Comparison of Two-Dimensional- and Three-Dimensional-Based U-Net Architectures for Brain Tissue Classification in One-Dimensional Brain CT

Meera Srikrishna[1,2], Rolf A. Heckemann[3], Joana B. Pereira[4,5], Giovanni Volpe[6], Anna Zettergren[7], Silke Kern[7,8], Eric Westman[4], Ingmar Skoog[7,8] and Michael Schöll[1,2,9,10]*

[1] Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden, [2] Department of Psychiatry and Neurochemistry, Institute of Physiology and Neuroscience, University of Gothenburg, Gothenburg, Sweden, [3] Department of Medical Radiation Sciences, Institute of Clinical Sciences, Sahlgrenska Academy, Gothenburg, Sweden, [4] Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institutet, Stockholm, Sweden, [5] Memory Research Unit, Department of Clinical Sciences, Malmö Lund University, Mälmo, Sweden, [6] Department of Physics, University of Gothenburg, Gothenburg, Sweden, [7] Neuropsychiatric Epidemiology, Institute of Neuroscience and Physiology, Sahlgrenska Academy, Centre for Ageing and Health (AgeCap), University of Gothenburg, Gothenburg, Sweden, [8] Region Västra Götaland, Sahlgrenska University Hospital, Psychiatry, Cognition and Old Age Psychiatry Clinic, Gothenburg, Sweden, [9] Dementia Research Centre, Institute of Neurology, University College London, London, United Kingdom, [10] Department of Clinical Physiology, Sahlgrenska University Hospital, Gothenburg, Sweden

Brain tissue segmentation plays a crucial role in feature extraction, volumetric quantification, and morphometric analysis of brain scans. For the assessment of brain structure and integrity, CT is a non-invasive, cheaper, faster, and more widely available modality than MRI. However, the clinical application of CT is mostly limited to the visual assessment of brain integrity and exclusion of copathologies. We have previously developed two-dimensional (2D) deep learning-based segmentation networks that successfully classified brain tissue in head CT. Recently, deep learning-based MRI segmentation models successfully use patch-based three-dimensional (3D) segmentation networks. In this study, we aimed to develop patch-based 3D segmentation networks for CT brain tissue classification. Furthermore, we aimed to compare the performance of 2D- and 3D-based segmentation networks to perform brain tissue classification in anisotropic CT scans. For this purpose, we developed 2D and 3D U-Net-based deep learning models that were trained and validated on MR-derived segmentations from scans of 744 participants of the Gothenburg H70 Cohort with both CT and T1-weighted MRI scans acquired timely close to each other. Segmentation performance of both 2D and 3D models was evaluated on 234 unseen datasets using measures of distance, spatial similarity, and tissue volume. Single-task slice-wise processed 2D U-Nets performed better than multitask patch-based 3D U-Nets in CT brain tissue classification. These findings provide support to the use of 2D U-Nets to segment brain tissue in one-dimensional (1D) CT. This could increase the application of CT to detect brain abnormalities in clinical settings.

Keywords: brain image segmentation, CT, MRI, deep learning, convolutional neural networks

INTRODUCTION

X-ray CT and MRI are the most frequently used modalities for structural assessment in neurodegenerative disorders (Wattjes et al., 2009; Pasi et al., 2011). MRI scans are commonly used for image-based tissue classification to quantify and extract atrophy-related measures from structural neuroimaging modalities (Despotović et al., 2015). Many software tools exist to perform automated brain segmentation in MR images, mainly for research purposes (Zhang et al., 2001; Ashburner and Friston, 2005; Cardoso et al., 2011; Fischl, 2012). Currently, CT scanning is used for the visual assessment of brain integrity and the exclusion of copathologies in neurodegenerative diseases (Musicco et al., 2004; Rayment et al., 2016). However, several studies suggest that visual assessment of brain volume changes derived from CT could also be used as the predictors of dementia, displaying comparable diagnostic properties to visual ratings of MRI scans (Sacuiu et al., 2018; Thiagarajan et al., 2018). In comparison with MR imaging, CT scanning is faster, cheaper, and more widely available. Despite these advantages, automated tissue classification in head CT is largely underexplored.

Brain tissue segmentation on CT is a challenging task due to lower soft-tissue contrast compared to MRI. Many existing CT data and some scanners collect anisotropic CT data or one-dimensional (1D) CT images. Several studies have recommended the quantification of brain tissue classes in CT using MR segmentation methods (Manniesing et al., 2017; Cauley et al., 2018), general image thresholding (Gupta et al., 2010) or segmentation methods (Aguilar et al., 2015), or probabilistic classification using Hounsfield units (HU) (Kemmling et al., 2012).

More recent studies have also started exploring the usage of deep learning in brain image segmentation (Akkus et al., 2017; Chen et al., 2018; Henschel et al., 2020; Zhang et al., 2021) mainly using MR images. For this purpose, fully convolutional neural networks (CNNs), residual networks, U-Nets, and recurrent neural networks are commonly used architectures. U-Net, developed in 2015, is a CNN-based architecture for biomedical image segmentation that performs segmentation by classifying at every pixel/voxel (Ronneberger et al., 2015). Many studies have used U-Nets to perform semantic segmentation in MRI (Wang et al., 2019; Wu et al., 2019; Brusini et al., 2020; Henschel et al., 2020; Zhang et al., 2021) and few in CT (Van De Leemput et al., 2019; Akkus et al., 2020). An influential and challenging aspect of deep learning-based studies, especially segmentation-based tasks is the selection of data and labels used for training (Willemink et al., 2020).

The two-dimensional (2D)-based deep learning models use 2D functions to train and predict segmentation maps for a single slice. To predict segmentation maps for full volumes, 2D models take predictions one slice at a time. The 2D model functions can capture context across the height and width of the slice to train and make predictions. The three-dimensional (3D)-based deep learning architectures can capture interslice context. However, this comes at a computational cost due to the increased number of parameters used by the models. To accommodate computational needs, patches are processed instead of whole 3D

volumes (Alom et al., 2019). The 3D patch-based processing is a useful method for processing large 3D volumes in 3D image classification and segmentation using deep learning, especially in MR image processing for deep learning (Baid et al., 2018; Largent et al., 2019).

