Using transfer learning for automated microbleed segmentation

Mahsa Dadar[1]*, Maryna Zhernovaia[2], Sawsan Mahmoud[2], Richard Camicioli[3], Josefina Maranzano[2] and Simon Duchesne[4,5]

[1] Department of Psychiatry, Faculty of Medicine, McGill University, Montreal, QC, Canada,
[2] Department of Anatomy, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada,
[3] Department of Medicine, Division of Neurology, University of Alberta, Edmonton, AB, Canada,
[4] CERVO Brain Research Center, Quebec City, QC, Canada, [5] Department of Radiology and Nuclear Medicine, Faculty of Medicine, Université Laval, Quebec City, QC, Canada

Introduction: Cerebral microbleeds are small perivascular hemorrhages that can occur in both gray and white matter brain regions. Microbleeds are a marker of cerebrovascular pathology and are associated with an increased risk of cognitive decline and dementia. Microbleeds can be identified and manually segmented by expert radiologists and neurologists, usually from susceptibility-contrast MRI. The latter is hard to harmonize across scanners, while manual segmentation is laborious, time-consuming, and subject to interrater and intrarater variability. Automated techniques so far have shown high accuracy at a neighborhood ("patch") level at the expense of a high number of false positive voxel-wise lesions. We aimed to develop an automated, more precise microbleed segmentation tool that can use standardizable MRI contrasts.

Methods: We first trained a ResNet50 network on another MRI segmentation task (cerebrospinal fluid vs. background segmentation) using T1-weighted, T2-weighted, and T2* MRIs. We then used transfer learning to train the network for the detection of microbleeds with the same contrasts. As a final step, we employed a combination of morphological operators and rules at the local lesion level to remove false positives. Manual segmentation of microbleeds from 78 participants was used to train and validate the system. We assessed the impact of patch size, freezing weights of the initial layers, mini-batch size, learning rate, and data augmentation on the performance of the Microbleed ResNet50 network.

Results: The proposed method achieved high performance, with a patch-level sensitivity, specificity, and accuracy of 99.57, 99.16, and 99.93%, respectively. At a per lesion level, sensitivity, precision, and Dice similarity index values were 89.1, 20.1, and 0.28% for cortical GM; 100, 100, and 1.0% for deep GM; and 91.1, 44.3, and 0.58% for WM, respectively.

Discussion: The proposed microbleed segmentation method is more suitable for the automated detection of microbleeds with high sensitivity.

Introduction

Cerebral microbleeds are defined as small perivascular deposits filled by hemosiderin leaking from the vessels (Greenberg et al., 2009). They are recognized as a marker of cerebral small vessel disease, alongside white matter hyperintensities (WMHs) and lacunar infarcts (Wardlaw et al., 2013). Cerebral microbleeds are commonly present in patients with ischaemic stroke and dementia and are more prevalent with increasing age (Roob et al., 1999; Sveinbjornsdottir et al., 2008; Vernooij et al., 2008). The presence of microbleeds has been linked to cognitive impairment and an increased risk of dementia (Werring et al., 2004; Greenberg et al., 2009).

In vivo, cerebral microbleeds can be detected as small, round, and well-demarcated hypointense areas on susceptibility-weighted (SWI) and T2* magnetic resonance images (MRIs) (Shams et al., 2015). Different studies used various size cutoff points to classify microbleeds, with maximum diameters ranging from 5 to 10 mm, and, in some cases, a minimum diameter of 2 mm (Cordonnier et al., 2007). These are well correlated to histopathological findings in the hemosiderin (Shoamanesh et al., 2011). In practice, microbleeds are labeled on MRI as being either "definite" or "possible" using a visual rating (Gregoire et al., 2009). However, visual detection and segmentation are time consuming and subject to inter- and intra-rater variabilities, particularly for smaller microbleeds, frequently overlooked by less experienced raters. Therefore, there is a need for reliable and practical automated microbleed segmentation tools that can produce sensitive and specific segmentations at the lesion level.

Most of the microbleed segmentation methodologies currently available are semi-automated, i.e., they require expert intervention in varying degrees to produce the final segmentation (Barnes et al., 2011; Kuijf et al., 2012, 2013; Bian et al., 2013; Fazlollahi et al., 2015; Morrison et al., 2018). There have also been a few attempts at developing automated microbleed segmentation pipelines based on SWI scans, which in general have a higher sensitivity and resolution for microbleed detection (Roy et al., 2015; Shams et al., 2015; Dou et al., 2016; Van Den Heuvel et al., 2016; Wang et al., 2017; Zhang et al., 2018a; Hong et al., 2020). These attempts were shown to have sensitivity (ranging from 93 to 99 %) and specificity (ranging from 92 to 99%) at a neighborhood ("patch") level but less so at the voxel-wise, lesion level (Dou et al., 2016).

To address these issues, we employed techniques from the deep learning literature. In particular, convolutional neural networks (CNNs) have been successfully employed in many image segmentation tasks. Very deep convolutional networks such as ResNet (He et al., 2016), GoogLeNet (Szegedy et al., 2015), AlexNet (Krizhevsky et al., 2017), and VGGNet (Simonyan and Zisserman, 2014) have shown impressive performances in image recognition tasks. CNNs have also exceeded the state-of-the-art performance in semantic segmentation, with significantly lower performance time, when sufficiently large training datasets with labels are available (Long et al., 2015; Rosenberg et al., 2020; Billot et al., 2021; Isensee et al., 2021). For example, when trained on a large dataset of 5,300 scans of six modalities, SythSeg by Billot et al. (2021) can provide robust segmentations of brain scans of different contrasts and resolutions. Recently, Girones Sanguesa et al. (2021) and Kuijf (2021) used R-CNN and nnU-net models to detect microbleeds in SWI and T2* images. In the absence of large sets of labeled training data, voxel-wise segmentation tasks can be transformed into patch-based segmentation tasks. For example, ResNet50 has recently been used by Hong et al. to detect microbleeds from SWI (Hong et al., 2019), achieving an accuracy of 97.46% at the patch level, outperforming other state-of-the-art methods (Roy et al., 2015; Zhang et al., 2018a,b).

As mentioned, most of these deep learning-based studies only report patch-level results, without assessing their techniques on voxel-wise lesions on a full brain scan. The reported specificities are generally between 92 and 99% (Lu et al., 2017; Wang et al., 2017; Zhang et al., 2018a; Hong et al., 2019, 2020). While high accuracy at a patch level is important, when applied to the whole brain, even a specificity of 99% might translate into thousands of false positive voxels. In fact, applying different microbleed segmentation methods at a voxel level, Dou et al. reported precision values ranging between 5 and 22% in some cases, leading to 280–2,500 false positives on average (Dou et al., 2016). The proposed method by Dou et al. had a much better performance in terms of precision and false positives, with a precision rate of 44.31% and an average false positive rate of 2.74; however, their reported sensitivity was relatively lower (93.16%).

Thus, improving performance at the lesion level would be desirable. Further, given that SWI is not always collected in either clinical or research settings and/or are hard to harmonize in multi-centric settings, it would be useful if a more versatile algorithm was proposed, able to segment microbleeds from other, used more general MRI contrasts (e.g., T1-weighted, T2-weighted, or T2*). To our knowledge, there is no published automated microbleed segmentation tool based on T1w/T2w/T2* acquisitions, which is the first contribution of our article.

The main challenge in developing automated microbleed segmentation tools using machine learning and in particular deep learning methods pertains to a general lack of reliable, manually labeled data. Our second contribution is how we address this problem by using transfer learning to deal with the relatively small number of manually labeled microbleeds in our dataset. Considering promising results from other authors, we employed a pre-trained ResNet50 network, further tailoring it for another, relevant MRI segmentation task, namely the classification of cerebrospinal fluid vs. brain tissue, for which we were able to generate a large number of training samples. The pre-trained weights were then used as the initial weights

for our microbleed segmentation network, allowing for a faster convergence with a smaller training sample.

A third contribution is how we employed post-processing to winnow out false positives. This strategy is described below, along with results on the performance of our technique, both at the patch level and the pixel level. Transfer learning is a powerful paradigm that makes our algorithm potentially versatile on several MRI contrasts for microbleeds detection.

Materials and methods

Participants

Data included 78 subjects (32 women, mean age = 77.16 ± 6.06 years) selected from the COMPASS-ND cohort (Chertkow et al., 2019) of the Canadian Consortium on Neurodegeneration in Aging (CCNA; www.ccna-ccnv.ca). The CCNA is a Canadian research hub for examining neurodegenerative diseases that affect cognition in aging. Clinical diagnosis was determined by participating clinicians based on longitudinal clinical, screening, and MRI findings (i.e., diagnosis reappraisal was performed using information from recruitment assessment, screening visit, clinical visit with physician input, and MRI). For details on clinical group ascertainment, refer to the study by Dadar et al. (2021a,b) and Pieruccini-Faria et al. (2021).

All COMPASS-ND images were read by a board-certified radiologist. Out of the whole cohort, participants in this study were selected based on the presence of WMHs on fluid attenuated inversion recovery MRIs as another indicator of cerebrovascular pathology and microbleeds on T2* images. Consequently, the sample was comprised of six individuals with subjective cognitive impairment (SCI), 30 individuals with mild cognitive impairment (MCI), six patients with Alzheimer's dementia (AD), eight patients with frontotemporal dementia (FTD), seven patients with Parkinson's disease (PD), three patients with Lewy body disease (LBD), five patients with vascular MCI (V-MCI), and 13 patients with mixed dementias. Given that ours is a study on segmentation performance, we assumed that there was no difference in the T2* appearance of a microbleed related to participants' diagnosis.

Ethical agreements were obtained for all sites. Participants gave written informed consent before enrollment in the study.

MRI acquisition

MRI data for all subjects in the CCNA were acquired with the harmonized Canadian Dementia Imaging Protocol [www.cdip-pcid.ca; (Duchesne et al., 2019)]. Table 1 summarizes the scanner information and acquisition parameters for the subjects included in this study.

MRI preprocessing

All T1-weighted, T2-weighted, and T2* images were preprocessed as follows: intensity non-uniformity correction (Sled et al., 1998) and linear intensity standardization to a range of [0–100]. Using a 6-parameter rigid registration, the three sequences were linearly co-registered (Dadar et al., 2018). The T1-weighted images were also linearly (Dadar et al., 2018) and nonlinearly (Avants et al., 2008) registered to the MNI-ICBM152-2009c average template (Manera et al., 2020). Nonlinear registrations were performed to generate the necessary inputs for BISON and were not applied to the data used for training the models. Brain extraction was performed on the T2* images using the BEaST brain segmentation tool (Eskildsen et al., 2012).

CSF segmentation

The Brain tISue segmentatiON (BISON) tissue classification tool was used to segment CSF based on the T1-weighted images (Dadar and Collins, 2020). BISON is an open source pipeline based on a random forests classifier that has been trained using a set of intensity and location features from a multi-center manually labeled dataset of 72 individuals aged 5–96 years (data unrelated to this study) (Dadar and Collins, 2020). BISON has been validated and used in longitudinal and multi-scanner multi-center studies (Dadar and Duchesne, 2020; Dadar et al., 2020; Maranzano et al., 2020).

