



Synthetic microbleeds generation for classifier training without ground truth

Saba Momeni^{a,b,*}, Amir Fazlollahi^a, Paul Yates^c, Christopher Rowe^d, Yongsheng Gao^b, Alan Wee-Chung Liew^e, Olivier Salvado^f

^a CSIRO Health and Biosecurity, Australian E-Health Research Centre, Brisbane, Australia

^b School of Engineering and Built Environment, Griffith University, Brisbane, Australia

^c Department of Aged Care, Austin Health, Heidelberg, Victoria, Australia

^d Department of Nuclear Medicine and Centre for PET, Austin Health, Heidelberg, Australia

^e School of Information & Communication Technology, Griffith University, Gold Coast, Australia

^f Data61, Brisbane, Australia

ARTICLE INFO

Article history:

Received 27 July 2020

Accepted 21 April 2021

Keywords:

Microbleeds detection

Data augmentation

Neural network

Synthetic data generation

Gaussian modeling

ABSTRACT

Background and Objective: Cerebral microbleeds (CMB) are important biomarkers of cerebrovascular diseases and cognitive dysfunctions. Susceptibility weighted imaging (SWI) is a common MRI sequence where CMB appear as small hypointense blobs. The prevalence of CMB in the population and in each scan is low, resulting in tedious and time-consuming visual assessment. Automated detection methods would be of value but are challenged by the CMB low prevalence, the presence of mimics such as blood vessels, and the difficulty to obtain sufficient ground truth for training and testing. In this paper, synthetic CMB (sCMB) generation using an analytical model is proposed for training and testing machine learning methods. The main aim is creating perfect synthetic ground truth as similar as reals, in high number, with a high diversity of shape, volume, intensity, and location to improve training of supervised methods.

Method: sCMB were modelled with a random Gaussian shape and added to healthy brain locations. We compared training on our synthetic data to standard augmentation techniques. We performed a validation experiment using sCMB and report result for whole brain detection using a 10-fold cross validation design with an ensemble of 10 neural networks.

Results: Performance was close to state of the art (~9 false positives per scan), when random forest was trained on synthetic only and tested on real lesion. Other experiments showed that top detection performance could be achieved when training on synthetic CMB only. Our dataset is made available, including a version with 37,000 synthetic lesions, that could be used for benchmarking and training.

Conclusion: Our proposed synthetic microbleeds model is a powerful data augmentation approach for CMB classification with and should be considered for training automated lesion detection system from MRI SWI.

© 2021 Elsevier B.V. All rights reserved.

Introduction

Cerebral Microbleeds (CMB) are small chronic brain hemorrhages, known as paramagnetic blood products, and likely caused by structural abnormalities of the small vessels. The concept of CMB is primarily a radiological construct describing small MRI signal voids [1],[2]. They are often present with cerebrovascular disease, dementia, Alzheimer's disease, and normal aging [3],[4]. Substantial progress has been made in recent years, in developing

MRI methodologies showing CMB, such as susceptibility-weighted imaging (SWI) [5]. SWI is generated from T2*-weighted gradient-recall echo (GRE) pulse sequence [5] which is sensitive to differences in tissue magnetic susceptibility. In SWI, residual paramagnetic blood products appear with a lower intensity and are therefore suitable for detecting CMB, showing as hypointense amorphous small semi-spherical blobs. However, on visual inspection of SWI scans, CMB appearance can easily be confounded with small blood vessels cross sections as shown in Fig. 1.

The detection of CMB from MRI is clinically important as an imaging marker of vessel diseases and cognitive dysfunctions [1]. The clinical observation of CMB based on the Microbleed Anatom-

* Corresponding author.

E-mail address: sabamomeni2020@gmail.com (S. Momeni).