Previously, we conducted a study exploring the possibility of using MR-derived brain tissue class labels to train deep learning models to perform brain tissue classification in head CTs (Srikrishna et al., 2021). We showed that 2D U-Nets could be successfully trained to perform automated segmentation of gray matter (GM), white matter (WM), cerebrospinal fluid (CSF), and intracranial volume in head CT. In this study, we planned to explore if incorporating interslice information using 3D deep learning models could improve CT brain segmentation performance. Patch-based 3D segmentation networks have been successfully used to develop 3D deep learning-based MRI segmentation models. In this study, we aimed to develop 3D U-Nets for the tissue classification of anisotropic head CT, with larger slice thickness (∼5 mm), and compared its CT brain tissue classification performance with 2D U-Nets. We trained both models using MR-derived segmentation labels and assessed the accuracy of the model by comparing CT segmentation results with MR segmentation results.
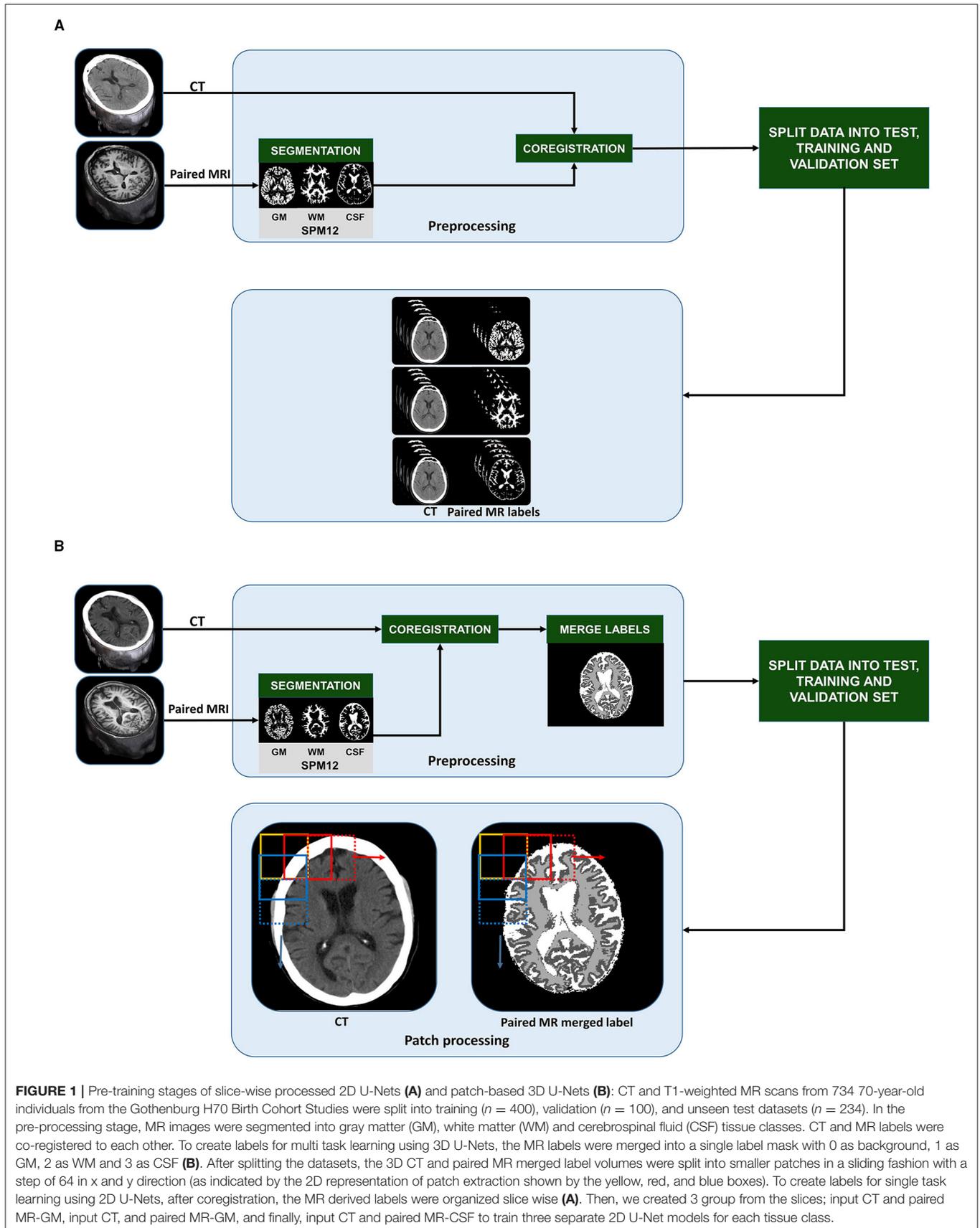
MATERIALS AND METHODS

Datasets

We derived 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 years to study the elderly population of Gothenburg in Sweden. For this study, we included same-day acquisitions of CT and MR images from 744 participants (52.6% female, mean age 70.44 ± 2.6 years) of the cohort born in 1944, collected from 2014 to 2016. The full study details are reported elsewhere (Rydberg Sterner et al., 2019). 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 × 512, and a voxel size of 0.5 × 0.5 × 5.0 mm (Philips Medical Systems, Best, Netherlands). MRI 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 × 256 × 160 voxels, voxel size: 1 × 1 × 1 mm, echo time: 3.2 ms, repetition time: 7.2 ms, and flip angle: 9° (Rydberg Sterner et al., 2019).

Model Development and Training
Image Preprocessing
We preprocessed all paired CT and MR images using SPM12 (http://www.fil.ion.ucl.ac.uk/spm), running on MATLAB 2020a (**Figure 1**), after first converting CT and MR images to NIfTI format and visually assessing the quality and integrity of all the scans. After that, each image was aligned to the anterior commissure- posterior commissure line.

We segmented MR images into GM, WM, and CSF labels using the unified segmentation algorithm in SPM12 (Ashburner and Friston, 2005). MR labels were used as training inputs to

FIGURE 1 | Pre-training stages of slice-wise processed 2D U-Nets **(A)** and patch-based 3D U-Nets **(B)**: CT and T1-weighted MR scans from 734 70-year-old individuals from the Gothenburg H70 Birth Cohort Studies were split into training ($n = 400$), validation ($n = 100$), and unseen test datasets ($n = 234$). In the pre-processing stage, MR images were segmented into gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) tissue classes. CT and MR labels were co-registered to each other. To create labels for multi task learning using 3D U-Nets, the MR labels were merged into a single label mask with 0 as background, 1 as GM, 2 as WM and 3 as CSF **(B)**. After splitting the datasets, the 3D CT and paired MR merged label volumes were split into smaller patches in a sliding fashion with a step of 64 in x and y direction (as indicated by the 2D representation of patch extraction shown by the yellow, red, and blue boxes). To create labels for single task learning using 2D U-Nets, after coregistration, the MR derived labels were organized slice wise **(A)**. Then, we created 3 group from the slices; input CT and paired MR-GM, input CT, and paired MR-GM, and finally, input CT and paired MR-CSF to train three separate 2D U-Net models for each tissue class.