Gray and white matter segmentation

All T1-weighted images were also processed using *FreeSurfer* version 6.0.0 (*recon-all-all*). *FreeSurfer* is an open-source software (https://surfer.nmr.mgh.harvard.edu/) that provides a full processing stream for structural T1-weighted data (Fischl, 2012). The final segmentation output (aseg.mgz) was then used to obtain individual masks for cortical GM, deep GM, cerebellar GM, WM, and cerebellar WM based on the *FreeSurfer* look up table available at https://surfer.nmr.mgh.harvard.edu/fswiki/FsTutorial/AnatomicalROI/FreeSurferColorLUT. Since *FreeSurfer* tends to segment some WMHs as GM (Dadar et al., 2021c), we also segmented the WMHs using a previously validated automated method (Dadar et al., 2017a,b) and used them to correct the tissue masks (i.e., WMH voxels that were segmented as cortical GM or deep GM by *FreeSurfer* were relabelled as WM, and WMH voxels that were segmented as cerebellar GM by *FreeSurfer* were relabelled as cerebellar WM).

TABLE 1 Acquisition parameters of the harmonized Canadian Dementia Imaging protocol.

Sequence	Scanner model	Matrix	Resolution (mm^3)	Number of slices	TR (msec)	TE (msec)	TI (msec)	Flip angle
T1w	GE	256 × 256	1.0 × 1.0 × 1.0	180	6.7	2.9	400	11
	Philips	256 × 248	1.0 × 1.0 × 1.0	180	7.3	3.3	935	9
	Siemens	256 × 256	1.0 × 1.0 × 1.0	192	2,300	2.98	–	9
T2w	GE	256 × 256	0.94 × 0.94 × 3.0	48	3,000	11	–	125
	Philips	256 × 254	0.94 × 0.94 × 3.0	48	3,000	13	–	90
	Siemens	256 × 256	0.94 × 0.94 × 3.0	48	3,000	10	–	165
T2*	GE	256 × 256	0.94 × 0.94 × 3.0	48	650	20	–	20
	Philips	256 × 256	0.94 × 0.94 × 3.0	48	650	20	–	20
	Siemens	256 × 256	0.94 × 0.94 × 3.0	48	650	20	–	20

TR, repetition time; TE, echo time; TI, inversion time.

Manual segmentation

The microbleeds were segmented by an expert rater (JM > 15 years of experience reading research brain MRI) using the interactive software package Display, part of the MINC toolkit (https://github.com/BIC-MNI) developed at the McConnell Brain Imaging Center of the Montreal Neurological Institute. The software allows visualization of co-registered MRI sequences (T1w, T2w, and T2*) in three planes simultaneously, cycling between sequences to accurately assess the signal intensity and anatomical location of an area of interest. Identification criteria were in accordance with the study by Cordonnier et al. (2007), comprised a round area of hypointensity on T2* within the brain tissue, and based on the exclusion of colocalization with blood vessels based on the T1w and T2w information (Cordonnier et al., 2007). A maximum diameter cutoff point of 10 mm was used to exclude large hemorrhages (Cordonnier et al., 2007). No minimum microbleed size cutoff was used. Eight cases with a varying number of microbleeds were segmented a second time by the same rater (JM) to assess intra-rater variability.

Quality control

We visually assessed the quality of the preprocessed images, as well as the BISON and *FreeSurfer* automated segmentations.

Generating training data

Transfer learning CSF segmentation task

400,000 randomly sampled two-dimensional (2D) image patches were generated from the in-plane (axial plane, i.e., the plane with the greatest resolution) preprocessed and co-registered T2*, T2-weighted, and T1-weighted image slices. Half of the generated patches contained a voxel segmented as CSF by BISON in the center voxel of the patch, and the other half contained either GM or WM in the center of the patch. The patches were randomly assigned to training, validation, and test sets (50, 25, and 25% respectively). To avoid leakage, patches that were generated from one participant were only included in the same set; i.e., the random split was performed at the participant level (Mateos-Pérez et al., 2018). Figure 1 shows examples of the generated CSF and background patches.

Microbleed segmentation task

Similarly, 11,570 2D patches were generated from the preprocessed and co-registered T2*, T2-weighted, and T1-weighted in-plane image slices for microbleed segmentation. One half of the generated patches contained a voxel segmented as microbleed by the expert rater in the center voxel of the patch, and the other half was randomly sampled to contain a non-microbleed voxel in the center. The patches were randomly assigned to the training, validation, and test sets (60, 20, and 20% respectively) also at the participant level. We further ensured to include similar proportions of participants with small (1–4 voxels), medium (5–15 voxels), and large (more than 15 voxels) microbleeds in the three sets. Figure 2 shows examples of the generated microbleed and background patches.

We further augmented the microbleed patch dataset by randomly rotating the patches to generate additional training data. The random rotations were performed on the full slice (not the patches) centering around the microbleed voxel; therefore, the corner voxels in the patches include information from different areas not present in other patches. Matching numbers of novel background patches were also added to balance the training dataset. The performance of the model was assessed using the training dataset with no augmentation and by adding 4, 9, 14, 19, 24, and 29 random rotations to the training set, respectively.

Note that the same MRI dataset was used to generate patches for both CSF segmentation and microbleed segmentation tasks.

FIGURE 1
Examples of CSF and background patches generated for the CSF segmentation task.
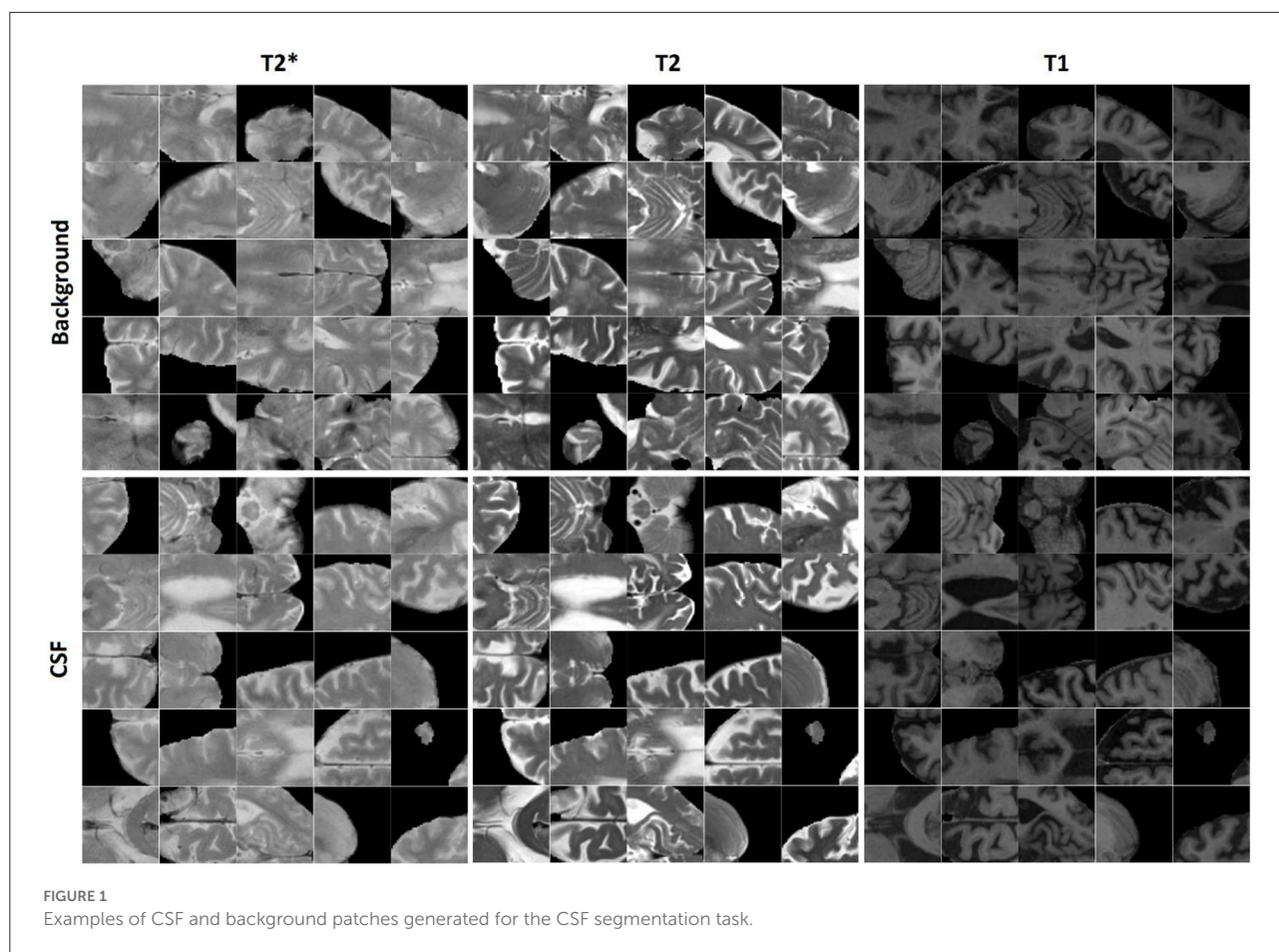
However, the individual patches were not the same, since the microbleed and CSF patches needed to have (respectively) either a microbleed or CSF voxel at their center, while the background patches were randomly sampled from the rest of the dataset.

## Training the ResNet50 network using transfer learning

Previous study by Hong et al. showed that using a pre-trained network based on natural images can be beneficial to microbleed segmentation tasks (Hong et al., 2019). Based on this study, we also used the ResNet50 network (He et al., 2016), a CNN pre-trained on over 1 million images from the ImageNet dataset (Russakovsky et al., 2015), to classify images into 1,000 object categories. This pre-training has allowed the network to learn rich feature representations for a wide range of images, which can also be useful in our task of interest. Our approach was to further train ResNet50 first on a task similar to microbleed segmentation (i.e., CSF vs. tissue) and then on microbleeds identification itself.

To satisfy the input size requirements of ResNet50 network, all patches were resized to 224 × 224 pixels, and the T2*,

T2-weighted, and T1-weighted patches were copied into three channels to generate an RGB image (Figure 3). The last fully connected layer of ResNet50, which contained 1,000 neurons (to perform the classification task for 1,000 object categories), was replaced with two neurons to adapt the network for performing a binary classification task (i.e., object vs. background). The weights of this last fully connected layer were initialized randomly. The network was first trained (all layers, no weight freezing) to perform the CSF vs. tissue segmentation task. We then retrained this network to perform microbleed vs. tissue segmentation. The training was performed on a single NVIDIA TITAN X with 12 GB GPU memory.

