Birth Cohort Studies. These multidisciplinary longitudinal epidemiological studies include six birth cohorts with baseline examinations at the age of 70 to study the elderly population of Gothenburg, Sweden. For the present study, we included same-day acquisitions of CT and MR images from 744 participants (52.6% female, mean age  $70.44 \pm 2.6$  years) of the cohort born in 1944, collected from 2014 to 2016. The full study details are reported elsewhere (Rydborg Sterner et al., 2019).

The H70 study was approved by the Regional Ethical Review Board and by the Radiation Protection Committee in Gothenburg, Sweden. Study participants gave informed consent in writing before data collection. Brain imaging was conducted at Aleris Röntgen Annedal in Gothenburg (Aleris Healthcare AB, Stockholm, Sweden). CT images were acquired on a 64-slice Philips Ingenuity CT system with a slice thickness of 0.9 mm, an acquisition matrix of  $512 \times 512$  and voxel size  $0.5 \times 0.5 \times 5.0$  mm<sup>3</sup> (Philips Medical Systems, Best, Netherlands). MR scanning was conducted on a 3-Tesla Philips Achieva system (Philips Medical Systems) using a T1-weighted sequence with the following parameters: field of view  $256 \times 256 \times 160$  voxels, voxel size:  $1 \times 1 \times 1$  mm<sup>3</sup>, echo time: 3.2 ms, repetition time: 7.2 ms, flip angle: 9° (Rydborg Sterner et al., 2019).

### 2.2. Deep learning model development

#### 2.2.1. Image pre-processing

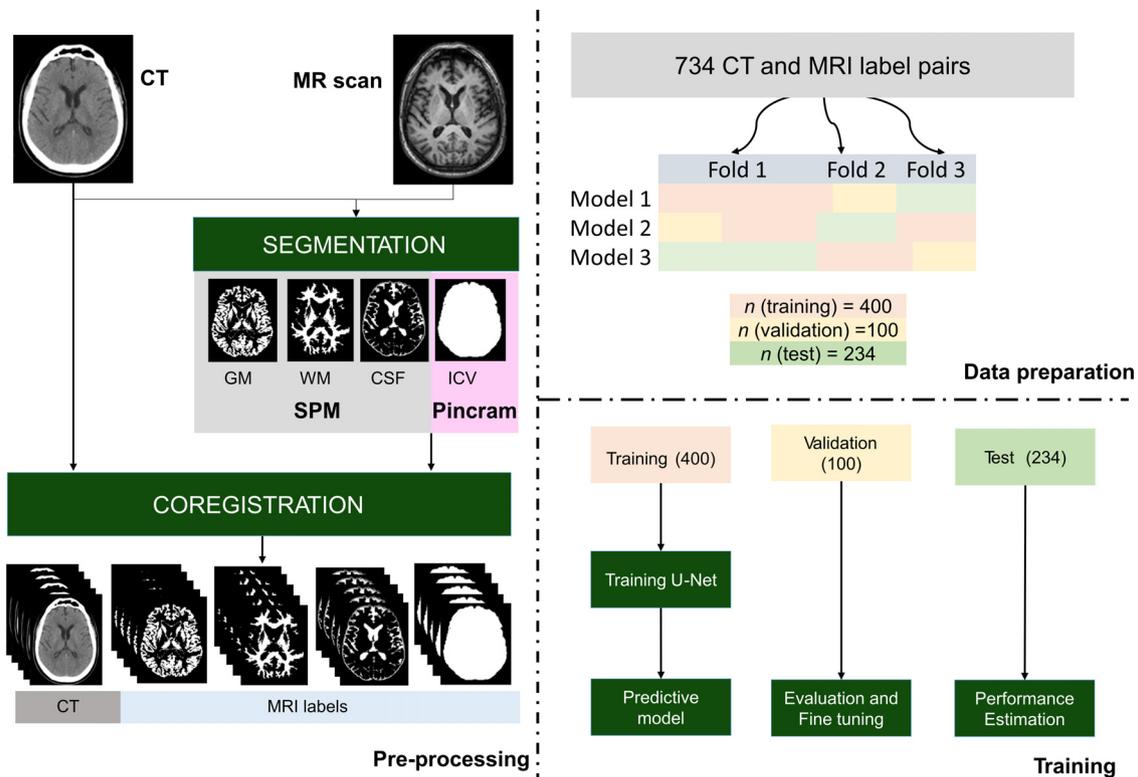
All images were pre-processed using SPM12 (<http://www.fil.ion.ucl.ac.uk/spm/>), running on MATLAB 2018a, and Pinfram (Heckemann et al., 2015). CT and MR images were converted to NIFTI format. Before pre-processing, the quality and integrity of all the scans were assessed. Each image was aligned to the AC-PC line.

**2.2.1.1. Segmentation.** The GM, WM, and CSF labels were derived from the MR images using the unified segmentation routine in SPM12 (Ashburner and Friston, 2005). We used Pinfram software to label the intracranial volume on each MR image. In the following, we will refer to the GM, WM, CSF, and ICV segmentation output from the MR image processing as *MR labels*.

**2.2.1.2. Co-registration.** To enable training of the model, the CT images and MR labels needed to be represented in a common image matrix. To achieve this, we paired each CT image with the corresponding (same participant) MR image and applied the co-registration function in SPM12 with 12 degrees of freedom (Ashburner and Friston, 2007). We applied the resulting affine transformation to the MR labels. SPM12 co-registration module optimises the transformation by minimizing or maximizing an objective function or cost function (Ashburner et al., 1997). Normalized mutual information was used for optimization. The alignment of the MR labels with the CT image was visually assessed for each data set. Ten data sets were excluded due to faulty co-registrations.

#### 2.2.2. Data preparation

The 734 pre-processed datasets consisting of CT images and their paired, co-registered MR labels were subdivided into training and cross-validation groups. We opted for three-fold cross-validation and developed three models for each tissue class. For this, the datasets were randomly split into three folds (Fig. 1). For each model, one fold was held



**Fig. 1.** Processing stages: CT and T1-weighted MR scans from 734 70-year-old individuals from the Gothenburg H70 Birth Cohort Studies were split into training, validation, and unseen test datasets. In the pre-processing stage, CT-MRI pairs were co-registered. MR images were pre-processed with SPM12 to extract grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) tissue class labels and with Pincrum to obtain intracranial volume (ICV) masks. A U-Net deep learning model was developed and underwent training to predict ICV, GM, WM, and CSF, with CT images and paired MR labels as training inputs.

out as the unseen test dataset, and two folds were further grouped into 400 training and 100 validation datasets. The model parameters were trained using the training datasets and fine-tuned with the validation dataset. The performance of the model was gauged on the unseen test datasets.

### 2.2.3. Model training

The deep learning model used in the training stage was developed in Python 3.7, using TensorFlow 2.0 and Keras 2.3.1. The models were trained using an Nvidia GeForce RTX 2080 Ti graphical processing unit with 11 GB of random-access memory (Nvidia Corp., Santa Clara, CA, USA). We used U-Net to carry out semantic segmentation, where each voxel in the input CT image is assigned a unique tissue class. The architecture of the model is depicted in Fig. 2.

The CT images were used as training inputs, and the MR labels were used as training labels. The U-Net was designed to accept 2D slices of  $512 \times 512$  pixels. Each input image was processed as a stack of 30 2D slices, and each slice was used as training input. In total, 12,000 training slices and 3,000 validation slices were used. Once the datasets were prepared and grouped, the model was created with 1,177,649 trainable hyper-parameters. The inputs were fed into the model, and learning was executed using the Keras module. The batch size was 16. Callback features were used (early stopping, automatic reduction of learning rate with respect to rate of training). The model was trained for 50 epochs with 750 samples per epoch in approximately 540 mins.

The decoder path of U-Net learned the specific features of the input images with the hyperparameters. The encoder path mapped these features to the respective pixel. In the initial epoch, the trainable parameters were selected at random for learning and feature specification (forward propagation). Once an epoch was completed, the model predicted the segmentation on the inputs. The loss between the prediction and inputs was calculated, and the trainable parameters were corrected

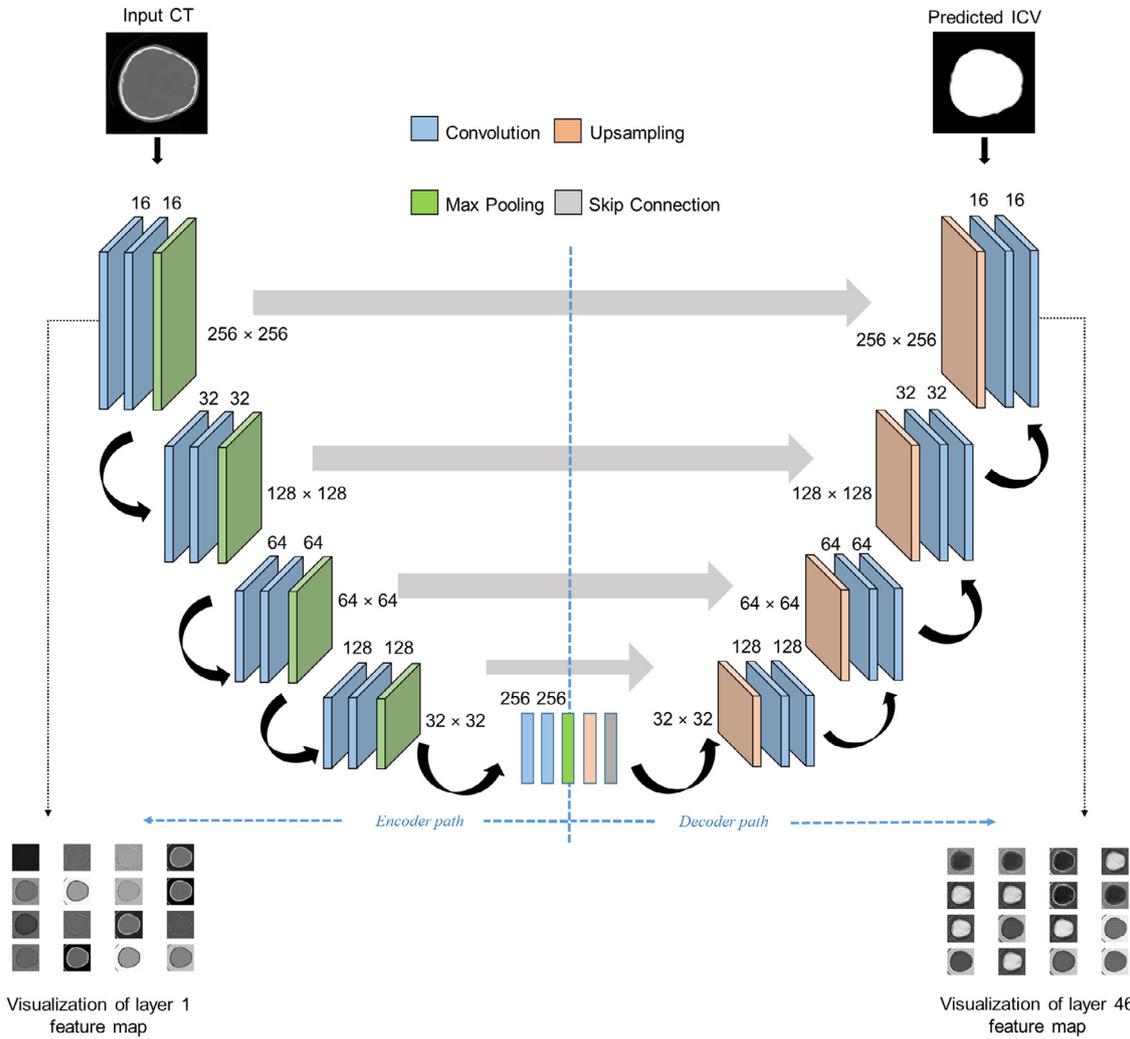
with respect to the loss (error backpropagation). As we are performing pixel-wise classification, binary cross-entropy was employed as the loss function. We used the adaptive moment estimator (Adam) optimizer (Kingma and Ba, 2014) with a learning rate of  $10^{-5}$  to estimate these parameters. All weights were initialized by a normal distribution with a mean of 0 and a standard deviation of 0.01, and all biases were initialized to 0. Once the trainable parameters were adjusted, the model proceeded to the next epoch. The propagation algorithms were repeated for all epochs, and parameters were updated. The training stopped when the model was saturated or when the number of epochs was completed.