develop our models. To represent the CT images and MR labels in a common image matrix, MR images were coregistered to their paired CT images using SPM12 (Ashburner and Friston, 2007). One of the most decisive steps in our study is the coregistration between MRI and CT. A successful coregistration of MRI labels to CT scans enables MRI-derived labels to be used

as training inputs. The difference in resolution affects the scaling and selection of starting points for the cost functions. SPM12 coregistration module optimizes the rigid transformation using a cost function, in our case, the normalized mutual information (Ashburner and Friston, 2007). Furthermore, the coregistration of each CT-MR pair was visually assessed, and 10 pairs were
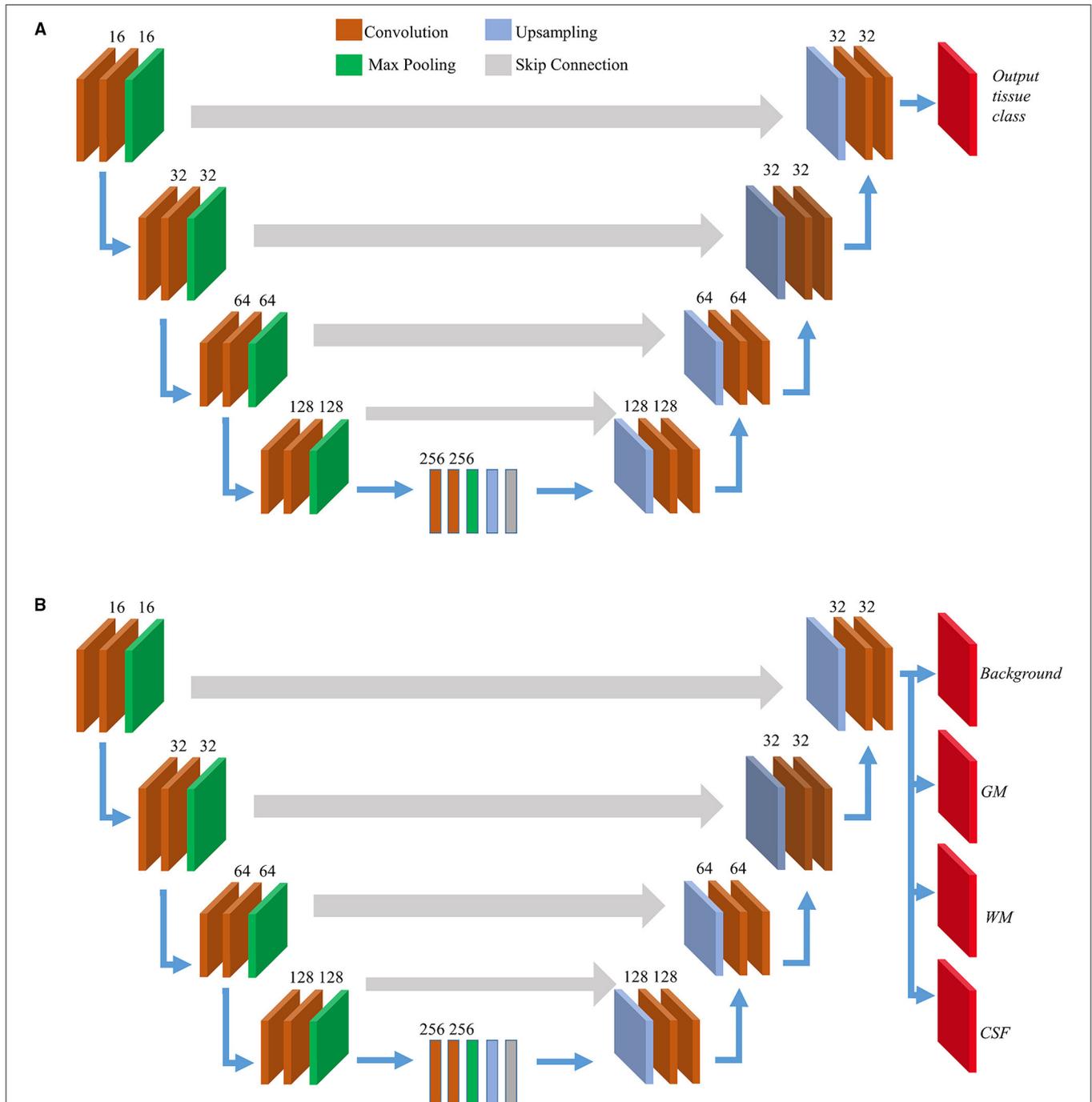
FIGURE 2 | Model architectures. Overview of internal layers in 2D (A) and 3D U-Net (B) utilized to perform brain segmentation. A two or three-dimensional version of internal layers is used depending on slice-wise or patch-wise inputs. The output layer(s) depends on single or multi-task learning. For 2D U-Nets we used single-task learning, hence there was a single output layer. For 3D U-Nets, three tissue classes were trained at a time along with background hence there were four output layers.

excluded based on faulty coregistration due to rotational or translational misregistration.

We developed 2D U-Nets to perform single-task learning and 3D U-Nets to perform multitask learning. To enable this, for 2D U-Net training, we paired each CT image with its corresponding MR-based GM, WM, and CSF labels (**Figure 1A**). For the 3D U-Net training, all the labels were merged into a single image and assigned the values 0 to the background, 1 to GM intensities, 2 to WM intensities, and 3 to CSF intensities (**Figure 1B**).

Data Preparation

We split the 734 processed datasets consisting of CT images and their paired, coregistered MR-label images into 500 training and 234 test datasets. We implemented a three-fold cross-validation in the training routine. The model hyperparameters were trained using the training datasets and fine-tuned with the validation dataset after each cycle or epoch and fold. The overall performance of the model was analyzed on the unseen test datasets. We used the same training and test dataset IDs for 2D and 3D U-Nets. Prior to the training routine, we thresholded all images within a 0–100 HU range, which is the recommended HU brain window. We resized the images to ensure that the length and width of the images remained 512, to accommodate the input size requirements of the U-Nets.

3D U-Net Structure and Training

We performed all the patch-based processing steps (**Figure 1A**) in Python 3.7. The MR-merged labels were converted into categorical data to enable multitask learning. The 3D $512 \times 512 \times 32$ sized CT images and categorical MR-merged labels were split into smaller 3D patches of size $128 \times 128 \times 32$ by sliding through various layers in the image. For each paired image/label, we obtained 49 pairs of patches, thus obtaining 24,500 patches in the training routine.

We trained the 3D U-Nets using an Nvidia GeForce RTX 2080 Ti graphical processing unit (GPU), 11 GB of random access memory (Nvidia Corp., Santa Clara, CA, USA). The model architectures were developed and trained in Python 3.7, using TensorFlow 2.0 and Keras 2.3.1. The 3D U-Nets were developed to accept 3D patches of size $128 \times 128 \times 32$.