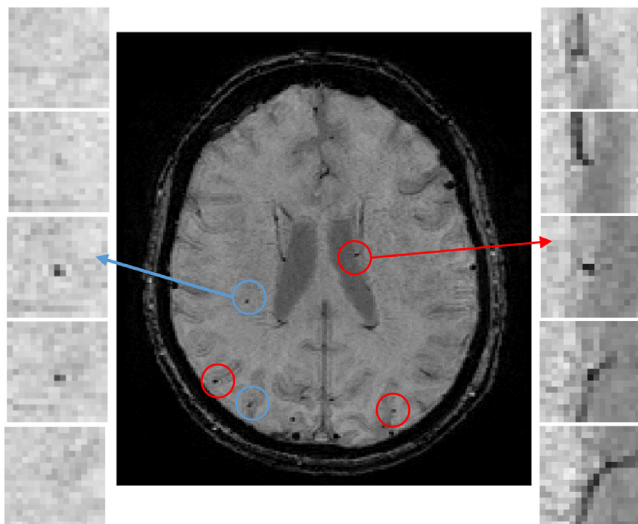


Fig. 1. Example of real CMB (blue circles, left panels) and vessel cross sections (red circles, right panels) in axial view across 5 consecutive slices.

ical Rating Scale (MARS) [6], and the brain observer microbleed scale (BOMBS) [7], is very time consuming and tedious because of large number of similar features among CMB and mimics that include vessel cross sections, calcifications, and cavernous malformation on SWI [8].

Automated detection of CMB is challenging because of their relatively low prevalence: only 10% of healthy individuals' brain in their sixth decades have typically a few CMB visible on MRI SWI [8]. Several computer-aided algorithms to detect CMB have been proposed, ideally with high sensitivity to detect those rare lesions and low false positive rates to avoid generating false alarms. This is a difficult challenge given the large number of pixels in MRI and the presence of numerous mimics [9]. Proposed algorithms are either using image processing, or classifier type machine learning methods, including recent deep learning approaches.

Kujif et al. [10] used Radial Symmetry Transform (RST) [11], which is capable of detecting symmetrical shapes, to detect CMB. Bian et al. [12] applied RST from T2*-weighted MR images with some post-processing to reduce the number of false positives. Machine learning methods [14]–[16] typically screen for candidate locations before extracting features, which are then classified using support vector machine (SVM) [15], random forest (RF) [13], or neural network [14].

To address low prevalence (imbalanced classification), a multi-step RF with bootstrapping and amplifying features from minority classes were applied in two studies [13],[15]. Typical data augmentation techniques including translation, rotation, flipping, adding noise, and gamma correction have been commonly used [9],[16],[17]. Cost sensitive matrix [18] and randomly undersampling majority class [19] are other two common methods.

Deep learning methods [9],[16]–[24] include convolutional layers (CL) followed by fully connected layers (FC). Liu et al. [9] used a CNN with SWI images and quantitative susceptibility mapping (QSM), both using phase (4 channel inputs). Chen et al. [20] exploited 3D deep residual networks, after applying a CMB candidates screening method as proposed in [12]. Lu et al. [21] applied pre-trained VGG, followed by an extreme learning machine (ELM). Hong et al. [22] used ResNet, with transfer learning from ResNet-50 pre-trained on ImageNet dataset.

An important shortcoming of published methods is the lack of validation using different studies, with different pathologies and MRI acquisitions. CMB can be concomitant with multiple pathologies such as Alzheimer disease, mild cognitive disorder

[14],[15], stroke [9],[16], traumatic brain injury [16], or cerebral autosomal-dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) [23],[17],[18],[24,21], and it is unclear whether a method would generalize to all diseases and various MRI parameters.

Indeed, the performance of reported CMB detection algorithms were different for each method, but also with different imaging acquisition conditions and different neurological diseases. In our review, the lowest performances were obtained by image processing algorithms [10],[12],[25],[26], and the best performances were reported from algorithms applying CNN [9],[16]. In most cases the number of used subjects and ground truth were very limited. The reported sensitivity and specificity ranged from 71% [25] to 99% [24], and from 93% [14] to 99% [13], respectively. In some works, the precision and number of FP were not reported. The lowest and highest available precision were 44% [16] and 70% [9], with corresponding average FP of 2.7 and 1.6 per scan and sensitivity of 93.1% and 95.8%. The performance of all those methods are summarized in Table 5.

In this study we propose generating synthetic microbleeds using an ad hoc model based on a range of parameters observed from real CMB: shape, location, intensity, and volume. The main contribution of this paper is three-fold: 1) we propose a new model to create synthetic CMB on SWI; 2) we compare several data augmentation strategies; and 3) we present an extensive validation framework and compare results of neural network classifiers with other state of art methods, including RF.