Parameter optimization

We assessed the performance of the network with five different patch sizes (14, 28, 52, 56, and 70 mm^2), with and without freezing the weights of the initial layers (no freezing as well as freezing the first 5, 10, 15, and 20 layers), different mini-batch sizes (20, 40, 60, 80, and 100), and different learning rates (0.002, 0.004, 0.006, 0.008, and 0.010). Each experiment was repeated five times by changing one hyper-parameter at a

FIGURE 2
Examples of microbleed and background patches generated for the microbleed segmentation task.
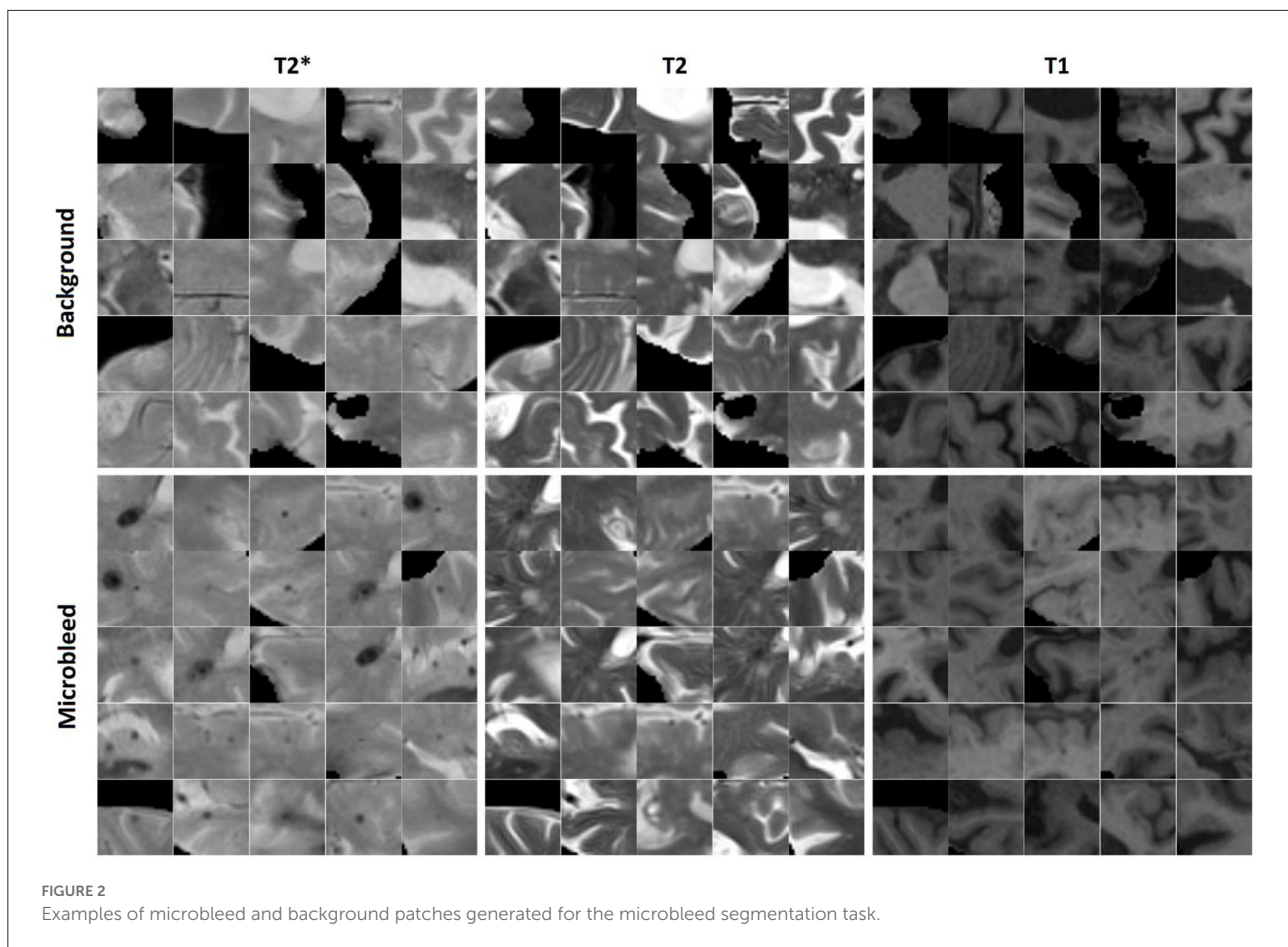
time, and the results were averaged to ensure their robustness. The epoch parameter was set to 10. Stochastic gradient descent with momentum (SGDM) of 0.9 was used to optimize the cross-entropy loss.

Post-processing

After applying ResNet50 to segment all microbleed candidate voxels in the brain mask and reconstructing the final map into a 3D segmentation map, we performed a post-processing step to reduce the number of false positives as well as to categorize the microbleeds into five classes depending on location. In this post-processing step, the microbleeds were first dilated by two voxels. Then, for each voxel at the border of the microbleed, if the ratio between the T2* intensity of the microbleed voxel and the surrounding dilated area ($\frac{Microbleed\ Intensity}{Dilated\ Mask\ Intensity}$) was lower than a specific threshold (to be specified by the user based on sensitivity/specificity preferences), the voxel was removed as a false positive. For each microbleed, the process was repeated iteratively until no voxel was removed as a false positive. The final remaining microbleeds were then categorized into regions (cortical GM,

deep GM, cerebellum WM, and cerebellum GM and WM based on their overlap with *FreeSurfer* segmentations). Figure 4 shows examples of *FreeSurfer*-based tissue categories, some segmented microbleeds, and some dilated surrounding areas.

Figure 5 shows a flow diagram of all the different steps performed in the microbleed segmentation pipeline. All implementations (i.e., generating training patches, training and validation of the model, and postprocessing) were performed using MATLAB version 2020b.
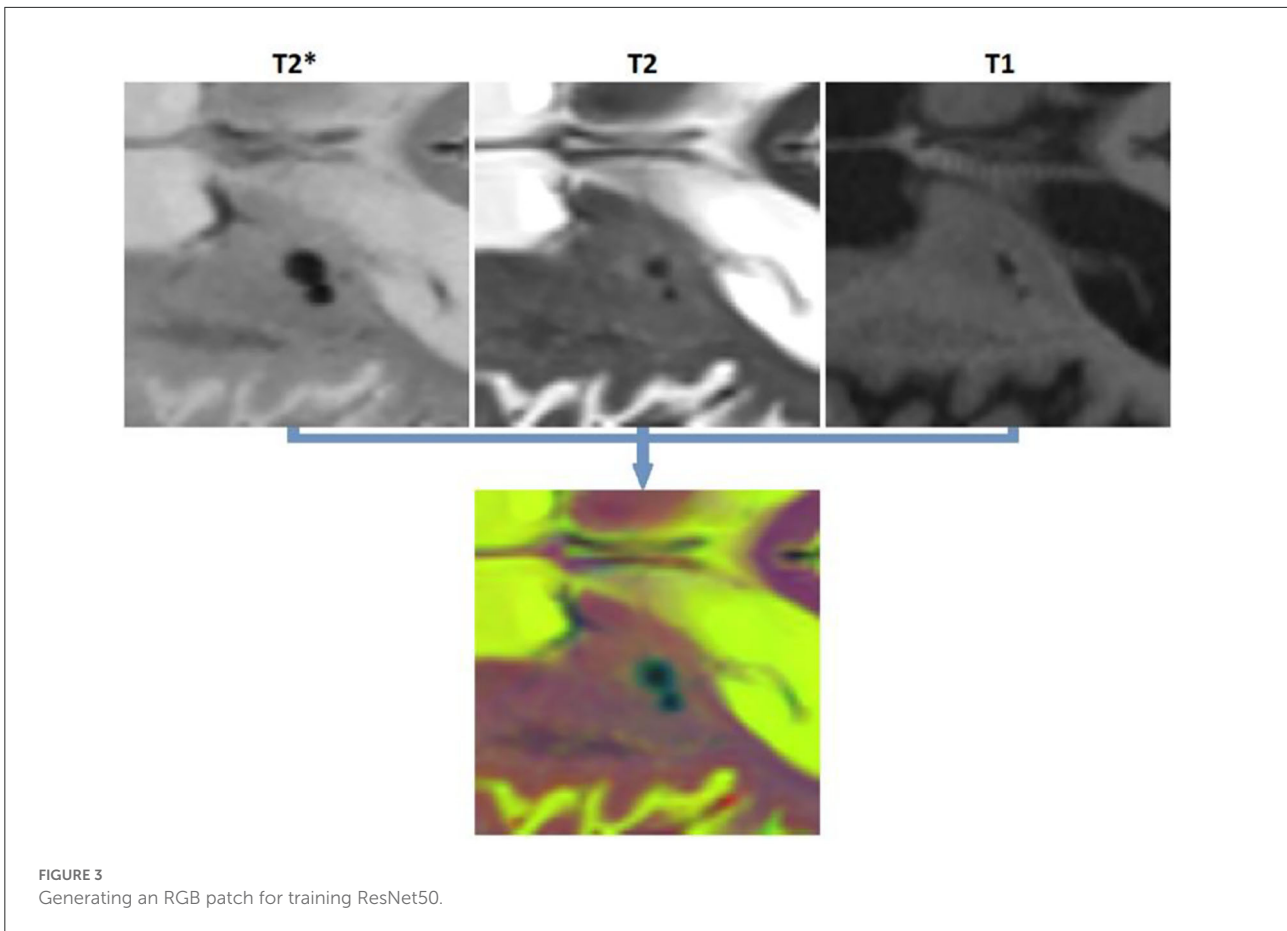
Performance validation

At the patch level, to enable comparisons against the results of other studies, we measured accuracy, sensitivity, and specificity to assess the performance of the proposed method.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (2)$$

$$Specificity = \frac{TN}{TN + FP}, \qquad (3)$$

FIGURE 3
Generating an RGB patch for training ResNet50.

Where TP, TN, FP, and FN denote the number of true positive, true negative, false positive, and false negative patch-level classifications, respectively. While high accuracy at the patch level is desirable, it does not necessarily ensure accurate segmentations at a voxel level. Applied across the entire brain (i.e., 100,000s of pixels), a patch-level specificity of 0.96%−0.99% e.g., as reported by Hong et al., 2019, 2020) might translate into thousands of false positive voxels. To assess the performance at a voxel-wise level, we applied the network to all patches in the brain mask, reconstructed the lesions in 3D across patches, and then measured per lesion sensitivity, precision, and Dice similarity index (Dice, 1945) values between manual (considered as the standard of reference) and automated segmentations.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \qquad (4)$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad (5)$$

$$\text{Dice Similarity Index} = 2\frac{MB1 \cap MB2}{MB1 + MB2} \qquad (6)$$

Here, TP (true positive) denotes the number of microbleeds (i.e., the number of distinct microbleeds segmented, regardless of their size) detected by both methods. FN (false negative)

denotes the number of microbleeds detected by the manual expert but not the automated method. FP (false positive) denotes the number of microbleeds detected by the automated method and not the manual rater. The Dice similarity index shows the proportion of microbleed detection by both methods over the number of microbleeds detected by each method. A Dice Similarity index of 1 shows a perfect agreement between the two methods. A microbleed is considered detected by both methods if there are any overlapping microbleed voxels between the two segmentations. Note that, in accordance with previous studies, specificity was used to reflect the proportion of true negative vs. all negative classifications for patch-level results. However, since specificity cannot be defined at the per-lesion level, we assessed precision instead of specificity for per-lesion results.