The architecture of the 3D U-Net-based network is shown in **Figure 2B**. We resized all images to $512 \times 512 \times 32$ to ensure that the spatial dimension of the output is the same as the input and to reduce the number of blank or empty patches. The U-Net architecture consisted of a contracting path or encoder path to capture context and a symmetric expanding path or decoder path that enables localization. In the first layer, the CT data were provided as input for training along with the corresponding ground truth, which in our case were the MR-merged labels. For 3D U-Nets, we programmed the internal functional layers using 3D functions. Each encoding block consists of two 3D convolutional layers with a kernel size 3 and a rectified linear unit (ReLU) activation function (Agarap, 2018), followed by 3D max-pooling and batch normalization layers. The number of feature maps increased in the subsequent layers to learn the structural

features of various tissue classes. Dropout layers were added with a ratio of 0.2 to the last two encoding blocks.

We used symmetric decoding blocks with skip connections from corresponding encoding blocks and concatenated features to the deconvolution outputs. The final output layers were expanded to accommodate the multiclass nature of the ground truth labels. Categorical cross-entropy was used as the loss function, and a uniform loss was applied across all classification labels. We balanced the number of encoding/decoding blocks and patch size to accommodate the GPU memory constraints. Finally, four output maps with a 3D convolutional layer and a softmax activation (Dunne and Campbell, 1997) corresponding to the background, GM, WM, and CSF were generated. For this study, the learning rate was initialized to 0.0001. The adaptive moment estimation (Adam) optimizer (Kingma and Ba, 2014) was used, and all weights were initialized using a normal distribution with a mean of 0, an SD of 0.01, and biases as 0. After every epoch of the training routine, the predicted probability map derived from the validation is compared with its corresponding MR-merged label. Each voxel of the output segmentation maps corresponds to the probability of that voxel belonging to a particular tissue class. It calculates the error using the loss function, and it is backpropagated to the training parameters of the 3D U-Net layers. The learning saturated after 40 epochs and after this early stopping was implemented. The training time was approximately 25 h.

2D U-Net Structure and Training

We developed and trained the 2D U-Net models using the procedures described in our previous study (Srikrishna et al., 2021). The architecture of the 2D U-Net-based network is shown in **Figure 2A**. We derived three models for GM, WM, and CSF tissue classification, respectively, using 2D U-Nets. In brief, we trained 2D U-Net-based deep learning models to differentiate between various tissue classes in the head CT. The segmentation patterns were studied from their paired MR labels. The 2D U-Net-based deep learning models were created and trained to accept CT scans and MR labels as input. The 2D U-Net follows the same structure as shown in **Figure 2** with the internal functional layers programmed using 2D functions. The 2D U-Net was designed to perform the slice-wise processing of input data where each input image was processed as a stack of 2D slices with a size of 512×512 pixels. In total, 12,000 training slices and 3,000 validation slices were used to train 1,177,649 trainable hyperparameters. The inputs were fed into the model, and learning was executed using the Keras module. The batch size was 16. The callback features were used such as early stopping and automatic reduction of learning rate with respect to the rate of training. The model was trained for 50 epochs with 750 samples per epoch in approximately 540 min. We trained the 2D U-Nets also on an Nvidia GeForce RTX 2080 Ti GPU using TensorFlow 2.0 and Keras 2.3.1.

Quantitative Performance Assessment and Comparison

For the quantitative assessment of the predictions from 3D U-Nets and 2D U-Nets trained on MR labels on the anisotropic
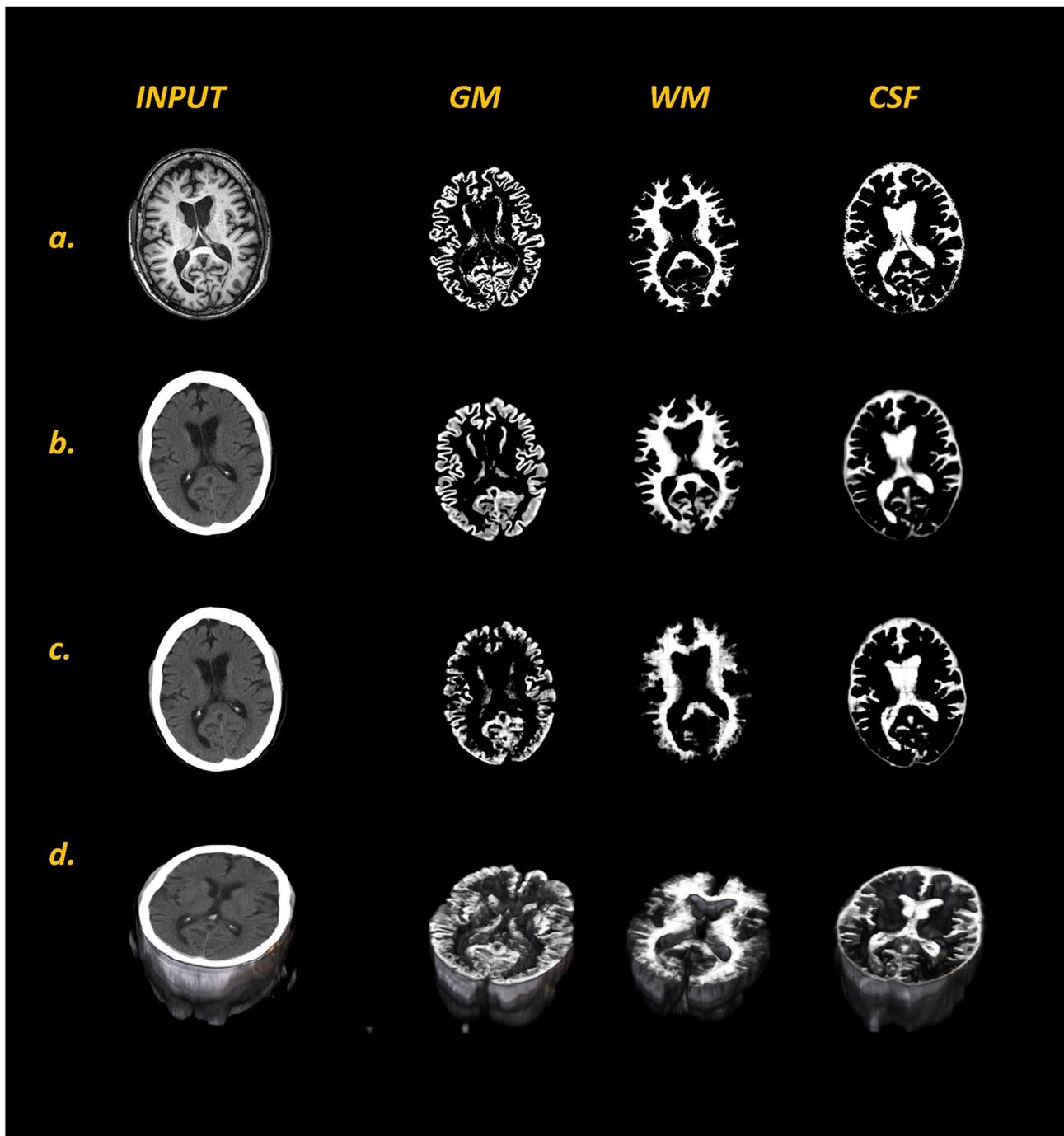
FIGURE 3 | Comparison of MR labels **(a)** with respective input CT images and predicted tissue class maps (GM, WM, and CSF) generated with 2D U-Net **(b)** and 3D U-Net **(c)** models from a representative dataset. Panel **(d)** shows the 3D visualization of input CT and 3D U-Net predicted tissue class maps. In comparison to 2D U-Nets, 3D U-Nets was not able to resolve the finer details of all three tissue classes, especially WM.