We have published preliminary results about our model [27], and in this paper we extend our findings to show that our proposed synthetic CMB are similar to real CMB, and that they can provide very large datasets, perfect ground truth, and an extended range of appearance.

Method

MRI acquisition and datasets

Data are from the Australian Imaging Biomarkers & Lifestyle (AIBL) [28]. Approval for the study was obtained from the Austin Health Human Research Ethics Committee and St Vincent's Health Research Ethics Committee, and written informed consent was obtained. All subjects underwent a SWI acquisitions and an anatomical T1-weighted (T1w) on a 3.0T Siemens TRIM TRIO scanner, where SWI were reconstructed online using the scanner system (software VB17). SWI scans were acquired with $0.93 \text{ mm} \times 0.93 \text{ mm}$ in-plane resolution and 1.75 mm slice thickness, repetition time/echo time of 27/20 ms, and flip angle 20° . T1w images were acquired using a standard 3D magnetization prepared rapid gradient echo (MPRAGE) sequence with in-plane resolution of $1.0 \text{ mm} \times 1.0 \text{ mm}$, slice thickness of 1.2 mm, repetition-time/echo-time/T1 = 2.300/2.98/900, flip angle of 9° , field of view of 240×256 , and 160 slices. The N4 bias field correction technique was applied on all the SWI dataset [29].

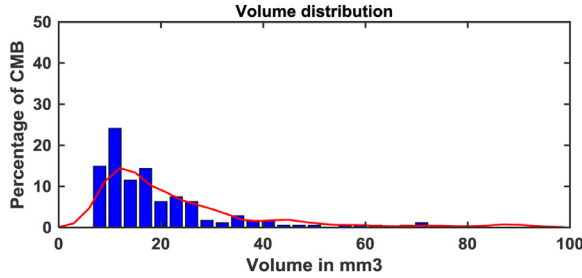
The subjects were classified as Alzheimer's disease (AD), mild cognitive impairment (MCI) and cognitively normal (CN) based on their clinical diagnosis as described in [30]. SWI images were manually inspected for CMB by two clinical experts using the MARS rating method [6]. CMB were considered as spherical hypointense lesions smaller than or equal to 10 mm [2]. CMB were labeled as either definite or possible, where possible ones were less hypointense and were not strictly spherical [31]. 263 possible and 175 definite CMB were marked. After independent reading by two clinical experts, the agreement between experts 1 and 2 was 83%, ($\kappa = 0.65$); expert 1 vs. consensus was 91% ($\kappa = 0.8$); and expert 2 vs. consensus was 91% ($\kappa = 0.83$). The dataset and related clinical information are summarized in Table 1.

Table 1

Clinical and demographics information about our dataset.

Data type	Scan # of AD/MCI/CN/Unknown	CMB # of AD/MCI/CN/Unknown	Subject # / Scan #/ CMB #	Age average \pm standard deviation (F/M)
Whole dataset	94/82/354/2646/43/144/2		214/556/235	75 \pm 6/74 \pm 5
At least one definite CMB	13/17/43/233/35/105/2		41/75/175	77 \pm 8/76 \pm 6
No CMB	72/59/284/240/0/0/0		143/439/0	76 \pm 9/75 \pm 7
Possible CMB only	9/6/27/013/8/39/0		30/42/60	74 \pm 5/73 \pm 3

F: female; M: male; AD: Alzheimer's disease; MCI: mild cognitive impairment; CN: cognitively normal.

**Fig. 2.** Real (blue) and smoothed (red) volume distribution histogram of real CMB.

Real CMB characterization

Real CMB (rCMB) present on MRI SWI as hypointense amorphous small semi-spherical blobs, and we have characterized the 175 definite rCMB present in our dataset using 4 criteria: volume, shape, intensity, and location.