### 2.3. Deriving CT based segmentation maps

Once the training was completed, the models were saved for the automated segmentation of CT scans. Deep learning-based CT segmentation is executed in Python 3.7, using TensorFlow 2.0 and Keras 2.3.1. The model prediction is independent of MRI, and pre-processing of CT images is not necessary. To segment a CT image using trained U-Net models, it is provided as an input to the model. The model then predicts various brain tissue class maps from CT scans using the trained hyper-parameters in less than one minute per dataset. The segmentation maps are probability maps or confidence maps, with each pixel specifying the probability of belonging to the particular tissue class. They are real images within the  $[0,1]$  interval. These segmentation maps can be used for either visualization or volumetric analysis. BV maps were derived by adding GM and WM tissue class maps.

### 2.4. Evaluation

To assess the method, we performed two sets of evaluations. In the first set, the model predictions were assessed using evaluation metrics



**Fig. 2.** Model architecture: Overview of internal layers in U-Net utilized to perform brain tissue class segmentation. U-Net comprises symmetrically aligned CNNs with two processing paths: the encoder path (for capturing the context of the input image using the label images) and the decoder path (for the localization and placement of extracted features to corresponding voxels). Intermediate layers capture various feature representations. Feature maps represent the output of intermediate layers.

used for segmentation similarity comparison. The second assessment involved comparing the CT-based deep learning-derived segmentation algorithm with various MR-based classification algorithms. We also estimated the loss of accuracy incurred when using CT instead of MRI for brain tissue classification.

#### 2.4.1. Assessment of model predictions

For this assessment, the MR labels were employed as the standard or reference criterion for comparison and evaluation. Following the approach suggested by the MRbrainS challenge (Mendrik et al., 2015), an online framework to evaluate segmentation models in MR imaging, the comparison of similarity between the predicted masks and standard criterion in this study was assessed using four measures: continuous Dice coefficient, Pearson correlation of volumetric measures, Hausdorff distances (HD) and volumetric error (VE).

**2.4.1.1. Continuous Dice coefficient ( $d_c$ ).** We used the continuous Dice score, a variant of the Dice coefficient that assesses spatial similarity between binary images and real-valued probability maps (Shamir et al., 2019).

**2.4.1.2. Correlation of volumetric measures ( $r$ ).** For the present study, we calculated the correlations between the volumetric measures derived from MR labels and the predicted segmentation labels. The MR labels

were binarized segmentations, whereas the predictions were real-valued probability maps. Hence, the predicted ICV masks were binarized by global thresholding at 0.5. The 30 slices were stacked to obtain a 3D image, and the sum of the pixels was computed to obtain the volume.

In case of the predicted GM, WM and CSF masks, we utilized a data-driven binarization technique. After stacking the slices to obtain a 3D image, at each voxel, we compared the intensities between all three tissue class probability maps. For each voxel, the tissue class with maximum intensity was assigned to it. This binarization approach logically assigns a single class to a particular voxel without discarding information.

**2.4.1.3. Volumetric error (VE).** Is a measure that considers the volumes of the segments to indicate similarity. It is the absolute volumetric difference divided by the sum of the compared volumes. For a given MR label  $T$  and the predicted CT segmentation  $P$ , if  $V_T$  and  $V_P$  are volumes of  $T$  and  $P$ , then  $VE$  is derived as:

$$VE = \frac{2||V_T - V_P||}{V_T + V_P}$$

**2.4.1.4. Hausdorff distance (HD).** HD measures the maximum Euclidean distance of a set to the nearest point in the other set. In this study, four HD variants were calculated for evaluating the model:

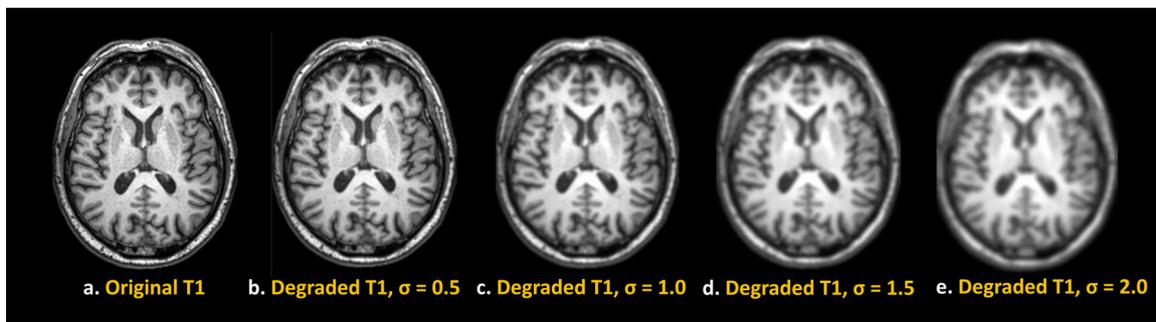


Fig. 3. Application of Gaussian filter with various standard deviation on T1 images. (a) depicts the original T1 weighted MR image and (b-e) shows the degraded T1 MR image by applying Gaussian blur with standard deviation  $\sigma$  of 0.5, 1, 1.5 and 2.

- i Average HD (AHD) – the average of directional HDs from  $T$  to  $P$  and  $P$  to  $T$
- ii Forward HD (FHD) – directional HD from  $T$  to  $P$
- iii Reverse HD (RHD) – directional HD from  $P$  to  $T$
- iv Modified HD (MHD) – maximum of directional HDs from  $T$  to  $P$  and  $P$  to  $T$

#### 2.4.2. Comparison between deep learning-based CT segmentation and established tissue classification algorithms

Since standard automated methods for CT brain tissue segmentation are not commonly available, we performed additional analyses using tissue classification algorithms that work on MR images for reference. We performed three comparisons using continuous Dice coefficients and correlation of volumetric measures with binarization by global thresholding.

**2.4.2.1. Comparison of SPM with other tissue classification algorithms.** To generate a benchmark for assessing the output of tissue classification algorithms, we compared MR labels generated with SPM12 ( $MRI_{spm}$ ) with those derived from two other segmentation algorithms: FSL FAST ( $MRI_{fsl}$ ) (Zhang et al., 2001) and NiftySeg ( $MRI_{niftyseg}$ ) (Cardoso et al., 2011) for a subset of the data ( $n=234$ ). Further details of the evaluation are provided in Supplementary material S1.

**2.4.2.2. Comparison of model performance when training with alternative MR labels.** To understand the impact of the MR image segmentation algorithm on model predictions, we trained three U-Net models with identical architectures and input CTs (from datasets not used in the previous analysis) but with MR labels generated with three distinct segmentation algorithms: SPM12, FSL FAST, and NiftySeg. For detailed steps, refer to Supplementary material S2. After training the models, we acquired three sets of predictions from SPM-trained models ( $CT_{spm}$ ), FSL-trained models ( $CT_{fsl}$ ), and NiftySeg-trained models ( $CT_{niftyseg}$ ) for the same subset (CTs of the MR images used in the previous analysis) of unseen data and compared these predictions.

**2.4.2.3. Comparison of CT based segmentation with MRI based segmentation and estimating the loss of accuracy in using CT instead of MRI.** To understand the difference in the segmentations derived from the two modalities, we first compared CT segmentations directly with MR segmentations in GM, WM and CSF tissue classes. After that, we compared the MRI-MRI overlap agreement to the CT-MRI overlap agreement. We compared  $CT_{spm}$  with  $MRI_{fsl}$ , using  $MRI_{spm}$  as the benchmark.  $MRI_{fsl}$  was an arbitrary choice. We observed that  $MRI_{fsl}$  agrees with the  $MRI_{spm}$  reference significantly more strongly than  $CT_{spm}$  (Fig. 4, Supplementary Fig. 6). To appraise the loss of accuracy incurred by using CT instead of MR imaging for tissue classification, we conducted an analysis where we run FSL on a degraded MR image. Degradation was achieved by introducing varying amounts of blur using Gaussian filtering with standard deviation ( $\sigma$ ) of 0.5, 1, 1.5 and 2 to the T1 MR image (Fig. 3). We then

determined which amount of blur degraded the image sufficiently to bring the  $MRI_{spm}$ - $MRI_{fsl}$  overlap down to the levels similar to  $CT_{spm}$ - $MRI_{spm}$  (Fig. 8, Supplementary Fig. 6, 8).