CT data, we derived segmentation maps from both the deep learning models on the test datasets ($n = 234$) in Python 3.7, using TensorFlow 2.0 and Keras 2.3.1. The segmentations from test CTs were derived using deep learning models without the intervention of MRI and any preprocessing steps. Both models took <1 min to acquire prediction tissue class maps for one dataset; however, 2D U-Nets were 45 s faster than 3D.

For the comparative study, we used the same test datasets for both the 2D U-Nets and 3D U-Nets. We compared the predictions to their corresponding MR labels, which we

TABLE 1 | Quantitative performance metrics in test datasets (n = 234).

Assessment metrics	Model	GM	WM	CSF
d_c	2D U-Nets	0.80 ± 0.03	0.83 ± 0.02	0.76 ± 0.06
	3D U-Nets	0.59 ± 0.02	0.65 ± 0.03	0.55 ± 0.05
r	2D U-Nets	0.92	0.94	0.9
	3D U-Nets	0.85	0.73	0.73
VE	2D U-Nets	0.07 ± 0.03	0.03 ± 0.03	0.07 ± 0.06
	3D U-Nets	0.09 ± 0.08	0.04 ± 0.12	0.11 ± 0.13
AHD (mm)	2D U-Nets	4.60 ± 1.6	4.10 ± 1.8	4.43 ± 1.5
	3D U-Nets	7.03 ± 2.6	13.49 ± 6.1	8.34 ± 7.2
MHD (mm)	2D U-Nets	1.67 ± 1	1.19 ± 0.8	1.42 ± 0.6
	3D U-Nets	2.59 ± 1.4	4.29 ± 2.9	4.65 ± 3.9
V_{CT} (liters)	2D U-Nets	0.55 ± 0.07	0.52 ± 0.08	0.27 ± 0.05
	3D U-Nets	0.81 ± 0.11	0.64 ± 0.12	0.44 ± 0.09

Continuous Dice score (d_c) expresses the extent of spatial overlap between CT predictions from 2D U-Nets/3D U-Nets and MR labels, Pearson's correlation coefficient (r) measures the linear relationship between volumetric measures, AHD and MHD are distance measures, and volumetric error (VE) expresses the absolute volumetric difference between CT segmentation maps and MR labels. The volumes of CT predictions are given by V_{CT}, respectively.

employed as the standard or reference criterion. We adopted the approach suggested by the MRbrainS challenge (Mendrik et al., 2015), for the comparison of similarity between the predicted masks and standard criterion. We assessed the predictions using four measures, namely, continuous Dice coefficient (d_c), Pearson's correlation of volumetric measures (r), Hausdorff distance (HD), and volumetric error (VE). The prediction maps from 2D U-Nets were obtained slice-wise. We stacked the slices to obtain 3D prediction maps for further comparison. We used the 3D version of all measures for the comparison of both 2D U-Net prediction maps and 3D U-Net predictions maps with their respective MR-derived labels.

To assess the spatial similarity between predicted probability maps and binary MR labels, we used the continuous Dice score, a variant of the Dice coefficient (Shamir et al., 2019). For distance similarity assessment, we used the average HD (AHD) and modified HD (MHD) (Dubuisson and Jain, 1994). We assessed the volumetric similarity using Pearson's correlation coefficient (Benesty et al., 2009) and VE. For this purpose, we binarized the prediction maps derived from both models. As the 3D U-Nets were trained using MR-merged labels, we used global thresholding with a threshold of 0.5 to binarize the predictions. For the prediction maps derived from 2D U-Nets, we used a data-driven approach for binarization. For image I, we derived segmentation maps I_{GM}, I_{WM}, and I_{CSF} using the 2D slice-wise predictions and stacked them to derive a 3D image. At each voxel (x,y,z), we compared the intensities of $I_{GM}(x,y,z)$, $I_{WM}(x,y,z)$, and $I_{CSF}(x,y,z)$. We assigned the voxel (x,y,z) to the tissue class with maximum intensity at that voxel. This method ensures that there are no overlapping pixels among various tissue classes. We measured VE by deriving the absolute volumetric difference divided by the sum of the compared volumes.

RESULTS

The representative predicted segmentations of different tissue classes derived from test CTs by 2D U-Nets and 3D U-Nets

are shown in **Figure 3**. **Table 1** presents the resulting metrics obtained from our comparative analysis. We compared the predictions from 2D U-Nets and 3D U-Nets separately with MR-derived labels.

In comparison with 3D U-Nets, 2D U-Nets performed better in the segmentation of all three tissue classes. The 2D U-Nets yielded d_c of 0.8, 0.83, and 0.76 in GM, WM, and CSF, respectively. The boundary measures expressed by AHD, MHD (in mm) were found to be (4.6, 1.67), (4.1, 1.19), and (4.43, 1.42) for GM, WM, and CSF, respectively. Pearson's coefficient for volumetric measures was observed to be 0.92, 0.94, and 0.9 for GM, WM, and CSF, respectively, with a VE of 0.07, 0.03, and 0.07. In the case of 3D U-Nets, we yielded d_c of 0.59, 0.65, and 0.55 in GM, WM, and CSF, respectively. The boundary measures expressed by AHD, MHD (in mm) were found to be (7.03, 2.59), (8.49, 4.29), and (8.34, 4.65) for GM, WM, and CSF, respectively. Pearson's coefficient for volumetric measures was observed to be 0.85, 0.73, and 0.73 for GM, WM, and CSF, respectively, with a VE of 0.09, 0.04, and 0.11. **Figure 4** shows the comparative performance of Dice scores and volumes between 2D U-Nets and 3D U-Nets. Volume predictions from 3D U-Nets were overestimated in comparison with predictions from 2D U-Nets in all tissue classes.