Volume

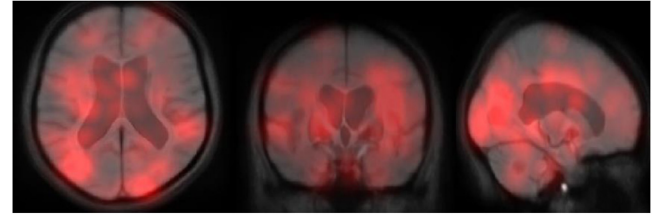
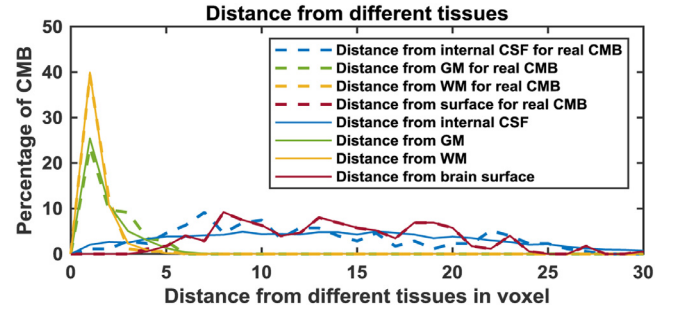
3D patch (size $7 \times 7 \times 7$ pixels) were extracted for each of the 175 definite CMB available. For each patch, a mixture of distributions was fitted to the histogram: a Gaussian to model brain tissues and a uniform distribution to model outliers due to partial volume effect and low intensity pixels from vessels and CMB. That is, we assumed that the intensity of a patch I , could be described as $I = G(\mu, \sigma) + U(0, 255)$, where G and U are a normal distribution and a uniform distribution, respectively. U was fixed with an equal probability for each intensity (1/256) and the parameters of the Gaussian distribution were estimated using the expectation maximization method.

After convergence (error variation $< 0.1\%$), a maximum a posteriori classification was performed to create a mask of the CMB in the patch, cleaned using connected components. Fractional CMB content was estimated for each pixel of the CMB mask by using a standard partial volume model and summed to obtain the rCMB volume. Using this method, Fig. 2 shows the distribution of the estimated volumes (in blue) for all the patches with rCMB, and a smoothed volume distribution (red curve) that we used for sampling synthetic sCMB volumes.

Location

The location of rCMB is shown as a density heat map in Fig. 3 where each definite location was modelled as a Gaussian with standard deviation of 9 voxels in the three directions, and presented as a maximum intensity projection along the 3 main axis overlaid on the average of all definite scans registered together.

Since the location distribution of rCMB was somewhat uniform throughout the brain and only slightly found more posterior, in the occipital lobe (Fig. 3c), we did not use this location information for generating sCMB in the rest of the paper. Instead, we uniformly sample location within the brain tissue. We then investigated location with respect to tissues. T1w scans were segmented into gray

**Fig. 3.** Gaussian distribution of real microbleeds for axial (a), coronal (b) and sagittal (c) views. The maximum intensity projection (red map) was overlaid on the average of all definite scans. The more red, the more likely to find a real lesion.**Fig. 4.** Distance from various tissues for rCMB and proposed sCMB.

matter (GM), white matter (WH), Cerebrospinal fluid (CSF) and internal CSF (ventricles) using a well-established method [32]. 4 distance metrics for CMB were defined as distance to the WM, GM, internal CSF (ventricles) and to the brain mask computed on the SWI using BET toolbox [33]. Fig. 4 shows the distance distribution for rCMB and sCMB.

Shape and Intensity

The average of all rCMB patches with size of $11 \times 11 \times 11$ was computed along with the standard deviation. Visualization along the 3 main directions shows symmetrical centered shape within ± 3 pixels (Fig. 5). The minimum intensity values were computed for all rCMB and used in our model, which resulted in similar distribution (Fig. 6), as explained in the next section.

Synthetic CMB generation

CMB were modelled using a 3D Gaussian function in high resolution (Fig. 7). More details are provided in the Appendix. Briefly, a high-resolution patch 10 times the actual MRI resolution was generated ($70 \times 70 \times 70$) (Fig. 7a). A Gaussian blob in this high resolution patch was randomized for shape and orientation. The high-resolution patch was then thresholded at half maximum value followed by a transformation to obtain 1 for the tissue and 0 at minimum corresponding to the lesion. We experimented with various profiles and found that squaring the resulting patch, thereby sharpening the Gaussian fall off, provided the best performance (Fig. 7b).