U-Net segmentations

A U-Net model was also trained on complete axial slices (256 mm^2) as well as smaller patches (patch sizes of 32, 64, 128 mm^2) from the same dataset for comparison (SGDM, Epoch = 10, initial learning rate = 0.001, mini-batch size = 20). The input images were similar to those used for the ResNet50
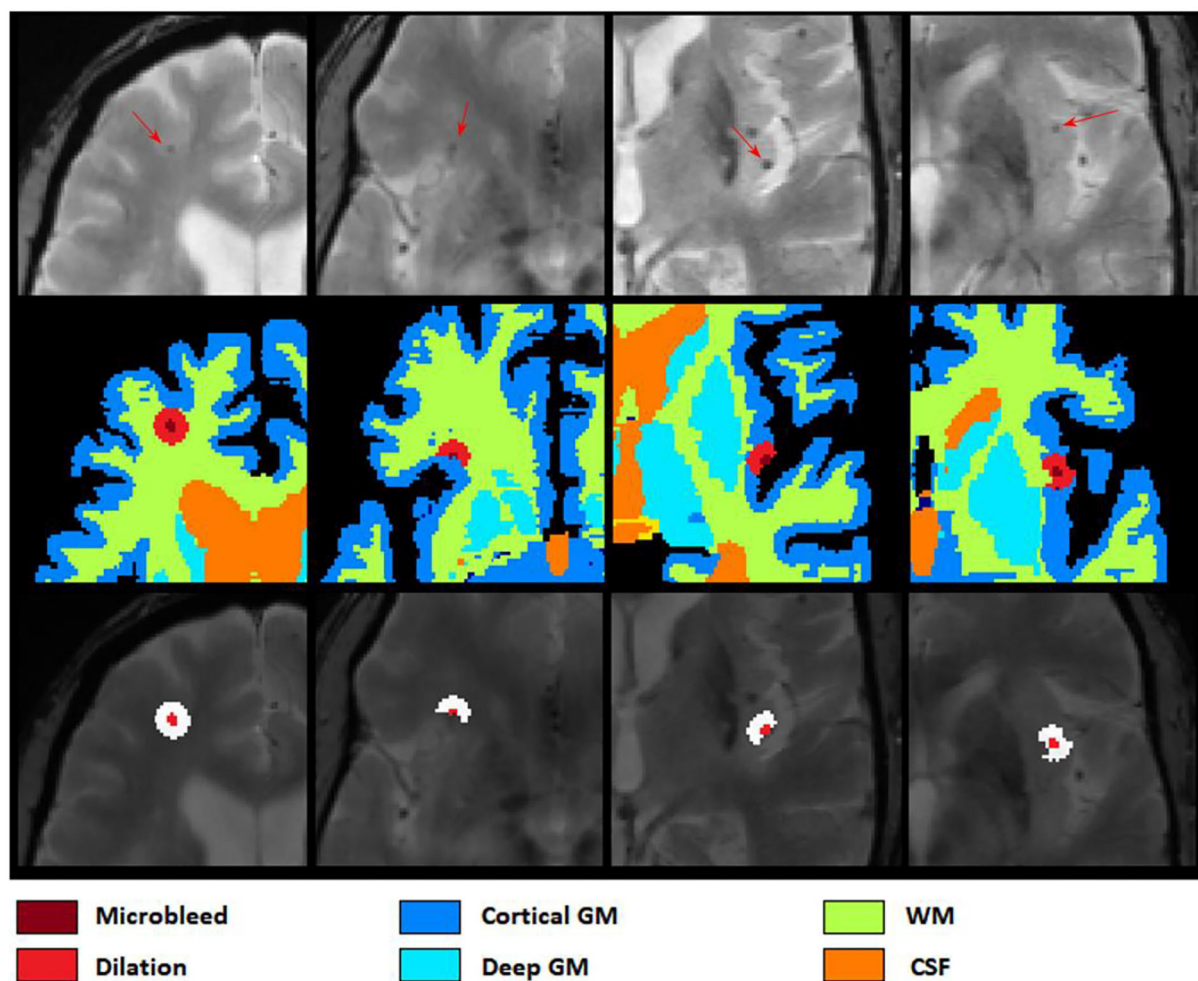
FIGURE 4
Axial slices showing *FreeSurfer* based tissue categories, segmented microbleeds, and their dilated surrounding areas.

model (i.e., T2*, T1-weighted, and T2-weighted axial slices), and the same number of training samples as those used to train the best ResNet50 models (i.e., with data augmentation) were generated for the segmentation tasks. The performance of the model in detecting the microbleeds (i.e., per lesion sensitivity) was then assessed for different patch sizes at the patch level. For the optimal patch size, the voxel level performance vs. the testing time was also assessed for overlapping patches (using the labels from the centering voxels of each patch) with different stride values from 1, 1/2, 1/4, 1/8, and 1/16 of the patch sizes to investigate whether performance improves when using overlapping patches at the expense of longer processing time.

## Data and code availability statement

All image processing steps were performed using the MINC tools, publicly available at: https://github.com/vfonov/ nist_mni_pipelines. BISON (used for tissue classification) is also publicly available at http://nist.mni.mcgill.ca/?p=2148. For more information on the CCNA dataset, please visit https:// ccna-ccnv.ca/. The microbleed segmentation methodology has been reported by the inventors to Université Laval (Report on invention 02351) and is now subject to patent protection (U.S. Trademark and Patent Office 63/257, 536).

Results

Manual segmentation and quality control

The distribution of manually segmented lesions for all 78 participants is shown in Figure 6. Based on the manual segmentations, 46.1, 14.10, and 58.9% of the cases had at least one microbleed in the cortical GM, deep GM, and WM regions, respectively. Only five cases (6.4%) had microbleeds in the cerebellum (in either GM or WM). The overall intra-rater

FIGURE 5
Flow diagram of the microbleed segmentation pipeline.
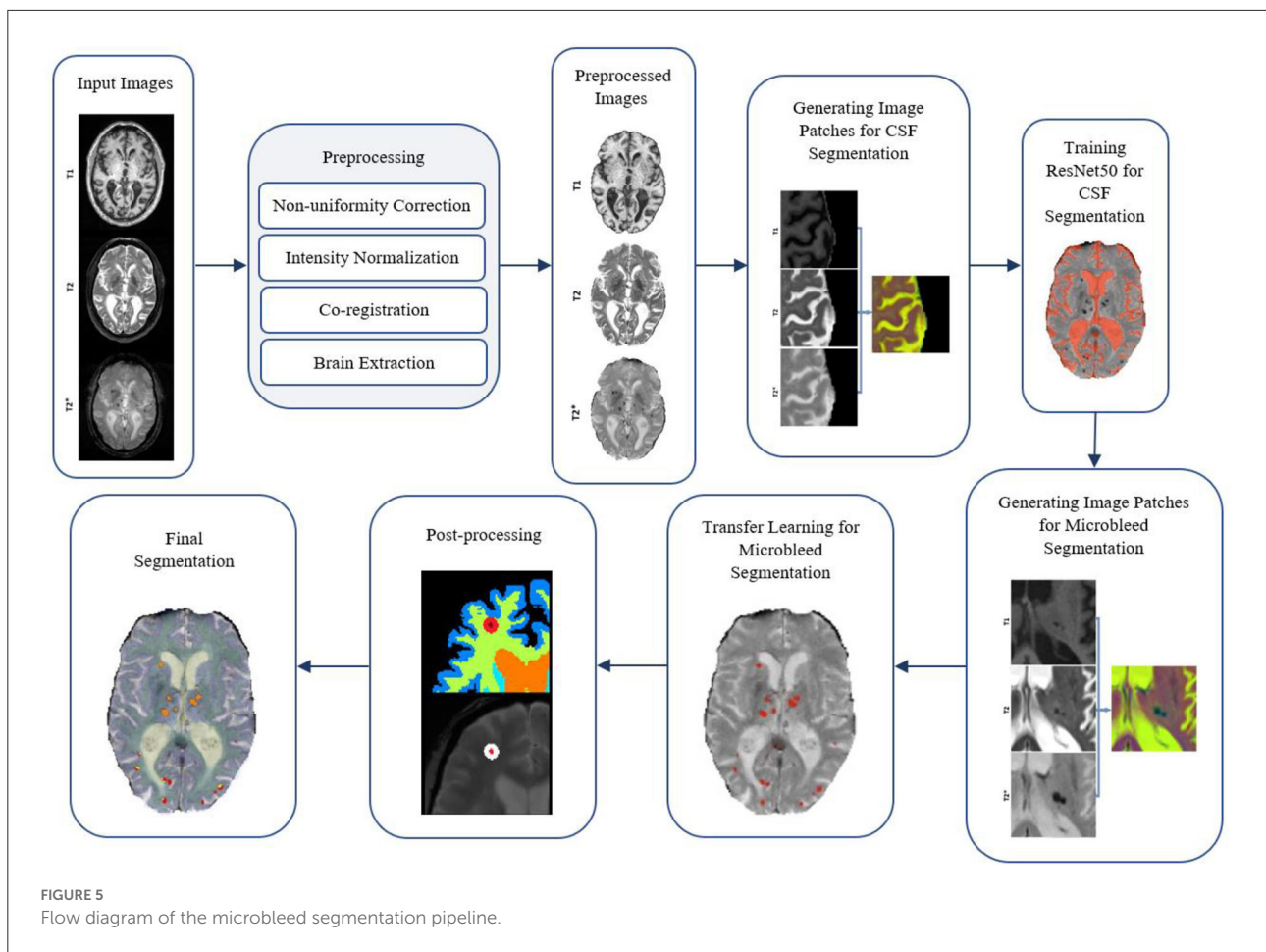
reliability (per lesion similarity index) for manual segmentation was 0.82 ± 0.14 ($\kappa_{\mathrm{CorticalGM}} = 0.78 \pm 0.22$, $\kappa_{\mathrm{DeepGM}} = 1.0 \pm 0.0$, $\kappa_{\mathrm{WM}} = 0.81 \pm 0.15$).

All MRIs passed the visual quality control step for preprocessing and BISON/*FreeSurfer* segmentation.

CSF segmentation

At the patch level, ResNet50 segmentations had accuracies of 0.946 (sensitivity = 0.955, specificity = 0.936) and 0.933 (sensitivity = 0.938, specificity = 0.928) for the validation and test sets, respectively. At a whole brain voxel level, the segmentations had a Dice similarity index of 0.913 ± 0.015 with BISON segmentations. Overall, ResNet50 CSF segmentations (mean volume = 129.06 ± 31.43 cm^3) were more generous in comparison with BISON (mean volume = 117.78 ± 29.51 cm^3). Figure 7 compares the two segmentations, with the color yellow indicating voxels that were segmented as CSF with both methods and the colors purple and green indicating voxels that were only segmented as CSF by BISON or ResNet50, respectively. The majority of the disagreements are in the borders of CSF

and brain tissue, where ResNet50 segments CSF slightly more generously than BISON.

Microbleed segmentation

Figure 8 shows the patch level performance results averaged over five repetitions for different patch sizes, freezing of initial layers, mini-batch sizes, and learning rates. Increasing patch size to more than 28 voxels leads to consistently lower accuracy for both the validation and test sets. Similarly, learning rates higher than 0.004 lowered the performance. The best performance (in terms of accuracy) was obtained with a patch size of 28, freezing the first five initial layers, a mini-batch size of 40, and a learning rate of 0.006.