DISCUSSION

Our previous study showed that 2D deep learning-based algorithms can be used to quantify brain tissue classes using only CT images (Srikrishna et al., 2021). In this study, we explored the possibility of using 3D-based deep learning algorithms for CT brain tissue classification and compared the performance of 2D- and 3D-based deep learning networks to perform brain tissue classification in CT scans. Our CT images are anisotropic, with low resolution (5 mm) along the z-axis. This study shows that 2D segmentation networks are appropriate to deal with datasets of this nature. This is the only study that compares 2D and 3D segmentation networks for 1D CT data showing that the type of
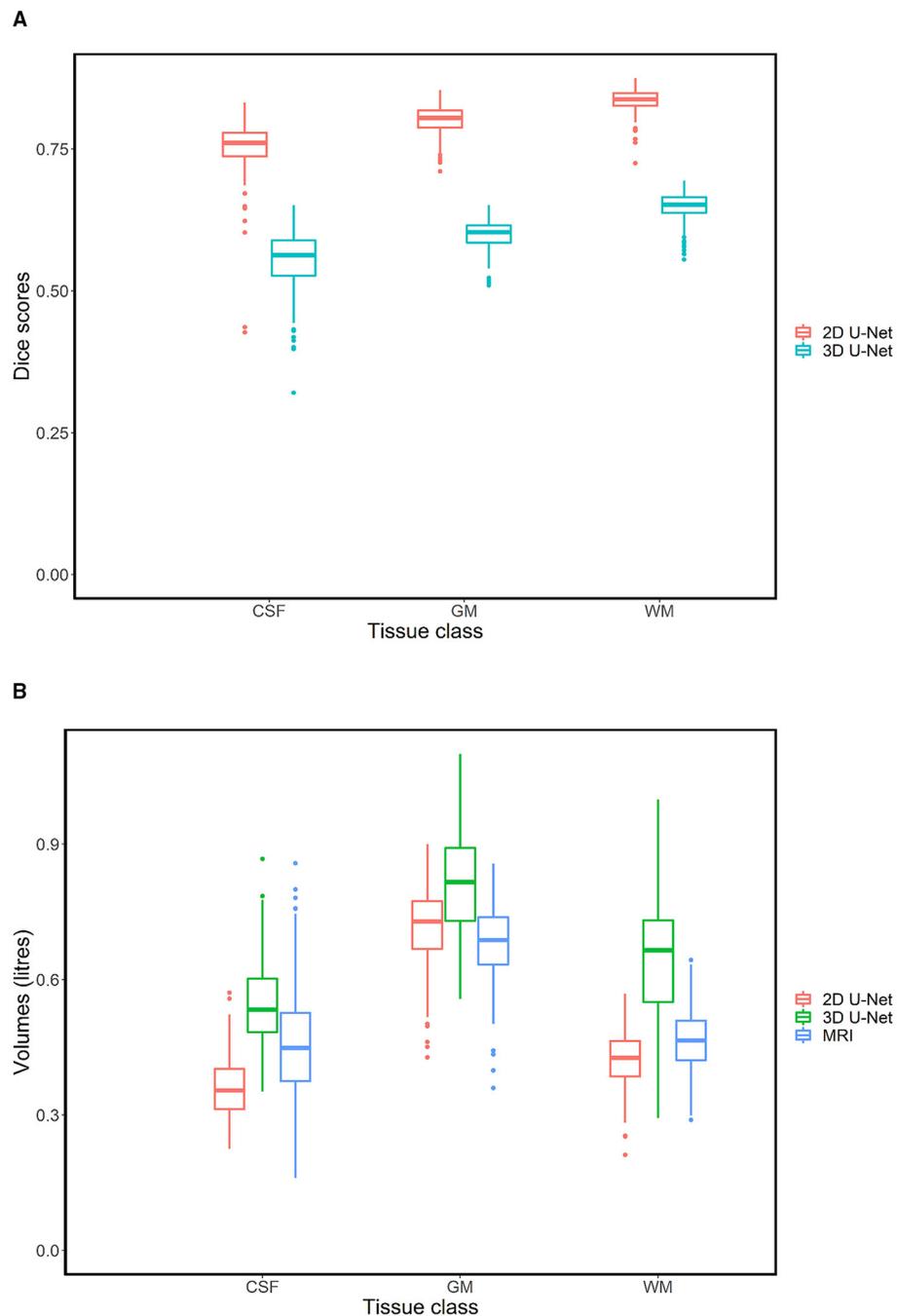
FIGURE 4 | Box plots showing differences in dice scores **(A)** and volumes **(B)**. CT-derived segmentations produced by 2D U-Nets had much better spatial overlap with its paired MR labels in comparison to CT-derived segmentations from 3D U-Nets. With respect to the MR-derived volumes, 3D U-Nets over estimated volumes in all tissue classes in comparison to 2D U-Nets.

model architecture and learning approach is dependent on the nature of the input CT datasets.

To process our CT datasets with 3D networks, either we would have to downsample the whole CT image to a size to suit the available memory or we would process the image in smaller

patches. In modalities such as MRI, where the contrast resolution is high, resizing would not affect the images. However, in the case of CT, where the contrast resolution is lower, we observed that resizing further reduced the contrast resolution of the CT images. Hence, we concluded that downsampling CT images

sacrifice potentially important information in the model training. Therefore, we opted for sliding patch-wise 3D processing. For our patch-based 3D U-Nets to train effectively, we attempted various patch sizes and U-Net depth. We opted for the patch size of 128 × 128 × 32, given that a lower patch size limits the receptive field the deep-learning network can notice, whereas a higher patch size has more memory requirements.

We compared the performance of 2D and 3D CT-based segmentation networks with respect to MR-derived labels using measures of distance, overlap, and volume. Overall, 2D U-Nets performed better than 3D U-Nets in all measures and all tissue classes. In terms of volume, the 2D and 3D U-Nets showed comparable performance (**Table 1**). However, 3D U-Net generally overestimated volume predictions for all tissue classes in comparison with 2D U-Nets. In terms of spatial overlap and distance measures, the single-task-based, slice-wise processed 2D U-Nets performed much better than multitask- and patch-wise processed 3D U-Nets. We attributed this to the nature of the dataset. The main purpose of using 3D U-Nets for 3D data is to capture contextual information in all planes and directions. However, for our datasets, where the thickness along the z-axis is large (5 mm), only limited contextual information can be derived from the z-axis. For such data, maximum information is present in one plane, in our case, the axial plane. Therefore, processing the data slice-wise and segmenting using 2D segmentation networks are more useful than processing patch-wise with 3D segmentation networks. In future, we aim to test this further by slicing the volume along 1D coronal or sagittal CT scans and by comparing the resulting U-Nets.