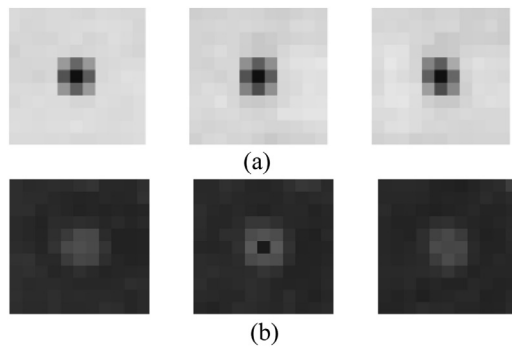


Fig. 5. Average and standard deviation of all definite rCMB. (a) average in top panel and (b) standard deviation in bottom panel of central sections of axial, coronal, and sagittal views (left, middle, right panels).

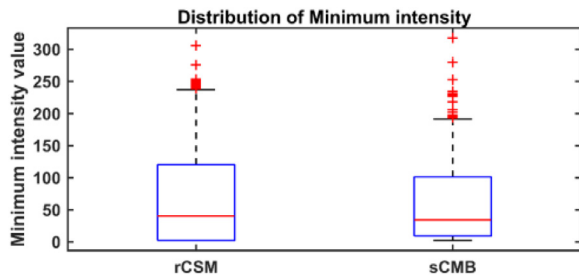


Fig. 6. Minimum intensity values for rCMB and the proposed sCMB.

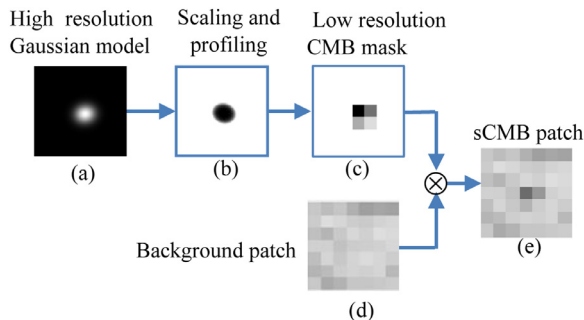


Fig. 7. Different steps of proposed synthetic CMB generation model. (a) Gaussian structure in high resolution with random shift, rotation, and shape. (b) thresholded, inverted, and scaled profile. (c) down sampling to low original SWI resolution to create a CMB mask. multiplication of the mask with a patch without any lesion (d) to create the final sCMB patch (e).

Finally, partial volume was simulated by averaging $10 \times 10 \times 10$ pixels into 1 pixel to yield a low resolution patch $7 \times 7 \times 7$ (Fig. 7c) that could be multiplied at any random location of a SWI scan to add a synthetic lesion as shown in Fig. 7e.

Matching sCMB to rCMB characteristics

An arbitrary large number of sCMB could be generated and their intensity matched the intensity of real lesions (Fig. 6). The location for each sCMB was generated randomly to match the observed location of real lesion in terms of proportion in white and gray matter tissues, but also in terms of distances from the various tissues (Fig. 4). Statistics about location also showed good match between sCMB and rCMB (Fig. 4). The sCMB were generated using random volume sampled from the observed volume distribution after smoothing (Fig. 2 red graph). In Fig. 8 examples of sCMB are shown for different volume, location and intensity.

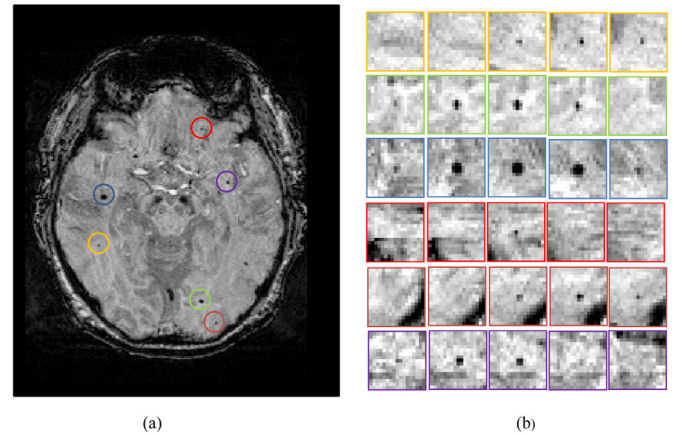


Fig. 8. Samples of proposed sCMB. (a) different shapes, sizes and locations are shown in an axial plane. The volumes are varied from 0.8 to 20 mm³. (b) sCMB are shown in $21 \times 21 \times 5$ patches located in WM (orange, green, blue color), near the skull and CSF (red and dark red), and near a vessel (purple color).