### 3. Results

The model output was assessed based on how well it predicted brain tissue segmentation labels in the test datasets. For each CT image, we determined the GM, WM, CSF, and ICV labels and their volumes. Examples of the predicted segmentations of different tissue class labels derived from test CTs are shown in Figs. 4 and 5.

#### 3.1. Assessment of model predictions

Table 1 presents all averaged resulting metrics obtained for this analysis. Continuous Dice coefficients were calculated to evaluate the spatial overlap between predicted CT tissue maps and MR labels of the test dataset, yielding  $d_c$  of 0.79, 0.82, 0.75, 0.93 and 0.98 in GM, WM, CSF, BV and ICV, respectively. These coefficients indicate very good to excellent spatial overlap between segmentations that were independently derived from CT and MRI. The boundary measures, AHD and MHD, were found to be low for WM ( $4.10 \pm 1.8$ ,  $1.19 \pm 0.8$  mm) and CSF ( $4.43 \pm 1.5$ ,  $1.42 \pm 0.6$  mm). Pearson  $r$  coefficients were calculated between the volumes of predicted masks and the MR labels (Fig. 6). Strong correlations of  $r=0.93$ ,  $r=0.96$ ,  $r=0.91$ ,  $r=0.98$  and  $r=0.97$  were observed for GM, WM, CSF, BV and ICV volumes, respectively. A trend of underestimation was observed mainly in predicting WM and CSF volumes (Fig. 7). Low VE of  $0.03 \pm 0.04$ ,  $0.03 \pm 0.02$ , and  $0.02 \pm 0.01$  was observed in WM, ICV, and BV, respectively, and somewhat higher VE of  $0.06 \pm 0.03$ ,  $0.06 \pm 0.04$  for GM and CSF, respectively.

#### 3.2. Comparison between deep learning-based CT segmentation and established tissue classification algorithms

$MRI_{spm}$  showed strong agreement both in terms of continuous Dice coefficients and in volumetric correlations with  $MRI_{fsl}$  and  $MRI_{niftyseg}$  (Table 2.a, Supplementary fig. 2, 3). When we compared the CT segmentations of U-Net model predictions trained MR labels derived from SPM, FSL FAST, and NiftySeg, there was strong volumetric correlation and spatial overlap between all pairings in GM and WM. There was a strong volumetric correlation in CSF but less spatial overlap in comparison to GM and WM (Table 2.b, Supplementary fig. 4, 5).

CT segmentations showed strong spatial overlap and volumetric correlation with MR segmentations derived from all three segmentation algorithms, notably GM and WM (Table 2.c, Supplementary fig. 6, 7). We compared the spatial overlap between  $CT_{spm}$ - $MRI_{spm}$  with  $MRI_{spm}$ - $MRI_{fsl}$  and  $MRI_{spm}$ - $dMRI_{fsl}$ , where  $d$  implies degraded T1 images by Gaussian filtering. By comparing  $d_c$  values for these pairings, we observed that performing tissue classification with CT instead of MR imaging incurs a loss of accuracy similar to or less than that of performing

**Table 1**

Model evaluation metrics in test datasets ( $n=234$ ). Continuous Dice score ( $d_c$ ) expresses the extent of spatial similarity between CT predictions and MR labels, Pearson's correlation coefficient ( $r$ ) measures the linear relationship between volumetric measures, AHD, MHD, RHD, FHD are boundary measures, and VE expresses the absolute volumetric difference between CT segmentation outputs and MR labels. The volumes of CT predictions and MR labels are given by  $V_{CT}$  and  $V_{MR}$ , respectively

| Metrics              | Model   | GM        | WM        | CSF       | ICV        | BV         |
|----------------------|---------|-----------|-----------|-----------|------------|------------|
| dc                   | Model 1 | 0.77±0.05 | 0.81±0.07 | 0.73±0.06 | 0.99±0.008 | 0.93±0.009 |
|                      | Model 2 | 0.79±0.03 | 0.83±0.02 | 0.75±0.06 | 0.98±0.01  | 0.94±0.008 |
|                      | Model 3 | 0.8±0.03  | 0.83±0.02 | 0.76±0.06 | 0.97±0.006 | 0.93±0.009 |
|                      | Mean    | 0.79      | 0.82      | 0.75      | 0.98       | 0.934      |
| r                    | Model 1 | 0.92      | 0.94      | 0.9       | 0.96       | 0.96       |
|                      | Model 2 | 0.96      | 0.98      | 0.91      | 0.98       | 0.99       |
|                      | Model 3 | 0.93      | 0.97      | 0.92      | 0.99       | 0.99       |
|                      | Mean    | 0.93      | 0.96      | 0.91      | 0.97       | 0.98       |
| VE                   | Model 1 | 0.06±0.3  | 0.03±0.04 | 0.06±0.05 | 0.02±0.02  | 0.03±0.02  |
|                      | Model 2 | 0.06±0.03 | 0.03±0.02 | 0.06±0.04 | 0.04±0.02  | 0.02±0.01  |
|                      | Model 3 | 0.07±0.03 | 0.03±0.03 | 0.07±0.06 | 0.02±0.01  | 0.02±0.02  |
|                      | Mean    | 0.06      | 0.03      | 0.06      | 0.03       | 0.02       |
| AHD                  | Model 1 | 4.85±2    | 4.36±2    | 4.52±1.5  | 3.75±2.1   | 4.13±2.24  |
|                      | Model 2 | 4.82±1.9  | 4.10±1.8  | 4.7±1.5   | 3.80±1.1   | 4.15±2.06  |
|                      | Model 3 | 4.60±1.6  | 4.10±1.8  | 4.43±1.5  | 3.86±1.1   | 4.07±1.7   |
|                      | Mean    | 4.7       | 4.19      | 4.57      | 3.8        | 4.12       |
| MHD                  | Model 1 | 1.72±1    | 1.24±0.8  | 1.45±0.7  | 0.91±1     | 1.34±0.5   |
|                      | Model 2 | 1.72±0.9  | 1.19±0.8  | 1.50±0.6  | 0.93±0.3   | 1.37±0.7   |
|                      | Model 3 | 1.67±1    | 1.19±0.8  | 1.42±0.6  | 0.92±0.3   | 1.38±0.8   |
|                      | Mean    | 1.71      | 1.21      | 1.46      | 0.92       | 1.36       |
| RHD                  | Model 1 | 1.67±1    | 1.16±0.8  | 1.38±0.7  | 0.865±1    | 1.21±0.7   |
|                      | Model 2 | 1.64±0.9  | 1.10±0.8  | 1.40±0.7  | 0.93±0.3   | 1.23±0.6   |
|                      | Model 3 | 1.60±1    | 1.11±0.8  | 1.41±0.7  | 0.92±0.3   | 1.38±0.7   |
|                      | Mean    | 1.7       | 1.2       | 1.45      | 0.91       | 1.28       |
| FHD                  | Model 1 | 1.72±0.9  | 1.24±0.8  | 1.45±0.7  | 0.91±1     | 1.34±0.5   |
|                      | Model 2 | 1.72±0.9  | 1.19±0.7  | 1.50±0.6  | 0.86±0.3   | 1.37±0.7   |
|                      | Model 3 | 1.67±0.9  | 1.19±0.8  | 1.42±0.6  | 0.85±0.3   | 1.24±0.7   |
|                      | Mean    | 1.64      | 1.12      | 1.37      | 0.87       | 1.31       |
| $V_{CT}$<br>(litres) | Model 1 | 0.61±0.06 | 0.51±0.06 | 0.28±0.05 | 1.65±0.19  | 1.13±0.12  |
|                      | Model 2 | 0.6±0.06  | 0.52±0.06 | 0.27±0.04 | 1.59±0.24  | 1.12±0.11  |
|                      | Model 3 | 0.59±0.07 | 0.5±0.08  | 0.27±0.06 | 1.67±0.19  | 1.1±0.15   |
|                      | Mean    | 0.71      | 0.41      | 0.38      | 1.64       | 1.12       |
| $V_{MR}$<br>(litres) | Model 1 | 0.58±0.05 | 0.54±0.06 | 0.29±0.04 | 1.64±0.18  | 1.11±0.11  |
|                      | Model 2 | 0.56±0.05 | 0.57±0.05 | 0.28±0.04 | 1.60±0.24  | 1.09±0.11  |
|                      | Model 3 | 0.55±0.07 | 0.52±0.08 | 0.27±0.05 | 1.66±0.18  | 1.08±0.14  |
|                      | Mean    | 0.56      | 0.54      | 0.28      | 1.63       | 1.09       |

**Table 2**

Comparison of CT segmentations and MR segmentation algorithm ( $n=234$ ). Continuous Dice score ( $d_c$ ) expresses the extent of the spatial similarity. Pearson's correlation ( $r$ ) measures the linear relationship between volumetric measures. (a-c) shows the MRI-MRI, CT-CT, and CT-MRI comparison, respectively.