Figure 9 shows the average performance of the model with these parameters trained with no augmentation as well as the same model trained on original data plus data augmented with 4, 9, 14, 19, 24, and 29 random rotations (no augmentation was performed on validation and test sets). All models with data augmentation performed better than the model without any data augmentation. The best performance was obtained using
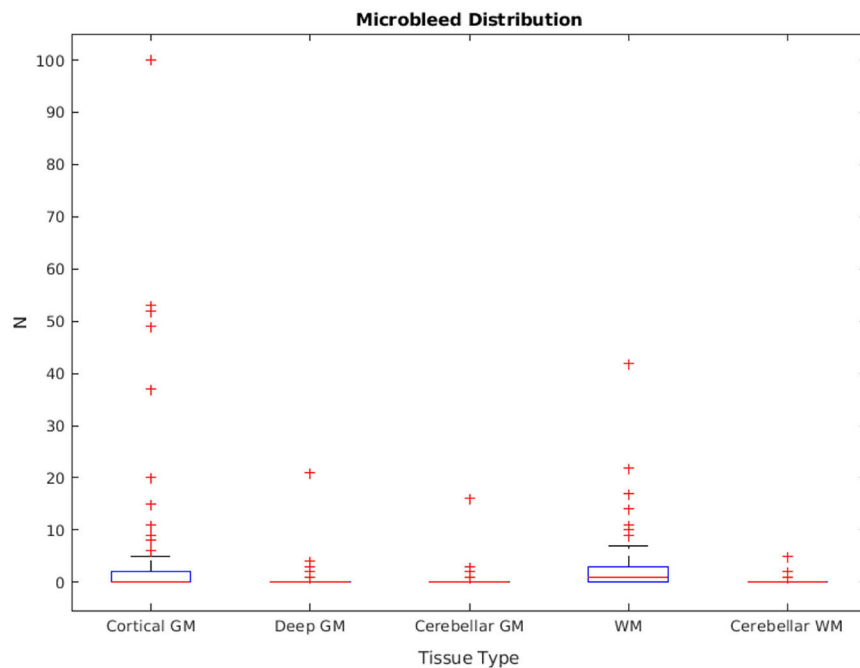
FIGURE 6
Microbleed distribution for each tissue type, based on the manual segmentations.

data augmented with 19 random rotations. For this network, patch-level accuracy, sensitivity, and specificity values were respectively 0.990, 0.979, and 0.999 for the validation set and 0.996, 0.992, and 0.999 for the test set.

Figure 10 shows sensitivity, precision, and similarity index values separately for the cortical, deep, and cerebellar GM and the cerebral and cerebellar WM, after applying post-processing with different thresholds to the voxel-wise segmentations. The threshold values can be selected by the user based on sensitivity and precision preferences.

Figure 11 shows examples of automated vs. manual segmentations for threshold = 1.4 for the deep GM and 1.2 for the rest of the regions, with examples of true positive (indicated in red), false positive (indicated in blue), and false negative (indicated in green) classifications. Most of the disagreements are in the voxels in the border of the microbleeds, where the automated tool sometimes performs a more generous (i.e. blue voxels) or more conservative (i.e. green voxels) segmentation than the manual expert. Such differences will not affect the overall microbleed counts since both methods have essentially identified the same microbleed.

U-Net segmentations

The U-Net model trained on full axial slices had similar accuracy for the CSF segmentation task to the patch-based classification model (mean Dice similarity index of 0.90 ± 0.07),

indicating that, with sufficient training data, this U-Net model is ideal for the CSF classification task, given its low processing time (i.e., <30 s). Figure 12A shows the per lesion sensitivity (applied at the patch level for all patches with microbleeds) of the transfer-learned U-Net model in microbleed segmentation for different patch sizes. The full-slice model was not able to provide accurate segmentations, missing many of the smaller microbleeds. Models trained on smaller patches had better performance, with sensitivity increasing as patch size decreased. Figures 12B,C show the performance and testing time for the most sensitive model at the patch level (i.e., patch size = 32) assessed with overlapping patches, showing increased sensitivity in detecting microbleeds for smaller stride values, at the expense of an increase in testing time. Taken together, these results indicate that, when sufficient training data are not available, the less efficient patch-based models have better performance for microbleed segmentation tasks.

Discussion

In this study, we presented a multi-sequence microbleed segmentation tool based on the ResNet50 network and routine T1-weighted, T2-weighted, and T2* MRI. To overcome the lack of availability of a large training dataset, we used transfer learning to first train ResNet50 for the task of CSF segmentation and then retrained the resulting network to perform microbleed segmentation.
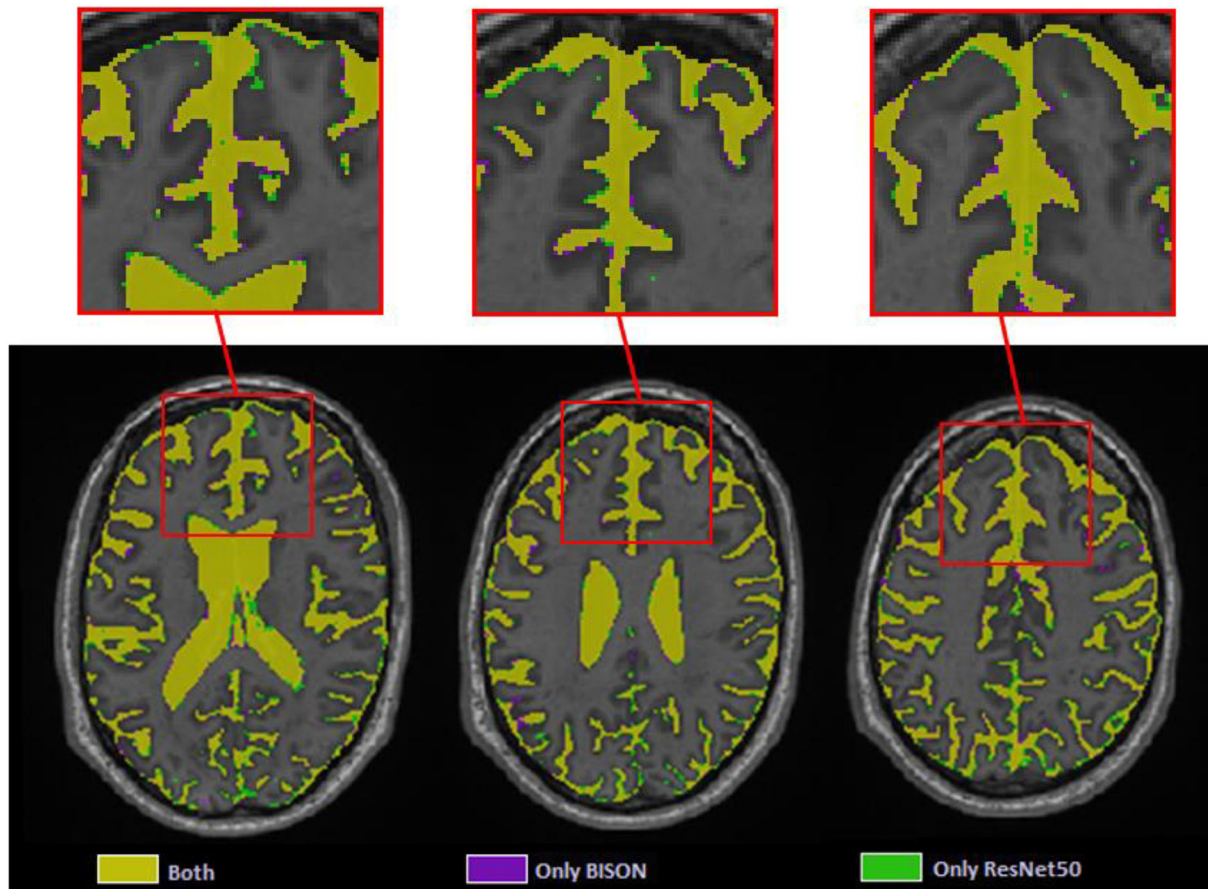
FIGURE 7
Axial slices comparing ResNet50 and BISON CSF segmentations.
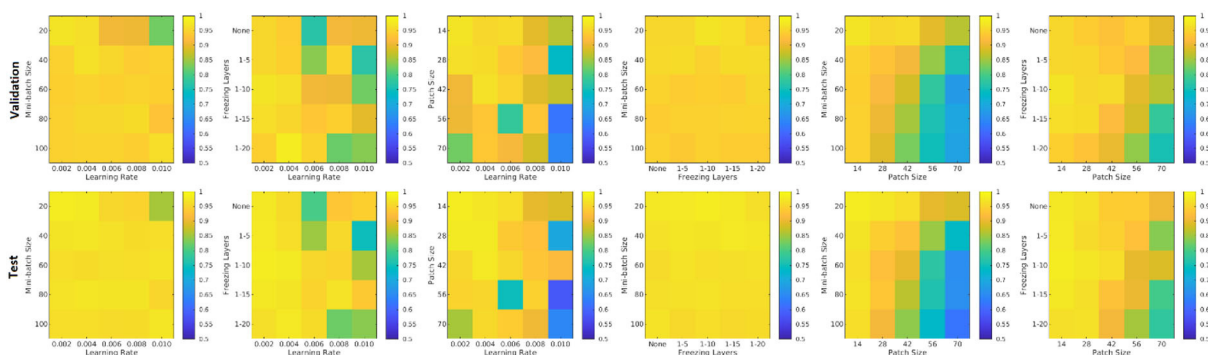
FIGURE 8
Performance accuracy as a function of patch size, freezing of initial layers, mini-batch size, and learning rate. Colors indicate the patch-level accuracy values, with warmer colors reflecting higher accuracy.

Due to the unavailability of ample training data for microbleed segmentation, we transformed the problem at hand to a patch-based classification task, allowing us to obtain better performance at the expense of reduced efficiency and an increase in testing time. For comparison, while a U-Net model trained on complete axial slices (as opposed to smaller

FIGURE 9

Impact of data augmentation on performance accuracy, sensitivity, and specificity at patch level. RR, Random Rotation.