Experiments

We compared our synthetic data generation method to 3 widely used data augmentation techniques and we conducted 6 validation experiments. All the metrics used for validation are described in an Appendix.

Data augmentation models

We considered 4 strategies. In the first model (M1) a majority class under sampling technique was used resulting in around 157 real CMB (157 because the 10-fold cross validation design was across subjects, each with different number of lesions). We used the same number of negative patches (Neg) from location without CMB. The other 3 data augmentation methods were as followed. M2: sCMB generated as described in the previous section for the synthetic positive class (sPos). M3: traditional augmentation method using the data from M1, including rotation (90° , 180° , 270°) and flipping around x, y, z axis resulting in around 5,000 Pos augmented rCMB (Aug). M4: synthetic minority oversampling technique (SMOTE) [34], by linear combination of the 3 closest neighbors in terms of volume.

In all experiments, a screening pass was performed to select candidate locations from the whole SWI, from which all the Neg patches were selected. Screening was done using the radial symmetry transform method [11]. Local minimum points from the RST symmetry map were considered as the CMB candidates. All the patches were normalized by applying zero mean and unit variance computed over the whole dataset.

Validation experiment

We considered an ensemble of artificial neural networks (ANN) with two layers for CMB detection and we performed validation experiments using only synthetic data to reserve the real CMB for testing. First, the number of neurons on the two hidden layers was optimized. We used a 10-fold cross validation design and averaged outputs from 5 ANN (ensemble of classifiers). 50% of the subjects with possible, and 30% of subjects with no CMB were used in this part, yielding 3,000 synthetic positive (sPos) and 3,000 negative (Neg) patches. Testing was done on an additional 600 samples (300 sPos and 300 Neg) not included in the training. For each combination of parameters, we fixed the sensitivity at 95% and computed specificity, AUC, and accuracy, which were averaged to yield a com-

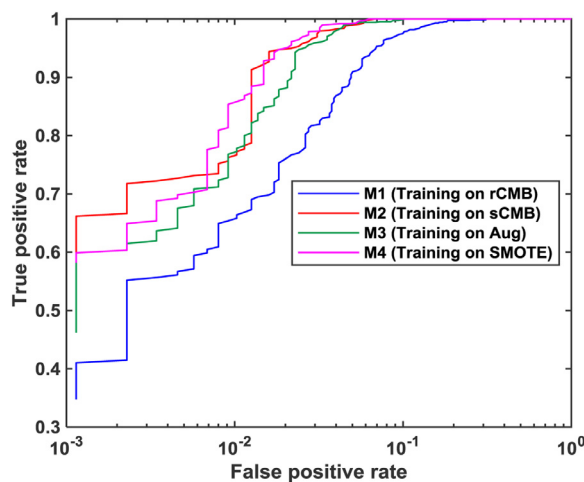


Fig. 9. Comparing ROC curves for 4 different training models on 8,000 samples.

posite score. The result of a grid search¹ resulted in the optimal combination of 10 neurons on hidden layer 1 and 70 on hidden layer 2.

We then computed the composite score for an increasing number of synthetic data. No improvement was noticed after 16,000 samples (8,000 sPos and 8,000 Neg), which we used as the size of the dataset required for saturating the performance of the ANN classifier (as defined above)². Similarly, we used the composite performance metric to test how many epochs were needed for training the ANN classifier, using 8,000 and 16,000 samples. We tested a range of epochs using 10-fold cross validation and an ensemble average of 5 ANNs. After 400 epochs no improvement could be seen³.