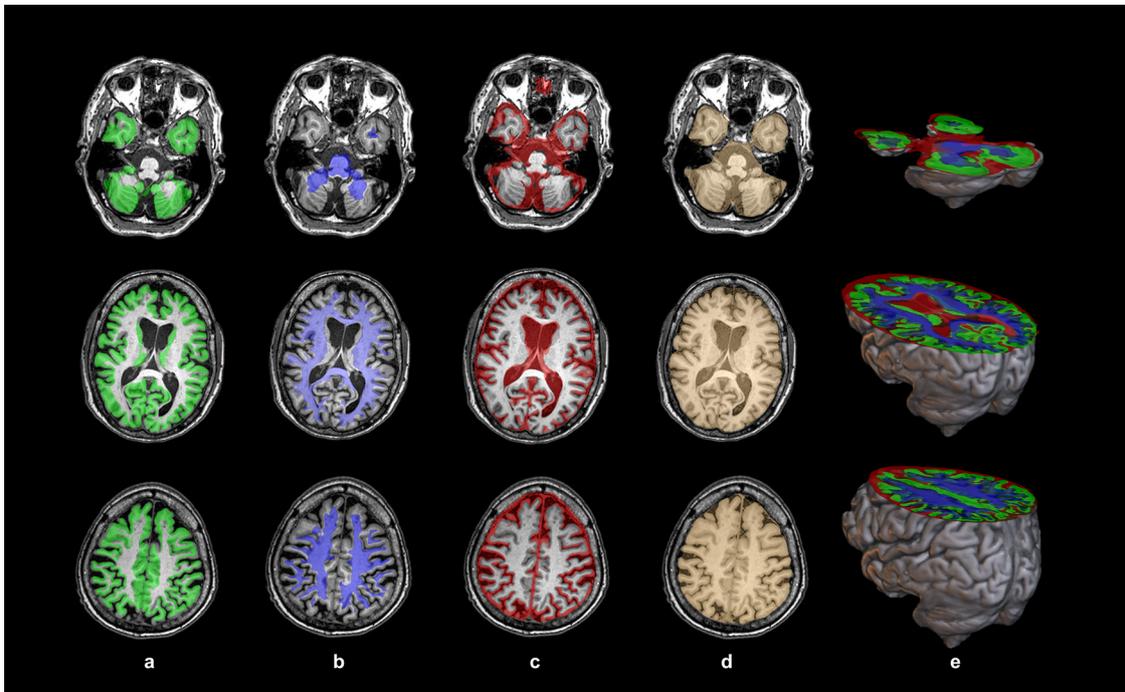
| a. Comparison of SPM, FSL FAST and NiftySeg |                                  |         |                                       |         |                                       |         |
|---|----------------------------------|---------|---------------------------------------|---------|---------------------------------------|---------|
| Tissue class                                | $MRI_{spm} - MRI_{fsl}$<br>$d_c$ | r value | $MRI_{spm} - MRI_{niftyseg}$<br>$d_c$ | r value | $MRI_{fsl} - MRI_{niftyseg}$<br>$d_c$ | r value |
| GM  | 0.88±0.015                       | 0.82    | 0.88±0.018                            | 0.82    | 0.86±0.02                             | 0.94    |
| WM  | 0.89±0.02                        | 0.95    | 0.94±0.01                             | 0.96    | 0.90±0.02                             | 0.98    |
| CSF   | 0.84±0.05                        | 0.78    | 0.84±0.05                             | 0.79    | 0.87±0.08                             | 0.91    |

| b. Comparison of model predictions from U-Nets developed from SPM-derived, FSL FAST-derived and NiftySeg-derived labels |                                |         |                                     |         |                                     |         |
|---|--------------------------------|---------|-------------------------------------|---------|-------------------------------------|---------|
| Tissue class  | $CT_{spm} - CT_{fsl}$<br>$d_c$ | r value | $CT_{spm} - CT_{niftyseg}$<br>$d_c$ | r value | $CT_{fsl} - CT_{niftyseg}$<br>$d_c$ | r value |
| GM  | 0.85±0.009                     | 0.98    | 0.85±0.009                          | 0.98    | 0.85±0.008                          | 0.98    |
| WM  | 0.84±0.013                     | 0.99    | 0.87±0.013                          | 0.99    | 0.85±0.009                          | 0.98    |
| CSF   | 0.76±0.04                      | 0.92    | 0.73±0.04                           | 0.94    | 0.77±0.02                           | 0.98    |

| c. Comparison of model predictions from U-Net developed from SPM-derived labels with MRI segmentations derived using SPM, FSL FAST and NiftySeg |                                 |         |                                 |         |                                      |         |
|---|---------------------------------|---------|---------------------------------|---------|--------------------------------------|---------|
| Tissue class  | $CT_{spm} - MRI_{spm}$<br>$d_c$ | r value | $CT_{spm} - MRI_{fsl}$<br>$d_c$ | r value | $CT_{spm} - MRI_{niftyseg}$<br>$d_c$ | r value |
| GM  | 0.8±0.02                        | 0.93    | 0.79±0.02                       | 0.94    | 0.78±0.02                            | 0.93    |
| WM  | 0.83±0.02                       | 0.96    | 0.85±0.02                       | 0.96    | 0.85±0.015                           | 0.97    |
| CSF   | 0.76±0.06                       | 0.79    | 0.76±0.04                       | 0.83    | 0.76±0.04                            | 0.7     |



**Fig. 4.** Segmentation across various sections. Columns (a-d) depicts GM, WM, CSF and ICV segmentations superimposed on the corresponding section of paired T1 weighted MR image. The last column (e) shows the 3D visualization of derived tissue class segmentations projected on the paired T1 MR image

tissue classification on an MR image that has been degraded through the application of Gaussian filtering with a standard deviation of 1.5 (Fig. 8). The overlap across various modality comparisons was inconsistent across tissue classes, notably CSF. In WM, the average and best results of  $CT_{spm}-MRI_{spm}$  are comparable to the worst results of  $MRI_{spm}-dMRI_{fsl}$  ( $\sigma = 1.5$ ) and  $MRI_{spm}-MRI_{fsl}$ . In CSF, the average results of  $CT_{spm}-MRI_{spm}$  overlapped with the worst results of  $MRI_{spm}-dMRI_{fsl}$ .

#### 4. Discussion

This study shows that deep learning can be applied to assess brain tissue classes quantitatively using only CT images. To this date, this kind of assessment required MR scanning. The significance of the proposed method is that it renders quantitative assessment of neurodegenerative change accessible to many more patients, as CT scans are far more accessible, cheaper, and more rapidly acquired than MR scans.

We evaluated our deep learning model by comparison with criterion standard MR labels obtained with established tissue classification tools using standard measures of label agreement. We found that, with proper optimization and validation, U-Net based deep learning makes CT-based brain tissue segmentation and quantification feasible. The resulting measures have significant potential as diagnostic biomarkers for radiologists and other clinicians.

We compared CT labels to MR labels using three measures (distance, overlap, and volume). The results indicate a strong performance of the deep learning model. High Dice coefficients and strong volumetric correlations show that the predictions were similar in terms of both spatial overlap and density. Across all models, volumetric correlation and overlap measures were highest for ICV, followed by WM and GM and lowest for CSF predictions. This does not imply that the model performed differently on CSF. As the labels differ in shape and surface-to-volume ratio, the same amount of perturbation will have substantially different effects on the overlap measure (Rohlfing et al., 2004). Strong volumetric correlation with low volumetric error observed in GM, WM, and ICV indicates there was a high quantitative similarity between the model-derived volumes with respect to MR images. The low values of AHD and MHD suggest that the predicted segmentations have good bound-

ary agreement with the reference. The high similarity between FHD and RHD shows that the boundary measures have low susceptibility to the directionality of distance measurements. The slight difference of performance within models indicates a minimal existence of training bias. To alleviate training bias, average of segmentation maps derived from each model can be used. Even though all three models' performances showed a good overlap with errors, we compensated for training bias by evaluating the mean metrics across the three models. Taken together, these distinct evaluation metrics indicate that our model performed well.

After assessing the model's performance as a tool for tissue classification or segmentation, we compared SPM-based MR segmentations to other MR image segmentation algorithms. The three algorithms showed strong agreement overall but an inconsistent picture across the tissue classes. Nevertheless, we found no reason to prefer one segmentation algorithm to another for generating training labels. In comparing model predictions from U-Nets trained with SPM-, FSL FAST-, and NiftySeg-derived MR labels, CT segmentation predictions trained with SPM input strongly agree with other predictions. Comparing CT segmentation from models trained with SPM labels also shows strong agreement to its MRI variants. This can be seen as an advantage. The U-Net model is trained to learn boundary patterns and intensity differences and to predict accordingly. Even though the model is provided with MR labels derived from one segmentation algorithm, the model detects the information present in CT and does not introduce label-dependent information outside the input CT image.

However, when we looked into the overlap agreement between  $CT_{spm}-MRI_{spm}$  and  $MRI_{spm}-MRI_{fsl}$ , we observed a large difference in the agreement in all three tissue classes. FSL agrees with the SPM reference significantly more strongly than CT. However, the best CT segmentation results for GM and WM were better than the worst FSL values; for CSF, even average CT-based values are better than bottom-quartile FSL results (Fig. 8). This shows that even though the CT segmentation results are in the same ballpark, we incur a loss of accuracy. This may be unavoidable, considering that CT images contain less information that distinguishes the tissues. To get a better idea of the loss of information, we asked what degradation process would reduce the amount of salient