FIGURE 10

Sensitivity, precision, and similarity index values for different post-processing threshold values in validation and test sets.
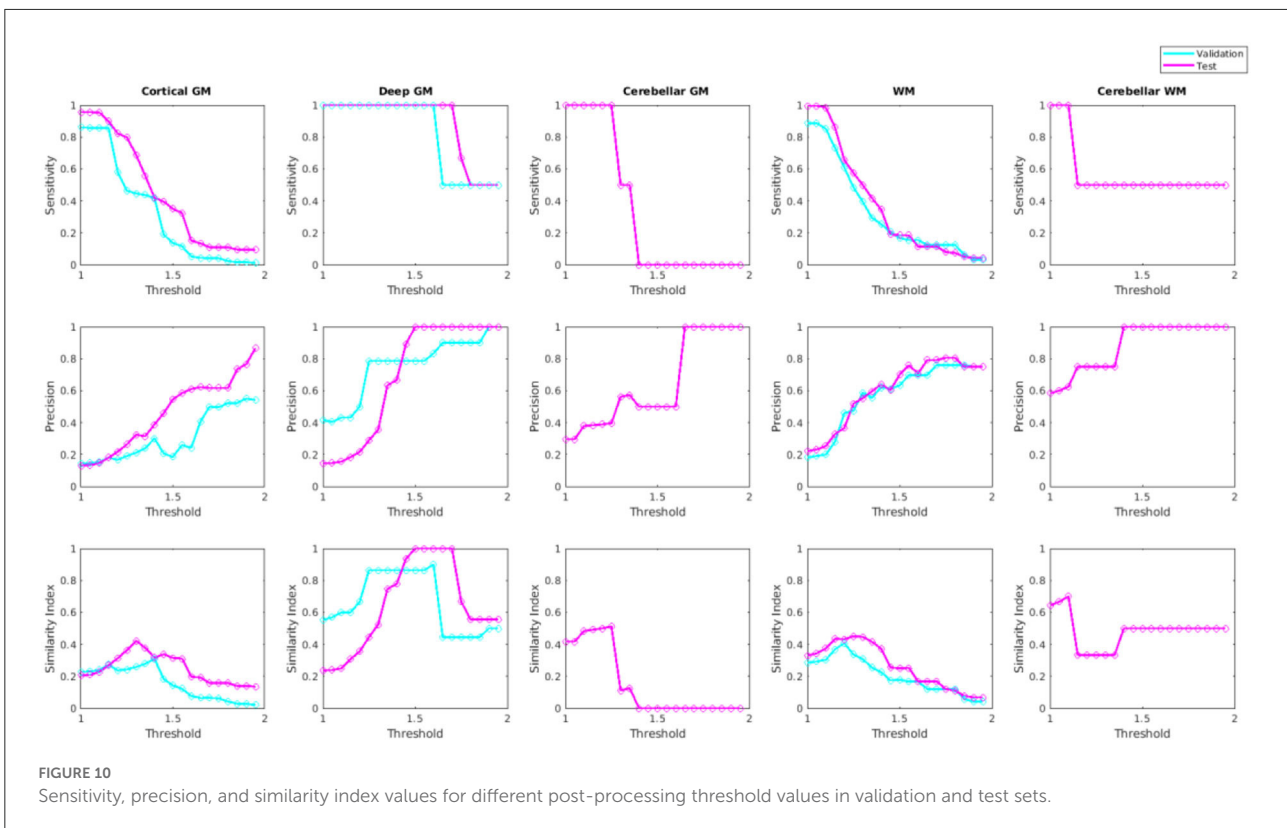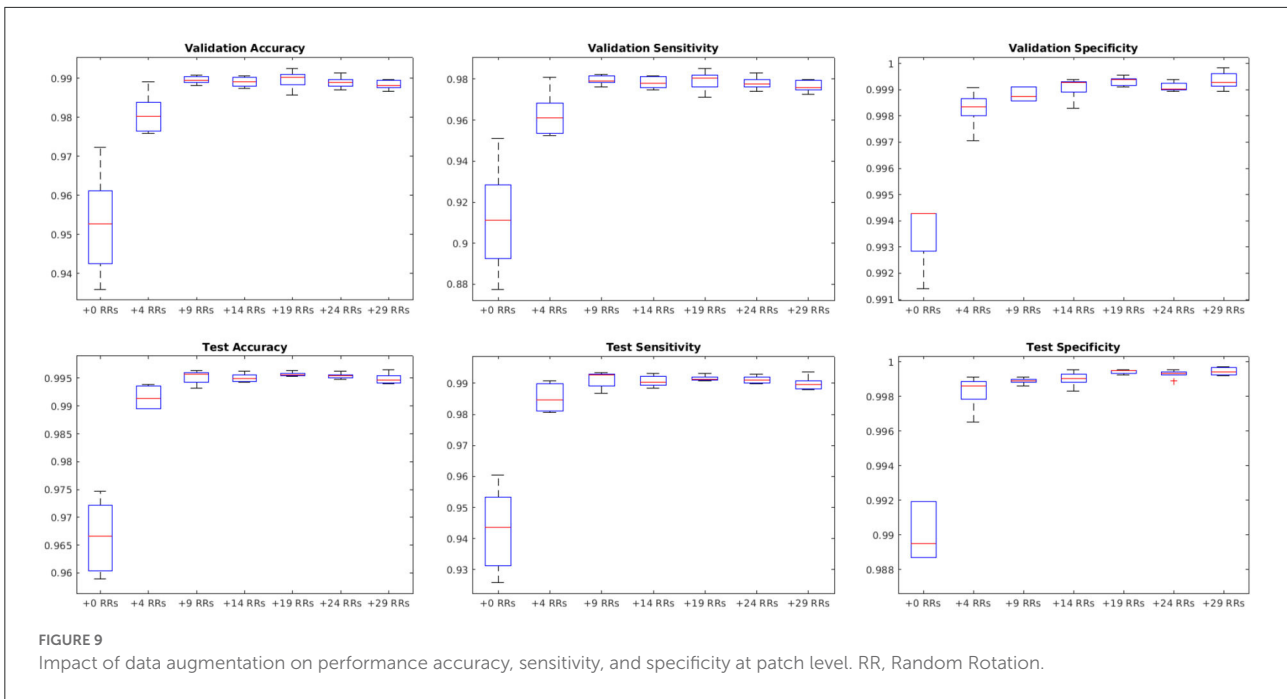
patches) from the same dataset was able to provide CSF segmentations with similar accuracy (mean Dice similarity index of 0.90 ± 0.07 vs. the patch-based equivalent of 0.91 ± 0.01), when applied in a transfer-learned model for microbleed segmentation, it was not able to provide sufficiently accurate microbleed segmentations, missing many of the smaller

FIGURE 11
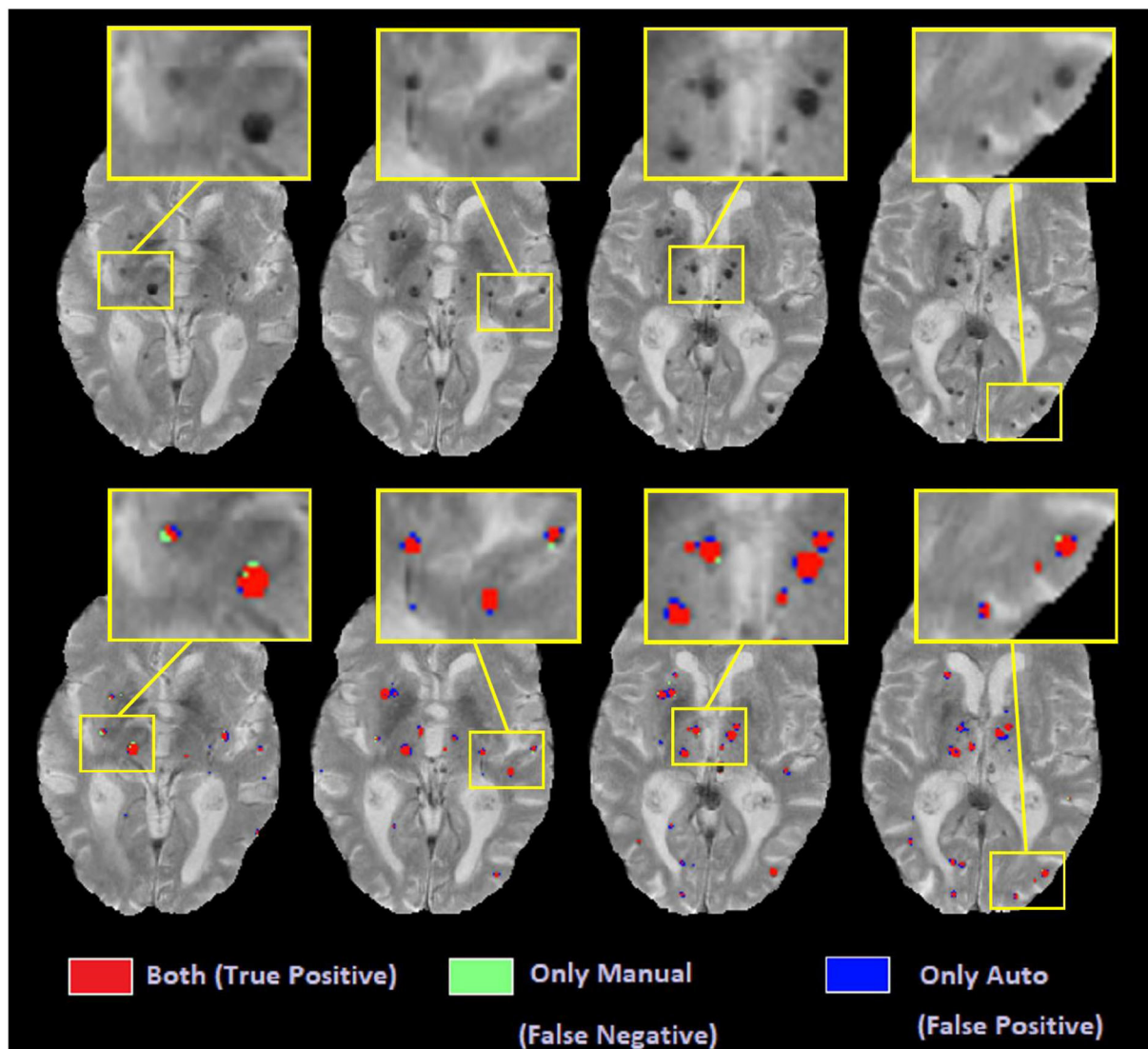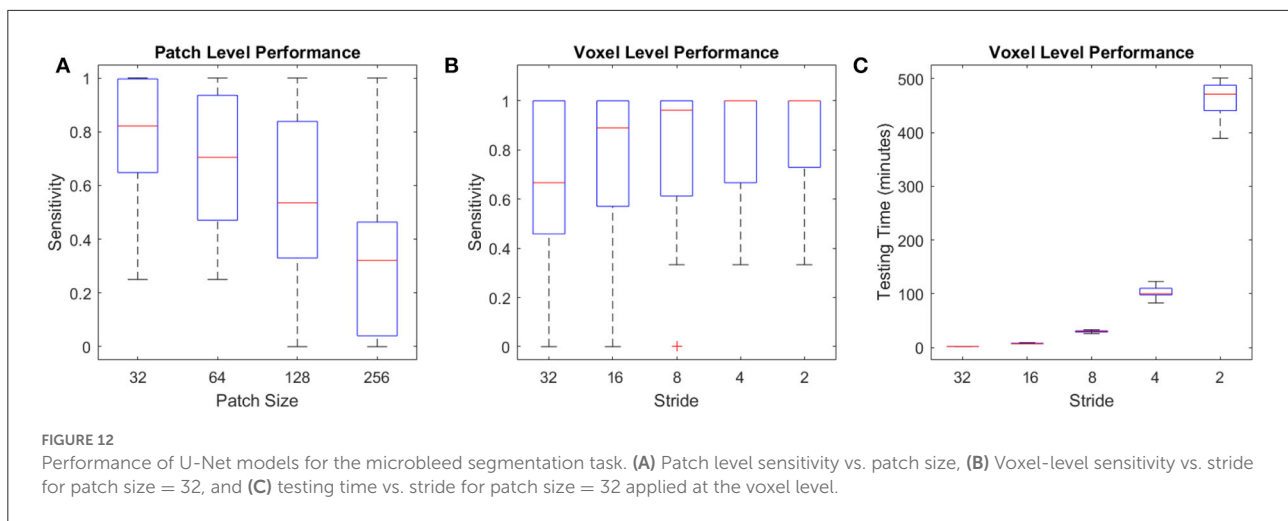Axial slices comparing automated and manual segmentations.

microbleeds (mean sensitivity of 0.22 and mean testing time of 18 s, compared to sensitivity values >0.9 for the proposed model). Along the same line, training the U-Net segmentation model on patches improved the ability of the model to detect microbleeds (mean sensitivity values ranging from 0.64 to 0.86 with stride values from 1, 1/2, 1/4, 1/8, and 1/16 of the patch size) at the expense of an increase in testing time (from 2 min to 7 h per case). Testing time can however be reduced to minutes by limiting the initial search mask (e.g., by excluding CSF voxels or areas that are hyperintense on the T2* image from the initial mask of interest since they would not include any microbleeds) or parallelizing the patch-based segmentation.