Various automated approaches and evaluation methods for CT segmentation have been described previously. Gupta et al. (2010) used domain knowledge to improve segmentation by adaptive thresholding, and Kemmling et al. (2012) created probabilistic atlas in standard MNI152 space from MR images. These atlases were transformed into CT image space to extract tissue classes. Cauley et al. (2018) explored the possibility of direct segmentation of CT using an MR segmentation algorithm employed in FSL software. CT segmentations derived from these algorithms lacked sharpness, and some of these results were not validated against manual or MR segmentations or using standard segmentation metrics for comparison. Our 2D U-Net model outperforms the study by Manniesing et al. (2017), which uses four-dimensional (4D) CT to create a weighted temporal average CT image, which is further subjected to CSF and vessel segmentation followed by support vector-based feature extraction and voxel classification for GM and WM segmentation, in terms of Dice coefficients and HDs. Our 3D U-Nets outperforms this study in terms of HDs, which were 14.85 and 12.65 mm for GM and WM, respectively, but underperformed in terms of Dice coefficients.

One of the unique features of our study is the nature of the datasets and the usage of MR-derived labels from MR-based automated segmentation tools as the standard criterion. Even though manual annotations are considered the "gold standard" for CT tissue classification, manual annotations are labor-intensive, rater-dependent, time-consuming, and challenging to reproduce. In the case of our study, we had access to a unique cohort with a large number of CT and MRI datasets collected close to each other in time. The MR-derived labels are apt for the localization and classification of brain tissue. They are easy to extract using reliable and readily available automated software such as FreeSurfer, SPM, and FSL even for large cohorts such as the H70 Birth Cohort.

In terms of neurodegenerative disease assessment, CT and structural MRI are both commonly used for visual assessments. However, even though both of these modalities are commonly used, we cannot ensure that deep learning networks, which have performed successfully in MR-based segmentation, can show similar performance in CT-based segmentation. For instance, in a few studies, patch-based 3D U-Nets have been successfully used for MR-based segmentation (Ballestar and Vilaplana, 2020; Qamar et al., 2020); however, in the case of anisotropic CT-based brain segmentation, 2D U-Nets performed better.

In addition to all these strengths, our study also has some limitations. The choice of using MR-based labels as training inputs has its advantages and disadvantages. In our previous study (Srikrishna et al., 2021), we attempted to understand the difference in the segmentations derived from the two modalities. Our previous studies show that performing brain segmentation with CT instead of MRI incurs a loss of accuracy similar to or less than that of performing brain segmentation on an MR image that has been degraded through the application of Gaussian filtering with an SD of 1.5. The effect of this difference on the development of segmentation models and comparison of MR labels and manual labels as training inputs for these segmentation models needs to be further explored. However, it is a challenging task to derive manual labels for tissue classes such as GM, WM, and CSF in large cohorts. Currently, we only compared U-Net-based segmentation networks. In future studies, we aim to train and compare various state-of-the-art 2D segmentation networks for CT datasets. We also plan to compare various 2D segmentation network architectures, loss functions, and learning methods, as well as find the optimum features for brain tissue class segmentation in CT. We also plan to increase our training data and validate our study in several cohorts since the training on one cohort might not necessarily give the same results in another (Mårtensson et al., 2020). We plan to explore the possibility of using 2.5D segmentation networks for 1D CT brain tissue classification and study the effect of a skull in the segmentation performance in CT scans. We plan to conduct the clinical validation of CT-derived measures for neurodegenerative disease assessment and compare the diagnostic accuracy of CT-based volumes derived using various segmentation algorithms.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data from the H70 cohort cannot be openly shared according to the existing ethical and data sharing approvals; however, relevant data can and will be shared with research groups following the submission of a research proposal to and subsequent approval by the study leadership.

Requests to access the datasets should be directed to Michael Schöll, michael.scholl@neuro.gu.se.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the H70 study was approved by the Regional Ethical Review Board in Gothenburg (Approval Numbers: 869-13, T076-14, T166-14, 976-13, 127-14, T936-15, 006-14, T703-14, 006-14, T201-17, T915-14, 959-15, and T139-15) and by the Radiation Protection Committee (Approval Number: 13-64). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MS contributed to conceptualization, methodology, software, formal analysis, data curation, and writing-original draft preparation. RH and JP contributed to supervision, methodology, and writing-reviewing and editing. GV contributed to methodology and writing-reviewing and editing. AZ and SK contributed to data generation and reviewing and editing. EW contributed to conceptualization and reviewing and editing. IS contributed to funding acquisition, data generation, and reviewing and editing. MS contributed to funding acquisition, conceptualization, supervision, project administration, and writing-reviewing and editing. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Agarap, A. F. (2018). Deep learning using rectified linear units (RELU). *arXiv preprint arXiv:1803.08375*.

Aguilar, C., Edholm, K., Simmons, A., Cavallin, L., Muller, S., Skoog, I., et al. (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., Galimzianova, A., Hoogi, A., Rubin, D. L., and Erickson, B. J. (2017). Deep learning for brain MRI segmentation: state of the art and future directions. *J. Digit. Imaging* 30, 449–459. doi: 10.1007/s10278-017-9983-4

Akkus, Z., Kostandy, P., Philbrick, K. A., and Erickson, B. J. (2020). Robust brain extraction tool for CT head images. *Neurocomputing* 392, 189–195. doi: 10.1016/j.neucom.2018.12.085

Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., and Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* 6:014006. doi: 10.1117/1.JMI.6.1.014006

Ashburner, J., and Friston, K. J. (2005). Unified segmentation. *Neuroimage* 26, 839–851. doi: 10.1016/j.neuroimage.2005.02.018

Ashburner, J., and Friston, K. J. (2007). "Rigid body registration," in *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, eds W. Penny, K. Friston, J. Ashburner, S. Kiebel, and T. Nichols (Amsterdam: Elsevier) 49–62. doi: 10.1016/B978-012372560-8/50004-8

Baid, U., Talbar, S., Rane, S., Gupta, S., Thakur, M. H., Moiyadi, A., et al. (2018). "Deep learning radiomics algorithm for gliomas (DRAG) model: a novel approach using 3D unet based deep convolutional neural network for predicting survival in gliomas," *Presented at the International MICCAI Brainlesion Workshop* (Granada: Springer), 369–379. doi: 10.1007/978-3-030-11726-9_33

Ballestar, L. M., and Vilaplana, V. (2020). MRI brain tumor segmentation and uncertainty estimation using 3D-UNet architectures. *arXiv preprint arXiv:2012.15294*. doi: 10.1007/978-3-030-72084-1_34

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). "Pearson correlation coefficient," in *Noise Reduction in Speech Processing* (London: Springer), 1–4. doi: 10.1007/978-3-642-00296-0_5