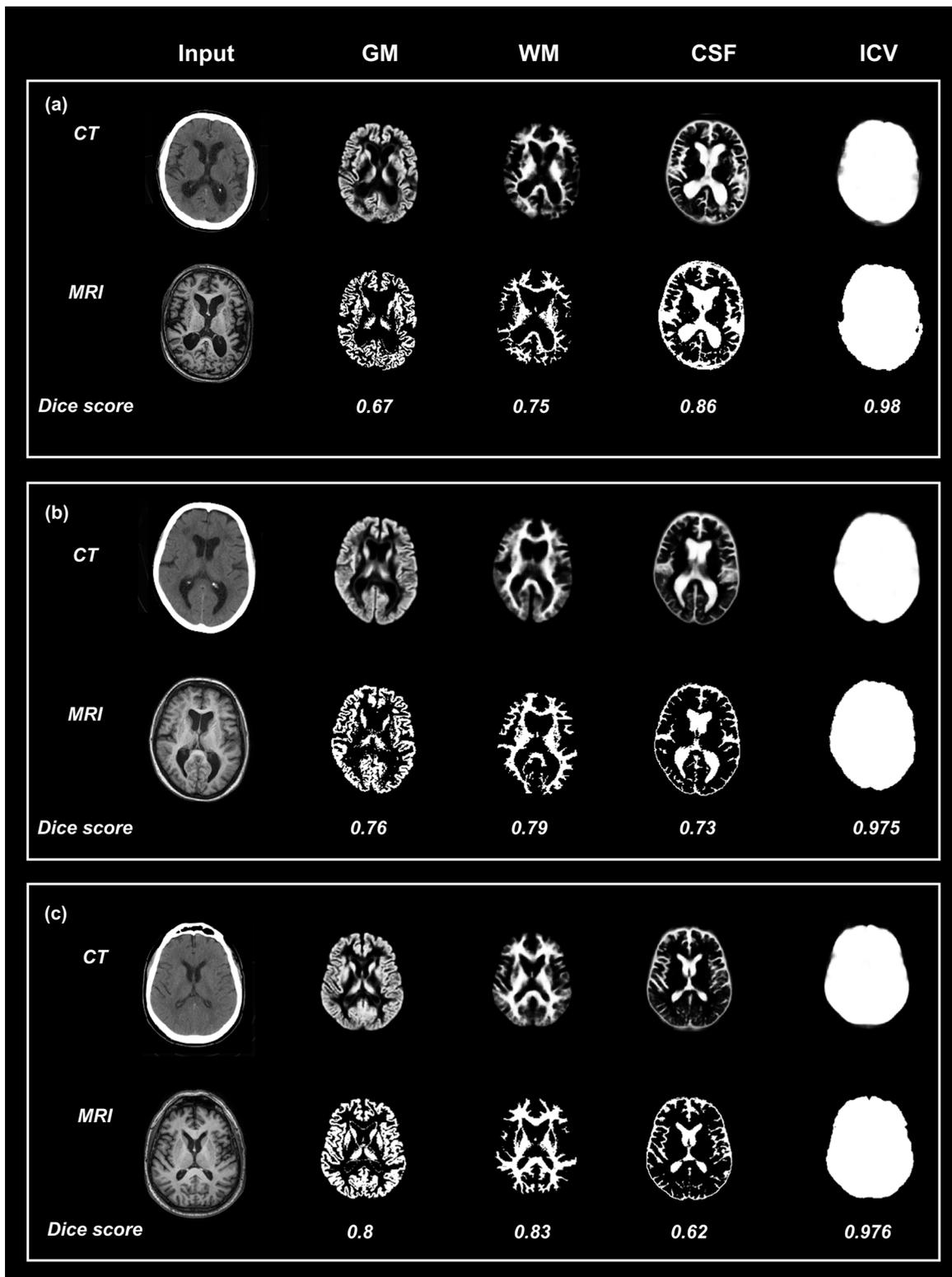
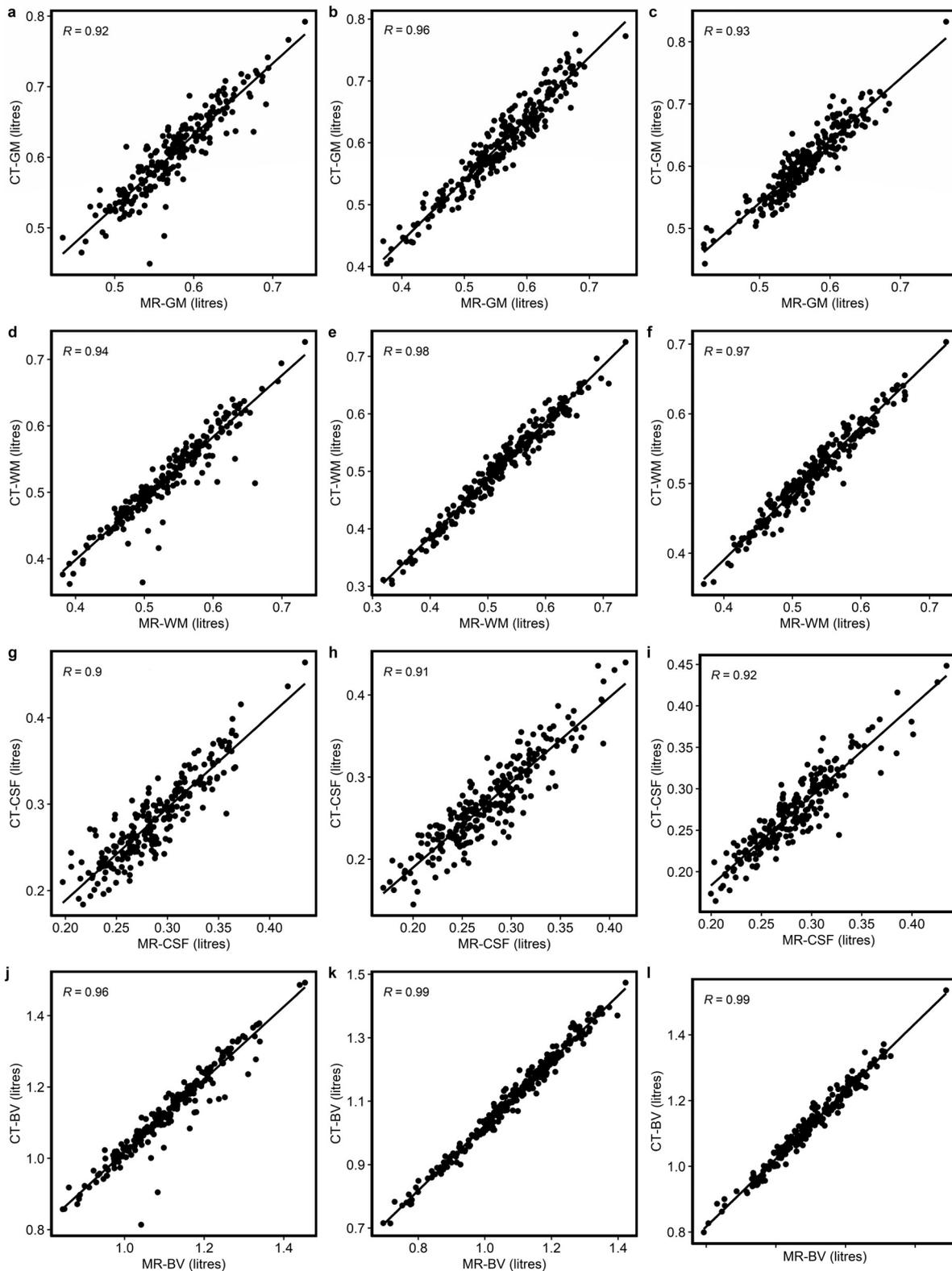


Fig. 5. Input CT images predicted tissue class maps (GM, WM, CSF and ICV) generated with U-Net models from three datasets (a, b, c) in comparison to respective MR labels. The Dice scores between the predicted CT segmentation and paired MR labels are indicated below the maps.

information in an MR image to the level of CT in this context. We found that applying Gaussian blurring with a sigma of 1.5 leads to approximately equivalent tissue classification accuracy. In future work, we will address the relevance of this loss of accuracy for diagnostic accuracy.

Various automated approaches and evaluation methods for CT segmentation have been described previously. Gupta et al., 2010 used

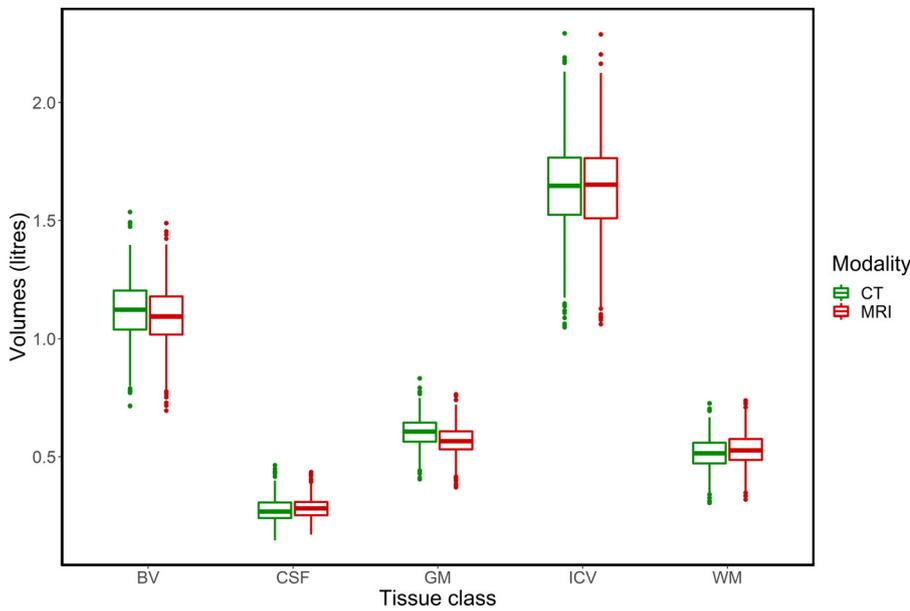
rough estimates of tissue probability based on domain knowledge to improve segmentation by adaptive thresholding and evaluated their method using manual contours on high confidence regions. Kemmling et al., 2012 used MR images to create a probabilistic atlas in standard MNI152 space that was transformed into the CT image space to extract GM, WM, and CSF. This method depended on both tissue speci-



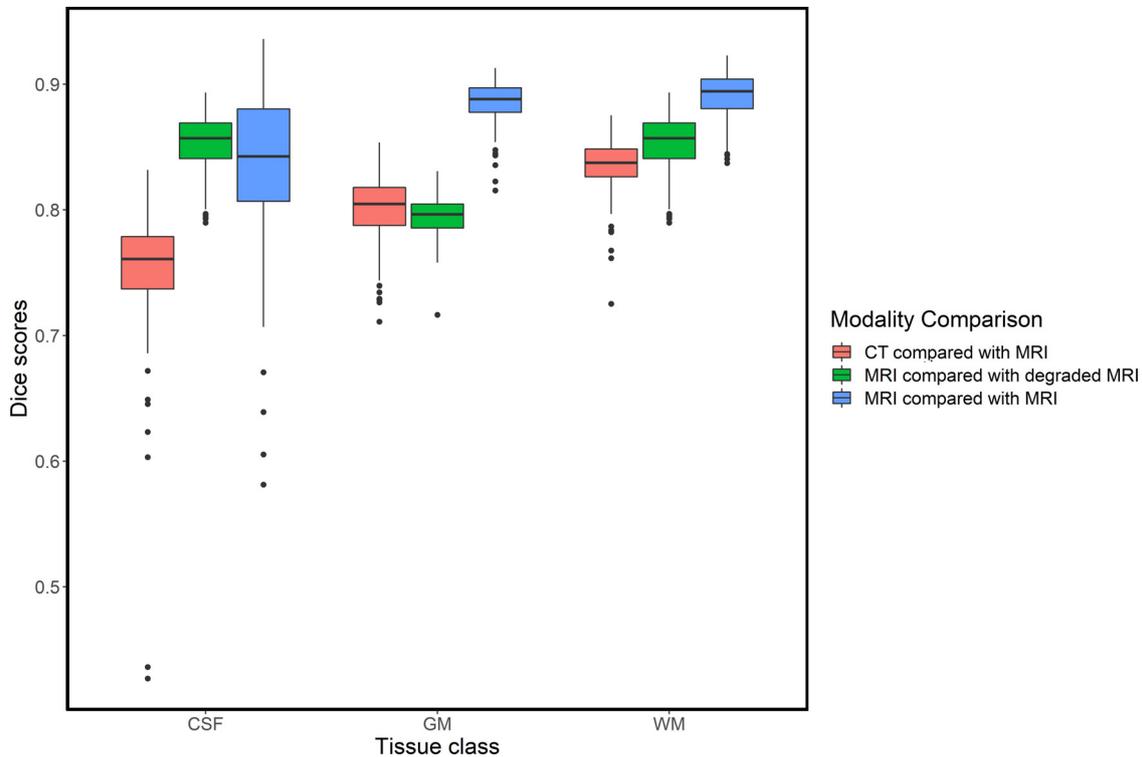
**Fig. 6.** Correlation between model-predicted CT-derived tissue class volumes and MRI-derived label volumes observed in the test datasets of several models ( $n=234$ , a-d: model 1, e-h: model 2 and i-l: model 3). The lines and  $r$ -values indicate the relationship between CT- and MRI-derived volumes.