Pre-training the model on the CSF segmentation task led to a significant improvement in the segmentation results. Without performing this pre-training step, using the same model and hyper-parameters, we had obtained patch-level accuracy values ranging between 0.74 and 0.96 for the validation and test sets. In comparison, the final transfer learned model was able to achieve patch level accuracies of 0.990 and 0.996 for the validation and test sets, respectively, strongly indicative of the need for this approach.

CSF segmentation was selected as the initial task for pre-training the model since a large number of CSF segmentations could be generated automatically without requiring manual segmentation. Furthermore, since there is no overlap between

FIGURE 12
Performance of U-Net models for the microbleed segmentation task. **(A)** Patch level sensitivity vs. patch size, **(B)** Voxel-level sensitivity vs. stride for patch size = 32, and **(C)** testing time vs. stride for patch size = 32 applied at the voxel level.

CSF and microbleed voxels and given that CSF voxels have a very different intensity profile than microbleeds on T2* images (CSF voxels are hyperintense in comparison with GM and WM, whereas microbleed voxels are hypointense), there would be no leakage between CSF segmentation and microbleed segmentation tasks.

The CSF classification experiments showed excellent agreement (Dice similarity index = 0.91) between ResNet50 and BISON segmentations. The majority of the disagreements were in the voxels bordering the CSF and brain tissue, where ResNet50 segmented CSF slightly more generously than BISON. Given that BISON segmentations were based only on T1-weighted images, whereas ResNet50 used information from T1-weighted, T2-weighted, and T2* sequences, these voxels might be CSF voxels correctly classified by ResNet50 method that did not have enough contrast on T1-weighted images to be captured by BISON.

At the patch level, with sensitivity and specificity values of 99.57 and 99.93%, the microbleed segmentation method outperforms most previously published results. At a per lesion level, the strategy yielded sensitivity, precision, and Dice similarity index values of 89.1, 20.1, and 0.28% for the cortical GM, 100, 100, and 1% for the deep GM, and 91.1, 44.3, and 0.58% for the WM, respectively. Post-processing improved the results (increased the similarity index) for all microbleed types (by successfully removing the false positives). The improvement was most evident for deep GM microbleeds, where most of the false positives were at the boundaries of the deep GM structures, which tend to have lower intensities compared to the neighboring tissue.

There are inherent challenges in comparing the performance of our proposed method against previously published results. Other studies have been mostly based on susceptibility-weighted scans, which in general have higher sensitivity and resolution levels (usually acquired at 0.5×0.5 mm^2 voxels vs. $1 \times$

1 mm^2 voxels in our case) and are better suited for the microbleed detection (Roy et al., 2015; Shams et al., 2015; Dou et al., 2016; Van Den Heuvel et al., 2016; Wang et al., 2017; Zhang et al., 2018a; Hong et al., 2020). Another concern in comparing results across studies regards the characteristics of the dataset used for training and validation of the results. Most previous methods used data from populations that are much more likely to have microbleeds, such as patients with cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy (also known as CADASIL) (Wang et al., 2017; Zhang et al., 2018a; Hong et al., 2020). In contrast, our dataset included non-demented aging adults and patients with neurodegenerative dementia, who do not necessarily have such a high cerebrovascular disease burden. In fact, 63% of the cases in our sample had fewer than three microbleeds. Since we use a participant-level assignment in the training, validation, and test sets, even one false positive would reduce the per-participant precision value by 33%–50% for those cases. In comparison, the training dataset used by Dou et al. (2016) included 1,149 microbleeds in 20 cases (i.e., 57.45 microbleeds per case on average), in which case one false positive detection would only change the reported precision by 1.7%. Along the same line, we included very small microbleeds with volumes between 1 and 4 mm^3 (i.e., 1–4 voxels) in our sample (~35% of the total microbleed count, distributed consistently between training, validation, and test sets), whereas others might choose to not include such very small lesions which are more challenging to detect and also have lower inter- and intra-rater reliabilities (Cordonnier et al., 2007). Regardless, even considering the disadvantage of fewer microbleeds per scan inherent to our population, the proposed method compares favorably against other published results.

Generalizability to data from other scanner models and manufacturers is another important point when developing automated segmentation tools. Automated techniques that have

been developed based on data from a single scanner might not be able to reliably perform the same segmentation task when applied to data from different scanner models and with different acquisition protocols (Dadar and Duchesne, 2020). To ensure the generalizability of our results, we used data from seven different scanner models across three widely used scanner manufacturers (i.e., Siemens, Philips, and GE) from a number of different sites.

Due to the inherently difficult nature of the task, even in manual ratings, inter-rater and intra-rater variabilities in microbleed detection is not very high. The per-lesion intra-rater similarity index for our dataset (based on eight randomly selected cases) was 0.82. Other studies also reported similar results, with one study reporting intra-rater and inter-rater agreements (similarity index) of 0.93 and 0.88, respectively, while others report more modest inter-rater agreements ranging between 0.33 and 0.66 (Cordonnier et al., 2007). In a dataset of 301 patients and using T2* images for microbleed detection, Gregoire et al. reported inter-rater similarity index values of 0.68 for the presence of microbleeds, where the two raters detected 375 (range: 0–35) and 507 (range: 0–49) microbleeds, respectively (Gregoire et al., 2009). Given the relatively high levels of inter-rater and intra-rater variability in microbleed segmentation results, it is also possible that some of the false positives detected by the automated method might be actual microbleed cases that were missed by the manual rater. Regarding the location of the microbleeds, Gregoire et al. reported higher levels of agreement (between the manual ratings) for microbleeds in the deep GM regions, similar to our results. This could be explained by the higher intensity contrast between the microbleeds (greater levels of hypointensity) and the background GM, which has a higher T2* signal than the WM, where the performance is usually less robust for manual raters as well (Gregoire et al., 2009). Finally, the relatively lower performance for cortical GM microbleed segmentation is also expected, given the close proximity to blood vessels (both in the sulci and on the surface of the brain), which show a hypointense signal that confounds with that of the microbleeds, leading to false positives and lowering the precision. However, despite the different levels of performance in the different brain areas, an automated segmentation method has the clear advantage of providing robust segmentations across different runs, essentially eliminating intra-rater variability, which is inevitable in manual segmentations.

Accurate and robust microbleed segmentation is necessary for assessing cerebrovascular disease burden in the aging and neurodegenerative disease populations, who may show a lower number of microbleeds than other pathologies (e.g., CADASIL), making the task more challenging. Additionally, an automated tool that can reliably detect microbleeds using data from different scanner models is highly advantageous. Our results suggest that the proposed method can provide accurate microbleed segmentations in multi-scanner data of a population with a low number of microbleeds per scan, making it applicable for use in large multi-center studies.

Data availability statement

The data analyzed in this study was obtained from the Canadian Consortium on Neurodegeneration in Aging (CCNA; https://ccna-ccnv.ca/), the following licenses/restrictions apply: Requests to access these datasets must first be granted by the CCNA. Requests to access these datasets should be directed to the CCNA, ccna.central@gmail.com.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

Author contributions

MD contributed to the study plan, analyzing the data, and writing of the manuscript. MZ, SM, and JM contributed to MRI manual segmentation and revision of the manuscript. RC and SD contributed to the study plan and revision of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

MD reports funding from the Canadian Consortium on Neurodegeneration in Aging (CCNA) in which SD and RC are co-investigators, the Alzheimer Society Research Program (ASRP), and the Healthy Brains for Healthy Lives (HBHL). CCNA is supported by a grant from the Canadian Institutes of Health Research with funding from several partners including the Alzheimer Society of Canada, Sanofi, and the Women's Brain Health Initiative.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

References

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image Anal.* 12, 26–41. doi: 10.1016/j.media.2007.06.004

Barnes, S. R., Haacke, E. M., Ayaz, M., Boikov, A. S., Kirsch, W., Kido, D., et al. (2011). Semiautomated detection of cerebral microbleeds in magnetic resonance images. *Magn. Reson. Imaging* 29, 844–852. doi: 10.1016/j.mri.2011.02.028

Bian, W., Hess, C. P., Chang, S. M., Nelson, S. J., and Lupo, J. M. (2013). Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images. *NeuroImage Clin.* 2, 282–290. doi: 10.1016/j.nicl.2013.01.012

Billot, B., Greve, D. N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., et al. (2021). Synthseg: Domain randomisation for segmentation of brain mri scans of any contrast and resolution. *ArXiv* Prepr. *ArXiv210709559*. doi: 10.48550/arXiv.2107.09559

Chertkow, H., Borrie, M., Whitehead, V., Black, S. E., Feldman, H. H., Gauthier, S., et al. (2019). The comprehensive assessment of neurodegeneration and dementia: Canadian Cohort Study. *Can. J. Neurol. Sci.* 46, 499–511. doi: 10.1017/cjn.2019.27

Cordonnier, C., Al-Shahi Salman, R., and Wardlaw, J. (2007). Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting. *Brain* 130, 1988–2003. doi: 10.1093/brain/awl387

Dadar, M., Camicioli, R., and Duchesne, S. (2021a). Multi-sequence average templates for aging and neurodegenerative disease populations. *Sci. Data* 9, 238. doi: 10.1101/2021.06.28.21259503

Dadar, M., and Collins, D. L. (2020). BISON: Brain tissue segmentation pipeline using T1-weighted magnetic resonance images and a random forest classifier. *Magn. Reson. Med.* 85, 1881–94.doi: 10.1101/747998

Dadar, M., and Duchesne, S. (2020). Reliability assessment of tissue classification algorithms for multi-center and multi-scanner data. *NeuroImage* 217, 116928. doi: 10.1016/j.neuroimage.2020.116928

Dadar, M., Fonov, V. S., Collins, D. L., and Initiative, A. D. N. (2018). A comparison of publicly available linear MRI stereotaxic registration techniques. *NeuroImage* 174, 191–200. doi: 10.1016/j.neuroimage.2018.03.025

Dadar, M., Mahmoud, S., Zhernovaia, M., Camicioli, R., Maranzano, J., Duchesne, S., et al. (2021b). White matter hyperintensity distribution differences in aging and neurodegenerative disease cohorts. *bioRxiv [Preprint]*. doi: 10.1101/2021.11.23.469690