Brusini, I., Lindberg, O., Muehlboeck, J., Smedby, Ö., Westman, E., and Wang, C. (2020). Shape information improves the cross-cohort performance of deep learning-based segmentation of the hippocampus. *Front. Neurosci.* 14:15. doi: 10.3389/fnins.2020.00015

Cardoso, M. J., Melbourne, A., Kendall, G. S., Modat, M., Hagmann, C. F., Robertson, N. J., et al. (2011). *Adaptive Neonate Brain Segmentation*. London: Springer. doi: 10.1007/978-3-642-23626-6_47

Cauley, K. A., Och, J., Yorks, P. J., and 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., and Heng, P.-A. (2018). VoxResNet: deep voxelwise residual networks for brain segmentation from 3D MR images. *Neuroimage* 170, 446–455. doi: 10.1016/j.neuroimage.2017.04.041

Despotović, I., Goossens, B., and Philips, W. (2015). MRI segmentation of the human brain: challenges, methods, and applications. *Comput. Math. Methods Med.* 2015:450341. doi: 10.1155/2015/450341

Dubuisson, M.-P., and Jain, A. K. (1994). "A modified Hausdorff distance for object matching," in *Proceedings of 12th International Conference on Pattern Recognition* (Jerusalem: IEEE), 566–568. doi: 10.1109/ICPR.1994.576361

Dunne, R. A., and Campbell, N. A. (1997). "On the pairing of the softmax activation and cross-entropy penalty functions and the derivation of the softmax activation function," *Presented at the Proc. 8th Aust. Conf. on the Neural Networks* (Melbourne, VIC), 185.

Fischl, B. (2012). FreeSurfer. *Neuroimage* 62, 774–781. doi: 10.1016/j.neuroimage.2012.01.021

Gupta, V., Ambrosius, W., Qian, G., Blazejewska, A., Kazmierski, R., Urbanik, A., et al. (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

Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., and Reuter, M. (2020). FastSurfer - A fast and accurate deep learning based neuroimaging pipeline. *Neuroimage* 219:117012. doi: 10.1016/j.neuroimage.2020.117012

Kemmling, A., Wersching, H., Berger, K., Knecht, S., Groden, C., and Nölte, I. (2012). Decomposing the hounsfield unit. *Clin. Neuroradiol.* 22, 79–91. doi: 10.1007/s00062-011-0123-0

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Largent, A., Barateau, A., Nunes, J.-C., Mylona, E., Castelli, J., Lafond, C., et al. (2019). Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int. J. Radiat. Oncol. Biol. Phys.* 105, 1137–1150. doi: 10.1016/j.ijrobp.2019.08.049

Mårtensson, G., Ferreira, D., Granberg, T., Cavallin, L., Oppedal, K., Padovani, A., et al. (2020). The reliability of a deep learning model in clinical out-of-distribution MRI data: a multicohort study. *Med. Image Anal.* 66:101714. doi: 10.1016/j.media.2020.101714

Manniesing, R., Oei, M. T., Oostveen, L. J., Melendez, J., Smit, E. J., Platel, B., et al. (2017). White matter and gray matter segmentation in 4D computed tomography. *Sci. Rep.* 7, 1–11. doi: 10.1038/s41598-017-00239-z

Mendrik, A. M., Vincken, K. L., Kuijf, H. J., Breeuwer, M., Bouvy, W. H., De Bresser, J., et al. (2015). MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. *Comput. Intell. Neurosci.* 2015:813696. doi: 10.1155/2015/813696

Musicco, M., Sorbi, S., Bonavita, V., and Caltagirone, C. (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., and Pantoni, L. (2011). The use of CT in dementia. *Int. Psychogeriatr.* 23, S6–S12. doi: 10.1017/S1041610211000950

Qamar, S., Jin, H., Zheng, R., Ahmad, P., and Usama, M. (2020). A variant form of 3D-UNet for infant brain segmentation. *Future Gen. Comput. Syst.* 108, 613–623. doi: 10.1016/j.future.2019.11.021

Rayment, D., Biju, M., Zheng, R., and Kuruvilla, T. (2016). Neuroimaging in dementia: an update for the general clinician. *Prog. Neurol. Psychiatry* 20, 16–20. doi: 10.1002/pnp.420

Ronneberger, O., Fischer, P., and Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Munich: Springer. doi: 10.1007/978-3-319-24574-4_28

Rydberg Sterner, T., Ahlner, F., Blennow, K., Dahlin-Ivanoff S., Falk, H., Havstam, L. (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., et al. (2018). Increased risk of dementia in subjective cognitive decline if CT brain changes are present. *J. Alzheimers Dis.* 66, 483–495. doi: 10.3233/JAD-180073

Shamir, R. R., Duchin, Y., Kim, J., Sapiro, G., and Harel, N. (2019). Continuous dice coefficient: a method for evaluating probabilistic segmentations. *arXiv preprint arXiv:1906.11031*. doi: 10.1101/306977

Srikrishna, M., Pereira, J. B., Heckemann, R. A., Volpe, G., van Westen, D., Zettergren, A., et al. (2021). Deep learning from MRI-derived labels

enables automatic brain tissue classification on human brain CT. *Neuroimage* 244:118606. doi: 10.1016/j.neuroimage.2021.118606

Thiagarajan, S., Shaik, M. A., Venketasubramanian, N., Ting, E., Hilal, S., and 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. doi: 10.1097/WAD.0000000000000227

Van De Leemput, S. C., Meijs, M., Patel, A., Meijer, F. J., Van Ginneken, B., and 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

Wang, L., Xie, C., and Zeng, N. (2019). RP-Net: a 3D convolutional neural network for brain segmentation from magnetic resonance imaging. *IEEE Access* 7, 39670–39679. doi: 10.1109/ACCESS.2019.29 06890

Wattjes, M. P., Henneman, W. J., van der Flier, W. M., de Vries, O., Träber, F., Geurts, J. J., et al. (2009). Diagnostic imaging of patients in a memory clinic: comparison of MR imaging and 64-detector row CT. *Radiology* 253, 174–183. doi: 10.1148/radiol.2531082262

Willemink, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., et al. (2020). Preparing medical imaging data for machine learning. *Radiology* 295, 4–15. doi: 10.1148/radiol.2020192224

Wu, J., Zhang, Y., Wang, K., and Tang, X. (2019). Skip connection U-Net for white matter hyperintensities segmentation from MRI. *IEEE Access* 7, 155194–155202. doi: 10.1109/ACCESS.2019.29 48476

Zhang, F., Breger, A., Cho, K. I. K., Ning, L., Westin, C.-F., O'Donnell, L. J., et al. (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., and 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. doi: 10.1109/42.906424