ficiency of the probabilistic atlas and spatial warping between CT and MR images. Since no quantitative evaluation was carried out in that study, we cannot directly compare the performance of their method with our models. [Cauley et al., 2018](#) focused mainly on the feasibility of direct segmentation of CT using FSL software, which is traditionally employed

to segment MR images. The CT segmentations derived from MR image segmentation softwares lacked sharpness, and hence it is challenging to apply these softwares directly to CT images. Moreover, the results were not evaluated against visual ratings or MR segmentations. Studies on 4D CT brain segmentation have been conducted using feature extrac-



**Fig. 7.** Box plots visualizing individual tissue class volumes. Overall, CT-derived segmentations produced a small number of outliers with respect to paired MR labels. CT-derived WM and CSF volume predictions were generally somewhat underestimated, and GM volume and BV predictions were slightly overestimated.



**Fig. 8.** Distribution of Dice scores between CT-MRI. Continuous Dice coefficients between CT segmentation derived from SPM, MRI segmented using FSL and degraded with Gaussian blur of standard deviation 1.5 in GM, WM, and CSF observed a subset of data (n=234).

tion along with support vector machine-based classifiers and modified U-Nets. Our model outperforms Manniesing et al., 2017 with respect to Dice coefficients and HDs. With temporal features, this method yielded Dice coefficients of 0.81 and 0.79 with HD of 12.65 and 14.85 mm for WM and GM, respectively. Without temporal features, this method achieves Dice coefficients of 0.79 and 0.78 for WM and GM. Our model’s performance is comparable to the one reported by Van De Leemput et al., 2019 in terms of overlap measures in WM and GM. This method used modified U-Nets trained and validated on fewer datasets with manually annotated labels. The networks trained with 4D CT combined spatial information from 3D images with lower noise and temporal features.

Despite having access to less information, our model achieved comparable results.

One of the strengths of our model development is the nature and number of data sets. To our knowledge, most of the methods used in previous studies were applied to a few data sets from specific case studies and used MR images or MR-based atlases to perform brain segmentation. In the present study, we trained models to learn MR labels from CT scans and then segmented tissue classes directly from CT images. Intermediate layers of CNN retain multiple subsets of feature sets like edge, contour, intensity, and uniformity. The model-predicted tissue maps are a result of the evaluation from these collected feature sets.

Another unique feature of our model development is the use of MR labels for training and comparison. One of the considerable limitations of applying deep learning methods to medical image analysis for diagnostic decisions is the lack of well-annotated data sets (Willeminck et al., 2020). Depending on the annotation protocol, annotation labels can be highly operator-dependent. Manual annotations are labour-intensive, rater-dependent, time consuming, and challenging to reproduce. Intra- and inter-operator variability are rarely considered and quantified. In the case of the Gothenburg H70 Birth cohort, we have a large number of paired CT and MRI datasets acquired within short intervals. Despite the cost and the longer scanning time, MRI is a widely used modality for brain tissue class segmentation. MR-derived labels are well-suited for localization and classification of brain tissue. Our model has the advantage that it was trained on MR labels generated with standard research tools in a reproducible manner. These labels are highly consistent and therefore more likely to enable replication of strong performance across different studies than manually drawn training labels. Despite its many limitations, manual annotations are by many considered the gold standard for CT tissue classification. In future work, we plan to compare manual CT annotations to deep learning-derived CT-based model predictions trained from MR-derived labels.

Aside from these strengths, our study also has some limitations. Currently, we train the U-Net with images representative of an elderly population, obtained from a single cohort, with little variance in age across participants. The model can be utilized for CT tissue classification of adult population mainly above 50 years of age. The CT tissue classification of paediatric to young adulthood brain by deep learning models predominantly trained on healthy elderly population will be challenging. Paediatric brain is characterized by dynamic changes in developmental trajectories of cortical, subcortical GM and WM, with white matter largely unmyelinated in normal new-born brain. CSF volumes increase about 2% per year during typical paediatric development (Giedd et al., 2015; Weisenfeld et al., 2006). Moreover, paediatric brain scans have low tissue contrast and exhibit varied signal intensity characteristics in comparison to fully developed brain (Gousias et al., 2013). In future work, we need to assess the usefulness of our model by applying it to other cohorts and age groups. We may then find that to be more generally applicable. The model needs to be trained and tested on a greater variety of data sets collected from populations that are more diverse in terms of age, gender, and acquired on various CT scanners. The applicability of the model may need to be improved by training on cohorts representative of the clinical target populations – for example, on persons referred to a memory clinic.

Although CNNs can be useful image analysis tools, they also present challenges in tracking the features recorded across intermediate layers. Hence, the transparency of the model development is limited, similarly to other models using deep learning.

In terms of deep learning-based architectures, fully CNNs (F-CNNs), U-Nets, residual networks and recurrent neural networks are available for organ/tumour segmentation in MRI. Recently, Zeng and Zheng, 2018 developed context-guided, multi-stream fully convolutional networks trained using T1 and T2 weighted MR images to map MR volumetric data to tissue class labels. Chen et al., 2018 utilized residual networks and summation of feature maps from different layers for brain segmentation. The study also proposed an auto-context residual network that uses multi-modality information from T1-inversion recovery pulse sequence, T2-weighted-fluid-attenuated inversion recovery, and T1 weighted images, which performed better than residual networks, trained using T1 weighted images alone. Roy et al., 2019 developed QuickNAT, which is inspired by an encoder/decoder-based U-Net architecture enhanced with unpooling layers along with dense connections within each encoder/decoder block to perform brain tissue classification on T1-weighted MRIs. This study used Freesurfer-derived labels for pre-training and manual labels for fine-tuning. Henschel et al., 2020 developed FastSurfer, a neuroimaging pipeline that includes a whole-brain MRI segmentation into 95 classes. FastSurfer is an improved version

of QuickNAT that introduces competitive dense blocks and spatial information aggregation. Wu et al., 2019; Zhang et al., 2021 use U-Net and its variants for tissue classification of MR brain images. Due to its encoder/decoder pattern and skip connections, U-Net and its variants perform fast and efficient semantic segmentation, even when the training examples are sparse and varied regarding their source, modality, and nature. Many of the MR-based deep learning algorithms use multi-modal information to train or improve their models for segmentation. The availability of publicly available datasets and standard segmentation algorithms aids testing, validation and comparison of developed models to existing models. In the future, we plan to compare other MR brain segmentation models for brain tissue classification in CT. In particular, we plan to assess the impact of transfer-learning based models with pre-trained models such as VGG as base models, as well as the impact of exchanging loss functions on semantic segmentation of head CT images.

The objective of the current study is to investigate the possibility of automated tissue classification in CT. As a next step, we plan to investigate the clinical usability and diagnostic value of CT-based measures primarily in neurodegenerative diseases and aim to study the correlation of these measures with neuropsychological test results, blood and CSF biomarkers, and expert visual assessments. By combining data sets of patients with neurodegenerative diseases, we intend to create a well-annotated database with variability. By cascading new CNN layers in the architecture and combining measures from other modalities, we aim to create a model for both tissue segmentation and disease indication using CT. Finally, we aim to continue to validate and optimize the model layers, parameters, and predictions to enable patch-based multi-task 3D learning to extract brain atrophy measures from CT.

## 5. Conclusion

In this study, we propose a method to perform semantic segmentation on head CT to distinguish brain tissue classes. The quantitative validation results indicate that the method has clinical potential and warrants further development. Thus, we will validate our approach on various cohorts, including patients with neurodegenerative disorders and adapt it for implementation as a clinical diagnostic support tool.

## Data availability

The H70 cohort cannot openly share data according to the existing ethical and data sharing approvals; however, relevant data can and will be shared with research groups after submitting a research proposal which has to be approved by the study leadership.

## Credit author statement

Meera Srikrishna: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing-Original draft preparation.

Joana B. Pereira: Supervision, Methodology, Writing-Reviewing and Editing.

Rolf A. Heckemann: Supervision, Methodology, Writing-Reviewing and Editing.

Giovanni Volpe: Methodology, Writing-Reviewing and Editing.

Danielle van Westen: Reviewing.

Anna Zettergren: Data generation, Reviewing and Editing

Silke Kern: Data generation, Reviewing and Editing

Lars-Olof Wahlund: Conceptualization, Reviewing

Eric Westman: Conceptualization, Reviewing and Editing

Ingmar Skoog: Funding acquisition, Data generation, Reviewing and Editing

Michael Schöll: Funding acquisition, Conceptualization, Supervision, Project administration, Writing-Reviewing and Editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors would like to thank Dr Alexis Moscoso for his insightful comments to improve this study.

The study was supported by the Knut and Alice Wallenberg Foundation (Wallenberg Centre for Molecular and Translational Medicine; KAW 2014.0363), the Swedish Research Council (#2017-02869), the Swedish state under the agreement between the Swedish government and the County Councils, the ALF-agreement (#ALFGBG-813971), and the Swedish Alzheimer Foundation (#AF-740191). This work used computing resources provided by the Swedish National Infrastructure for Computing (SNIC) at Chalmers Centre for Computational Science and Engineering (C3SE), partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.neuroimage.2021.118606](https://doi.org/10.1016/j.neuroimage.2021.118606).