Dadar, M., Maranzano, J., Misquitta, K., Anor, C. J., Fonov, V. S., Tartaglia, M. C., et al. (2017a). Performance comparison of 10 different classification techniques in segmenting white matter hyperintensities in aging. *NeuroImage* 157, 233–249. doi: 10.1016/j.neuroimage.2017.06.009

Dadar, M., Narayanan, S., Arnold, D. L., Collins, D. L., and Maranzano, J. (2020). Conversion of diffusely abnormal white matter to focal lesions is linked to progression in secondary progressive multiple sclerosis. *Mult. Scler. J.* 27, 208–19. doi: 10.1101/832345

Dadar, M., Pascoal, T., Manitsirikul, S., Misquitta, K., Tartaglia, C., Brietner, J., et al. (2017b). Validation of a regression technique for segmentation of white matter hyperintensities in Alzheimer's disease. *IEEE Trans. Med. Imaging.* 36, 1758–68. doi: 10.1109/TMI.2017.2693978

Dadar, M., Potvin, O., Camicioli, R., Duchesne, S., and Initiative, A. D. N. (2021c). Beware of white matter hyperintensities causing systematic errors in FreeSurfer gray matter segmentations! *Hum. Brain Mapp.* 42, 2734–45. doi: 10.1101/2020.07.07.191809

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297–302. doi: 10.2307/1932409

Dou, Q., Chen, H., Yu, L., Zhao, L., Qin, J., Wang, D., et al. (2016). Automatic detection of cerebral microbleeds from MR images *via* 3D convolutional neural networks. *IEEE Trans. Med. Imaging* 35, 1182–1195. doi: 10.1109/TMI.2016.2528129

Duchesne, S., Chouinard, I., Potvin, O., Fonov, V. S., Khademi, A., Bartha, R., et al. (2019). The Canadian Dementia Imaging Protocol: Harmonizing National Cohorts. *J. Magn. Reson. Imaging* 49, 456–465. doi: 10.1002/jmri.26197

Eskildsen, S. F., Coup,é, P., Fonov, V., Manjón, J. V., Leung, K. K., Guizard, N., et al. (2012). BEaST: brain extraction based on nonlocal segmentation technique. *NeuroImage* 59, 2362–2373. doi: 10.1016/j.neuroimage.2011.09.012

Fazlollahi, A., Meriaudeau, F., Giancardo, L., Villemagne, V. L., Rowe, C. C., Yates, P., et al. (2015). Computer-aided detection of cerebral microbleeds in susceptibility-weighted imaging. *Comput. Med. Imaging Graph.* 46, 269–276. doi: 10.1016/j.compmedimag.2015.10.001

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

Girones Sanguesa, M., Kutnar, D., van der Velden, B. H., and Kuijf, H. J. (2021). MixMicrobleed: multi-stage detection and segmentation of cerebral microbleeds. *ArXiv.* E-Prints arXiv-2108. doi: 10.48550/arXiv.2108.02482

Greenberg, S., Vernooij, M. W., Cordonnier, C., Viswanathan, A., Al-Shahi Salman, R., Warach, S., et al. (2009). Cerebral microbleeds: a guide to detection and interpretation. *Lancet Neurol.* 8, 165–174. doi: 10.1016/S1474-4422(09)70013-4

Gregoire, S. M., Chaudhary, U. J., Brown, M. M., Yousry, T. A., Kallis, C., Jäger, H. R., et al. (2009). The Microbleed Anatomical Rating Scale (MARS): reliability of a tool to map brain microbleeds. *Neurology* 73, 1759–1766. doi: 10.1212/WNL.0b013e3181c34a7d

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 770–778. doi: 10.1109/CVPR.2016.90

Hong, J., Cheng, H., Zhang, Y.-D., and Liu, J. (2019). Detecting cerebral microbleeds with transfer learning. *Mach. Vis. Appl.* 30, 1123–1133. doi: 10.1007/s00138-019-01029-5

Hong, J., Wang, S.-H., Cheng, H., and Liu, J. (2020). Classification of cerebral microbleeds based on fully-optimized convolutional neural network. *Multimed. Tools Appl.* 79, 15151–15169. doi: 10.1007/s11042-018-6862-z

Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., and Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18, 203–211. doi: 10.1038/s41592-020-01008-z

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90. doi: 10.1145/3065386

Kuijf, H. J. (2021). MixMicrobleedNet: segmentation of cerebral microbleeds using nnU-Net. *ArXiv* Prepr. ArXiv210801389. doi: 10.48550/arXiv.2108.01389

Kuijf, H. J., Bresser, d. e., Geerlings, J., Conijn, M. I., Viergever, M. M., Biessels, M. A., et al. K.L. (2012). Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform. *Neuroimage* 59, 2266–2273. doi: 10.1016/j.neuroimage.2011.09.061

Kuijf, H. J., Brundel, M., Bresser, J., d. e., Veluw, S. J., van, Heringa, S. M., et al. (2013). Semi-automated detection of cerebral microbleeds on 3.0 T MR images. *PLOS ONE* 8, e66610. doi: 10.1371/journal.pone.0066610

Long, J., Shelhamer, E., and Darrell, T. (2015). "Fully convolutional networks for semantic segmentation," in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 3431–3440. doi: 10.1109/CVPR.2015.7298965

Lu, S., Lu, Z., Hou, X., Cheng, H., and Wang, S. (2017). "Detection of cerebral microbleeding based on deep convolutional neural network," in: *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. (Chengdu: IEEE) 93–96. doi: 10.1109/ICCWAMTIP.2017.8301456

Manera, A. L., Dadar, M., Fonov, V., and Collins, D. L. (2020). CerebrA, registration and manual label correction of Mindboggle-101 atlas for MNI-ICBM152 template. *Sci. Data* 7, 1–9. doi: 10.1038/s41597-020-0557-9

Maranzano, J., Dadar, M., Arnold, D. L., Collins, D. L., and Narayanan, S. (2020). Automated separation of diffusely abnormal white matter from focal white matter lesions on MRI in multiple sclerosis. *NeuroImage* 213, 116690. doi: 10.1016/j.neuroimage.2020.116690

Mateos-Pérez, J. M., Dadar, M., Lacalle-Aurioles, M., Iturria-Medina, Y., Zeighami, Y., Evans, A. C., et al. (2018). Structural neuroimaging as clinical predictor: a review of machine learning applications. *NeuroImage Clin.* 20, 506–22. doi: 10.1016/j.nicl.2018.08.019

Morrison, M. A., Payabvash, S., Chen, Y., Avadiappan, S., Shah, M., Zou, X., et al. (2018). A user-guided tool for semi-automated cerebral microbleed detection and volume segmentation: evaluating vascular injury and data labelling for machine learning. *NeuroImage Clin.* 20, 498–505. doi: 10.1016/j.nicl.2018.08.002

Pieruccini-Faria, F., Black, S. E., Masellis, M., Smith, E. E., Almeida, Q. J., Li, K. Z. H., et al. (2021). Gait variability across neurodegenerative and cognitive disorders: results from the Canadian Consortium of Neurodegeneration in Aging (CCNA) and the Gait and Brain Study. *Alzheimers Dement.* 17, 1317–28. doi: 10.1002/alz.12298

Roob, G., Schmidt, R., Kapeller, P., Lechner, A., Hartung, H.-P., et al. (1999). MRI evidence of past cerebral microbleeds in a healthy elderly population. *Neurology* 52, 991–991. doi: 10.1212/WNL.52.5.991

Rosenberg, A., Mangialasche, F., Ngandu, T., Solomon, A., and Kivipelto, M. (2020). Multidomain interventions to prevent cognitive impairment, Alzheimer's disease, and dementia: from FINGER to world-wide FINGERS. *J. Prev. Alzheimers Dis.* 7, 29–36. doi: 10.14283/jpad.2019.41

Roy, S., Jog, A., Magrath, E., Butman, J. A., and Pham, D. L. (2015). "Cerebral microbleed segmentation from susceptibility weighted images," in: *Medical Imaging 2015: Image Processing*. (Orlando, FL: International Society for Optics and Photonics), 94131E. doi: 10.1117/12.2082237

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al. (2015). "ImageNet large scale visual recognition challenge," in *IJCV*.

Shams, S., Martola, J., Cavallin, L., Granberg, T., Shams, M., Aspelin, P., et al. (2015). SWI or T2*: which MRI sequence to use in the detection of cerebral microbleeds? The Karolinska Imaging Dementia Study. *Am. J. Neuroradiol.* 36, 1089–1095. doi: 10.3174/ajnr.A4248

Shoamanesh, A., Kwok, C. S., and Benavente, O. (2011). Cerebral microbleeds: histopathological correlation of neuroimaging. *Cerebrovasc. Dis.* 32, 528–534. doi: 10.1159/000331466

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *ArXiv.* Prepr. ArXiv14091556. doi: 10.48550/arXiv.1409.1556

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE Trans. Med. Imaging* 17, 87–97. doi: 10.1109/42.668698

Sveinbjornsdottir, S., Sigurdsson, S., Aspelund, T., Kjartansson, O., Eiriksdottir, G., Valtysdottir, B., et al. (2008). Cerebral microbleeds in the population based AGES-Reykjavik study: prevalence and location. *J. Neurol. Neurosurg. Psychiatry* 79, 1002–1006. doi: 10.1136/jnnp.2007.121913

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9. doi: 10.1109/CVPR.2015.7298594

Van Den Heuvel, T. L. A., Van Der Eerden, A. W., Manniesing, R., Ghafoorian, M., Tan, T., Andriessen, T., et al. (2016). Automated detection of cerebral microbleeds in patients with traumatic brain injury. *NeuroImage Clin.* 12, 241–251. doi: 10.1016/j.nicl.2016.07.002

Vernooij, M. W., van der Lugt, A., Ikram, M. A., Wielopolski, P. A., Niessen, W. J., Hofman, A., et al. (2008). Prevalence and risk factors of cerebral microbleeds: the Rotterdam Scan Study. *Neurology* 70, 1208–1214. doi: 10.1212/01.wnl.0000307750.41970.d9

Wang, S., Jiang, Y., Hou, X., Cheng, H., and Du, S. (2017). Cerebral microbleed detection based on the convolution neural network with rank based average pooling. *IEEE Access* 5, 16576–16583. doi: 10.1109/ACCESS.2017.2736558

Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., et al. (2013). STandards for ReportIng Vascular changes on nEuroimaging (STRIVE v1). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol.* 12, 822–838. doi: 10.1016/S1474-4422(13)70124-8

Werring, D. J., Frazer, D. W., Coward, L. J., Losseff, N. A., Watt, H., Cipolotti, L., et al. (2004). Cognitive dysfunction in patients with cerebral microbleeds on T2*-weighted gradient-echo MRI. *Brain* 127, 2265–2275. doi: 10.1093/brain/awh253

Zhang, Y.-D., Hou, X.-X., Chen, Y., Chen, H., Yang, M., et al. (2018a). Voxelwise detection of cerebral microbleed in CADASIL patients by leaky rectified linear unit and early stopping. *Multimed. Tools Appl.* 77, 21825–21845. doi: 10.1007/s11042-017-4383-9

Zhang, Y.-D., Zhang, Y., Hou, X.-X., Chen, H., Wang, S.-H., et al. (2018b). Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. *Multimed. Tools Appl.* 77, 10521–10538. doi: 10.1007/s11042-017-4554-8