## References

- Aguilar, C., Edholm, K., Simmons, A., Cavallin, L., Muller, S., Skoog, I., Larsson, E.M., Axelsson, R., Wahlund, L.O., Westman, E., 2015. Automated CT-based segmentation and quantification of total intracranial volume. *Eur. Radiol.* 25, 3151–3160. doi:10.1007/s00330-015-3747-7.
- Akkus, Z., Kostandy, P., Philbrick, K.A., Erickson, B.J., 2019. Robust brain extraction tool for CT head images. *Neurocomputing* doi:10.1016/j.neucom.2018.12.085.
- Ashburner, J., Friston, K.J., 2007. Rigid body registration. *Statistical parametric mapping: The analysis of functional brain images* 49–62.
- Ashburner, J., Friston, K.J., 2005. Unified segmentation. *Neuroimage* 26, 839–851.
- Ashburner, J., Neelin, P., Collins, D., Evans, A., Friston, K., 1997. Incorporating prior knowledge into image registration. *Neuroimage* 6, 344–352.
- Cardoso, M.J., Melbourne, A., Kendall, G.S., Modat, M., Hagmann, C.F., Robertson, N.J., Marlow, N., Ourselin, S., 2011. Adaptive neonate brain segmentation. In: *Presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 378–386.
- Cauley, K., Hu, Y., Fielden, S., 2020. Aging and the brain: a quantitative study of clinical CT images. *Am. J. Neuroradiol.* 41, 809–814.
- Cauley, K.A., Och, J., Yorks, P.J., Fielden, S.W., 2018. Automated Segmentation of Head Computed Tomography Images Using FSL. *J. Comput. Assist. Tomogr.* 42, 104–110. doi:10.1097/RCT.0000000000000660.
- Chen, H., Dou, Q., Yu, L., Qin, J., Heng, P.-A., 2018. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage* 170, 446–455.
- Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., Wang, G., 2017. Low-dose CT via convolutional neural network. *Biomed. Optics Express* 8, 679. doi:10.1364/boe.8.000679.
- Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., Lladó, X., 2019. Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks. *Comput. Biol. Med.* 115. doi:10.1016/j.combiomed.2019.103487.
- Rayment, Dane, Biju, Maya, Zheng, Rui, 2016. Neuroimaging in dementia: an update for the general clinician. *Prog. Neurol. Psychiatry* 20, 16–20. doi:10.1002/pnp.420.
- Despotović, I., Goossens, B., Philips, W., 2015. MRI segmentation of the human brain: challenges, methods, and applications. *Comput. Math. Methods Med.* 2015.
- Driscoll, L., Davatzikos, C., An, Y., Wu, X., Shen, D., Kraut, M., Resnick, S., 2009. Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. *Neurology* 72, 1906–1913.
- Erickson, K.I., Leckie, R.L., Weinstein, A.M., 2014. Physical activity, fitness, and gray matter volume. *Neurobiol. Aging* 35, S20–S28.
- Eurostat, 2021. Medical technology [WWW Document], n.d. URL [https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlt\\_hrs equip&lang=en](https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=hlt_hrs equip&lang=en)
- Fotinos, A.F., Mintun, M.A., Snyder, A.Z., Morris, J.C., Buckner, R.L., 2008. Brain volume decline in aging: evidence for a relation between socioeconomic status, preclinical Alzheimer disease, and reserve. *Arch. Neurol.* 65, 113–120.
- Gao, X.W., Hui, R., Tian, Z., 2017. Classification of CT brain images based on deep learning networks. *Comput. Methods Programs Biomed.* 138, 49–56. doi:10.1016/j.cmpb.2016.10.007.
- Gautam, P., Nuñez, S., Narr, K., Kan, E., Sowell, E., 2014. Effects of prenatal alcohol exposure on the development of white matter volume and change in executive function. *NeuroImage: Clin.* 5, 19–27.
- Gibson, E., Li, W., Sudre, C., Fidon, L., Shaker, D.I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D.C., Ourselin, S., Cardoso, M.J., Vercauteren, T., 2018. NiftyNet: a deep-learning platform for medical imaging. *Comput. Methods Programs Biomed.* 158, 113–122. doi:10.1016/j.cmpb.2018.01.025.
- Giedd, J.N., Raznahan, A., Alexander-Bloch, A., Schmitt, E., Gogtay, N., Rapoport, J.L., 2015. Child psychiatry branch of the National Institute of Mental Health longitudinal structural magnetic resonance imaging study of human brain development. *Neuropsychopharmacology* 40, 43–49.
- Gousias, I.S., Hammers, A., Counsell, S.J., Srinivasan, L., Rutherford, M.A., Heckemann, R.A., Hajnal, J.V., Rueckert, D., Edwards, A.D., 2013. Magnetic resonance imaging of the newborn brain: automatic segmentation of brain images into 50 anatomical regions. *PLoS One* 8, e59990.
- Grieve, S.M., Korgaonkar, M.S., Koslow, S.H., Gordon, E., Williams, L.M., 2013. Widespread reductions in gray matter volume in depression. *NeuroImage: Clin.* 3, 332–339.
- Gupta, V., Ambrosius, W., Qian, G., Blazejewska, A., Kazmierski, R., Urbanik, A., Nowinski, W.L., 2010. Automatic segmentation of cerebrospinal fluid, white and gray matter in unenhanced computed tomography images. *Acad. Radiol.* 17, 1350–1358. doi:10.1016/j.acra.2010.06.005.
- Han, X., 2017. MR-based synthetic CT generation using a deep convolutional neural network method. *Med. Phys.* 44, 1408–1419. doi:10.1002/mp.12155.
- He, L., Yu, H., Shi, L., He, Y., Geng, J., Wei, Y., Sun, H., Chen, Y., 2018. Equity assessment of the distribution of CT and MRI scanners in China: a panel data analysis. *Int. J. Equity Health* 17, 157.
- Heckemann, R.A., Ledig, C., Gray, K.R., Aljabar, P., Rueckert, D., Hajnal, J.V., Hammers, A., 2015. Brain extraction using label propagation and group agreement: Pin-cram. *PLoS One* 10, e0129211.
- Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., Reuter, M., 2020. FastSurfer - a fast and accurate deep learning based neuroimaging pipeline. *Neuroimage* 219, 117012. doi:10.1016/j.neuroimage.2020.117012.
- Hort, J., O'Brien, J.T., Gainotti, G., Pirttila, T., Popescu, B.O., Rektorova, I., Sorbi, S., Scheltens, P., 2010. EFNS guidelines for the diagnosis and management of Alzheimer's disease. *Eur. J. Neurol.* 17, 1236–1248. doi:10.1111/j.1468-1331.2010.03040.x.
- Jacobs, M.A., Ibrahim, T.S., Ouwerkerk, R., 2007. MR imaging: brief overview and emerging applications. *Radiographics* 27, 1213–1229.
- Kemmling, A., Wersching, H., Berger, K., Knecht, S., Groden, C., Nölte, I., 2012. Decomposing the Hounsfield unit: Probabilistic segmentation of brain tissue in computed tomography. *Clin. Neuroradiol.* 22, 79–91. doi:10.1007/s00062-011-0123-0.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., Biller, A., 2016. Deep MRI brain extraction: a 3D convolutional neural network for skull stripping. *Neuroimage* 129, 460–469. doi:10.1016/j.neuroimage.2016.01.024.
- Lee, H., Yune, S., Mansouri, M., Kim, M., Tajmir, S.H., Guerrier, C.E., Ebert, S.A., Pomerantz, S.R., Romero, J.M., Kamalian, S., Lev, M.H., Do, S., 2019. An explainable deep-learning algorithm for the detection of acute intracranial haemorrhage from small datasets. *Nature Biomed. Eng.* 3, 173–182. doi:10.1038/s41551-018-0324-9.
- Liu, F., Jang, H., Kijowski, R., Zhao, G., Bradshaw, T., McMillan, A.B., 2018. A deep learning approach for (18)F-FDG PET attenuation correction. *EJNMMI Phys.* 5, 24. doi:10.1186/s40658-018-0225-8.
- Lu, D., Popuri, K., Ding, G.W., Balachandrar, R., Beg, M.F., Weiner, M., Aisen, P., Petersen, R., Jack, C., Jagust, W., Trojanowski, J., Toga, A., Beckett, L., Green, R., Saykin, A., Morris, J., Shaw, L., Kaye, J., Quinn, J., Silbert, L., Lind, B., Carter, R., Dolen, S., Schneider, L., Pawluczyk, S., Beccera, M., Teodoro, L., Spann, B., Brewer, J., Vandersweg, H., Fleisher, A., Heidebrink, J., Lord, J., Mason, S., Albers, C., Knopman, D., Johnson, Kris, Doody, R., Villanueva-Meyer, J., Chowdhury, M., Rountree, S., Dang, M., Stern, Y., Honig, L., Bell, K., Ances, B., Carroll, M., Creech, M., Franklin, E., Mintun, M., Schneider, S., Oliver, A., Marson, D., Grifith, R., Clark, D., Geldmacher, D., Brockington, J., Roberson, E., Love, M.N., Grossman, H., Mitsis, E., Shah, R., deToledo-Morrell, L., Duara, R., Varon, D., Greig, M., Roberts, P., Albert, M., Onyeko, C., D'Agostino, D., Kielb, S., Galvin, J., Cerbone, B., Michel, C., Pogorelec, D., Rusinek, H., de Leon, M., Glodzik, L., Santi, S., De, Doraiswamy, P., Petrella, J., Borges-Neto, S., Wong, T., Coleman, E., Smith, C., Jicha, G., Hardy, P., Sinha, P., Oates, E., Conrad, G., Porsteinsson, A., Goldstein, B., Martin, K., Makino, K., Ismail, M., Brand, C., Mulnard, R., Thai, G., Mc-Adams-Ortiz, C., Womack, K., Mathews, D., Quiceno, M., Levey, A., Lah, J., Cellar, J., Burns, J., Swerdlow, R., Brooks, W., Apostolova, L., Tingu, K., Woo, E., Silverman, D., Lu, P., Bartzokis, G., Graf-Radford, N., Parfitt, F., Kendall, T., Johnson, H., Farlow, M., Hake, A.M., Matthews, B., Brosch, J., Her-ring, S., Hunt, C., Dyck, C., Carson, R., MacAvoy, M., Varma, P., Chertkow, H., Bergman, H., Hosein, C., Black, S., Stefanovic, B., Caldwell, C., Hsiung, G.Y.R., Feldman, H., Mudge, B., Assaly, M., Finger, E., Pasternack, S., Rachisky, I., Trost, D., Kertesz, A., Bernick, C., Munic, D., Mesulam, M.M., Lipowski, K., Weintraub, S., Bonakdarpour, B., Kerwin, D., Wu, C.K., Johnson, N., Sadowsky, C., Villena, T., Turner, R.S., Johnson, Kathleen, Reynolds, B., Sperling, R., Johnson, Keith, Marshall, G., Yesavage, J., Taylor, J., Lane, B., Rosen, A., Tinklenberg, J., Sabbagh, M., Belden, C., Jacobson, S., Sirrel, S., Kowall, N., Killiany, R., Budson, A., Norbash, A., Johnson, P.L., Obisesan, T., Wolday, S., Allard, J., Lerner, A., Ogrocki, P., Tatsuoka, C., Fatica, P., Fletcher, E., Maillard, P., Olichney, J., DeCarli, C., Carmichael, O., Kittur, S., Borrie, M., Lee, T.Y., Bartha, R., Johnson, S., Asthana, S., Carlsson, C., Potkin, S., Preda, A., Nguyen, D., Tariot, P., Burke, A., Trncic, N., Reeder, S., Bates, V., Capote, H., Ranka, M., Scharre, D., Katakami, M., Adeli, A., Zimmerman, E., Celmins, D., Brown, A., Pearlson, G., Blank, K., Anderson, K., Flashman, L., Seltzer, M., Hynes, M., Santulli, R., Sink, K., Gordiner, L., Williamson, J., Garg, P., Watkins, F., Ott, B., Querfurth, H., Tremont, G., Salloway, S., Malloy, P., Correia, S., Rosen, H., Miller, B., Perry, D., Mintzer, J., Spicer, K., Bachman, D., Pomara, N., Hernando, R., Sarrael, A., Relkin, N.,

- Chaing, G., Lin, M., Ravdin, L., Smith, A., Raj, B.A., Fargher, K., 2018. Multimodal and Multiscale Deep Neural Networks for the Early Diagnosis of Alzheimer's Disease using structural MR and FDG-PET images. *Sci. Rep.* 8. doi:10.1038/s41598-018-22871-z.
- Manniesing, R., Oei, M.T.H., Oostveen, L.J., Melendez, J., Smit, E.J., Platel, B., Sanchez, C.I., Meijer, F.J.A., Prokop, M., van Ginneken, B., 2017. White Matter and Gray Matter Segmentation in 4D Computed Tomography. *Sci. Rep.* 7, 119. doi:10.1038/s41598-017-00239-z.
- Mendrik, A.M., Vincken, K.L., Kuijff, H.J., Breeuwer, M., Bouvy, W.H., de Bresser, J., Alansary, A., de Bruijne, M., Carass, A., El-Baz, A., Jog, A., Katyal, R., Khan, A.R., van der Lijn, F., Mahmood, Q., Mukherjee, R., van Opbroek, A., Paneri, S., Pereira, S., Persson, M., Rajchl, M., Sarikaya, D., Smedby, Ö., Silva, C.A., Vrooman, H.A., Vyas, S., Wang, C., Zhao, L., Biessels, G.J., Vieregger, M.A., 2015. MRBrainS Challenge: Online Evaluation Framework for Brain Image Segmentation in 3T MRI Scans. *Comput. Intell. Neurosci.* 2015, 813696. doi:10.1155/2015/813696.
- Mlynarski, P., Delingette, H., Criminisi, A., Ayache, N., 2019. Deep learning with mixed supervision for brain tumor segmentation. *J. Med. Imaging* 6. doi:10.1117/1.JMI.6.3.034002.
- Musiccio, M., Sorbi, S., Bonavita, V., Caltagirone, C., Aguglia, U., Burzomati, D., Appoloni, I., Isella, V., Bergamasco, B., Ferrero, P., Rota, E., Bergonzi, P., Cancelli, I., Bertolani, L., Bionda, E., Strumia, S., Palazzini, E., Bonavita, Vincenzo, Fragazzi, A., Bono, G., Bonuccelli, U., Cipriani, G., Agostini, S., Sorbi, Sandro, Bracco, L., Bessi, V., Bruni, A.C., Curcio, S., Puccio, G., Caffarra, P., Messa, G., Vezzadini, G., Caltagirone, Carlo, Perri, R., Scalmata, S., Canal, N., Alberoni, M., Mantovani, F., Farina, E., Pinardi, G., Cannata, A., Cappa, S., Marcone, A., Casale, R., Fundarò, C., Comi, G., Magnani, G., Coppola, G., De Leo, M.S., Vasquez, A., Cossa, F., Simone, P., Sinforiani, E., Tabaton, M., Giliberto, L., Tanganelli, P., Partinico, D., Tola, M., Toso, V., Billo, G., Sgarabella, T., Parisen, P., Valluzzi, F., Zaccara, G., Piccininni, M., Zagnoni, P., Peano, M., Carena, G., Tommaso, B., Francesco, P., 2004. Validation of the Guidelines for the Diagnosis of Dementia and Alzheimer's Disease of the Italian Neurological Society. Study in 72 Italian neurological centres and 1549 patients. *Neurol. Sci.* 25, 289–295. doi:10.1007/s10072-004-0356-7.
- Pasi, M., Poggesi, A., Pantoni, L., 2011. The use of CT in dementia. *Int. Psychogeriatr.* 23, S6–S12. doi:10.1017/S1041610211000950.
- Pini, L., Pievani, M., Bocchetta, M., Altomare, D., Bosco, P., Cavedo, E., Galluzzi, S., Marzoni, M., Frisoni, G.B., 2016. Brain atrophy in Alzheimer's disease and aging. *Ageing Res. Rev.* 30, 25–48.
- Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr, C.R., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *Neuroimage* 21, 1428–1442.
- Roy, A.G., Conjeti, S., Navab, N., Wachinger, C., Alzheimer's Disease Neuroimaging Initiative, 2019. QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. *Neuroimage* 186, 713–727.
- Rydberg Sterner, T., Ahlner, F., Blennow, K., Dahlin-Ivanoff, S., Falk, H., Havstam Johansson, L., Hoff, M., Holm, M., Hörder, H., Jacobsson, T., Johansson, B., Johansson, L., Kern, J., Kern, S., Machado, A., Mellqvist Fässberg, M., Nilsson, J., Ribbe, M., Rothenberg, E., Rydén, L., Sadeghi, A., Sacuiu, S., Samuelsson, J., Sigström, R., Skoog, J., Thorvaldsson, V., Waern, M., Westman, E., Wetterberg, H., Zetterberg, H., Zetterberg, M., Zettergren, A., Östling, S., Skoog, I., 2019. The Gothenburg H70 Birth cohort study 2014–16: design, methods and study population. *Eur. J. Epidemiol.* 34, 191–209. doi:10.1007/s10654-018-0459-8.
- Sacuiu, S., Eckerström, M., Johansson, L., Kern, S., Sigström, R., Xinxin, G., Östling, S., Skoog, I., 2018. Increased risk of dementia in subjective cognitive decline if CT brain changes are present. *J. Alzheimer's Dis.* 66, 483–495.
- Sandor, T., Metcalf, D., Kim, Y.-J., 1991. Segmentation of brain CT images using the concept of region growing. *Int. J. Biomed. Comput.* 29, 133–147. doi:10.1016/0020-7101(91)90004-X.
- Shamir, R.R., Duchin, Y., Kim, J., Sapir, G., Harel, N., 2019. Continuous dice coefficient: a method for evaluating probabilistic segmentations. arXiv preprint arXiv:1906.11031.
- Stewart, R., 2001. Neuroimaging in dementia and depression. *Current Opin. Psychiatry* 14, 371–375.
- Suzuki, K., 2017. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* 10, 257–273. doi:10.1007/s12194-017-0406-5.
- Thiagarajan, S., Shaik, M.A., Venketasubramanian, N., Ting, E., Hilal, S., Chen, C., 2018. Coronal CT is comparable to MR imaging in aiding diagnosis of dementia in a memory clinic in Singapore. *Alzheimer Dis. Assoc. Disord.* 32, 94–100.
- Van De Leemput, S.C., Meijs, M., Patel, A., Meijer, F.J.A., Van Ginneken, B., Manniesing, R., 2019. Multiclass brain tissue segmentation in 4D CT using convolutional neural networks. *IEEE Access* 7, 51557–51569. doi:10.1109/ACCESS.2019.2910348.
- Wattjes, M.P., Henneman, W.J., van der Flier, W.M., de Vries, O., Träber, F., Geurts, J.J., Scheltens, P., Vrenken, H., Barkhof, F., 2009. Diagnostic imaging of patients in a memory clinic: comparison of MR imaging and 64–detector row CT. *Radiology* 253, 174–183.
- Weisenfeld, N.I., Mewes, A.U., Warfield, S.K., 2006. Highly accurate segmentation of brain tissue and subcortical gray matter from newborn MRI. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 199–206.
- Westbrook, C., Talbot, J., 2018. *MRI in Practice*. John Wiley & Sons.
- Willeminck, M.J., Koszek, W.A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L.R., Summers, R.M., Rubin, D.L., Lungren, M.P., 2020. Preparing medical imaging data for machine learning. *Radiology* 295, 4–15.
- Wu, J., Zhang, Y., Wang, K., Tang, X., 2019. Skip Connection U-Net for White Matter Hyperintensities Segmentation From MRI. *IEEE Access* 7, 155194–155202. doi:10.1109/ACCESS.2019.2948476.
- Zeng, G., Zheng, G., 2018. Multi-stream 3D FCN with multi-scale deep supervision for multi-modality isointense infant brain MR image segmentation. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. Presented at the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 136–140. doi:10.1109/ISBI.2018.8363540.
- Zhang, F., Breger, A., Cho, K.I.K., Ning, L., Westin, C.-F., O'Donnell, L.J., Pasternak, O., 2021. Deep learning based segmentation of brain tissue from diffusion MRI. *Neuroimage* 233, 117934. doi:10.1016/j.neuroimage.2021.117934.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* 20, 45–57.