Comparison of data augmentation models for patch classification

In this experiment, 50% of subjects with possible, 70% of subjects with no rCMB, and all the subjects with at least one definite CMB were used. RST [11] was used to screen CMB candidate locations to extract negative patches. Eventually, 8,000 patches were extracted without CMB that made up the negative class (4000) and the sPos class (4000) once sCMB were added using model M2. The rCMB were used (as described in the previous paragraph) for the data augmentation model M3 and M4. For this experiment, 10-fold cross validation and averaging an ensemble of 10 ANNs were used. In addition, results from 5 randomized initializations (network weights and data order) were averaged. M1 was trained with around 157 samples, M3 could be trained up to 8,000 (maximum number of combination for traditional data augmentation), whereas M2 and M4 training was done on 16,000 samples. Results are shown in Fig. 9 for 8,000 (and around 157 for M1)⁴ and for 16,000 (M2 and M4 only)⁵.

Quantitative performance is summarized in Table 2 for the four models. AUC were very high (~99%) with little difference in performance between the two best methods: M2 and M4. When the sensitivity was set to 95%, M2 and M4 specificities were also similar followed by M3 and M1. When the dataset size was increased to 16,000, overall performance increased slightly with the same order between methods (right side of Table 2).

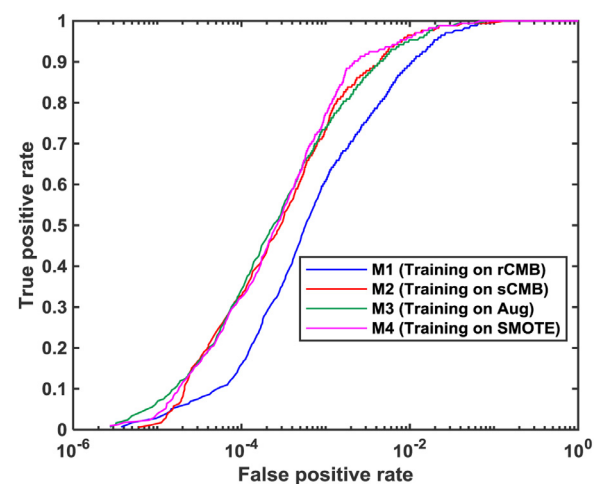


Fig. 10. Comparing ROC curves of 4 different training models for CMB classification on the whole SWI after training on 8,000 samples.

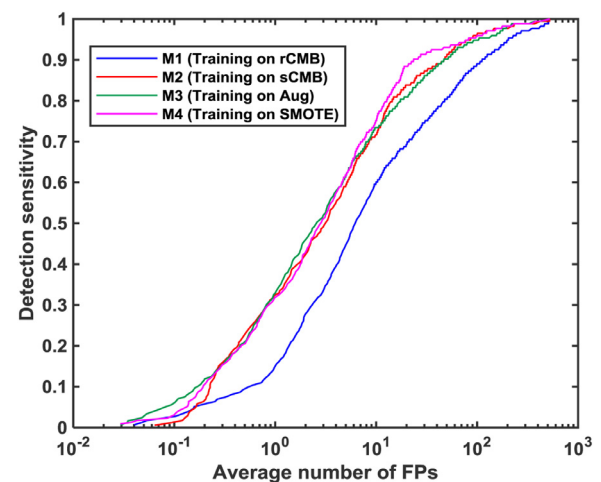


Fig. 11. Comparing number of FP from 4 different training models for CMB detection on the whole SWI after training on 8,000 samples.

Whole SWI CMB detection

CMB detection was performed on whole MRI SWI, a more relevant clinical situation: the method would be applied to a new subject's scan to estimate how many CMB are present, and where. The mix of subjects was similar as above, with synthetic CMB randomly added on healthy brain locations. Like previously, RST was used first as a screening step to reduce the number of candidate locations and thereby accelerating the testing. RST missed only one rCMB overall (sensitivity of 99.43%) and identified approximately 7,000 candidate locations for each scan, down from about 750,000 possible locations per scans. Figs. 10–11 show the results for training on 8,000 and 16,000 samples, using ROC and Free-response Receiver Operating Characteristic (FROC) curves (average number of false positive per scan). All the trainings were done using 10 fold-cross validations (on the subjects) with an ensemble average of 10 ANNs. For the 8,000 samples case, the average over 5 random initializations (ANN weights and data order) is shown.

To compute the number of true positives for each threshold, all the positive detections within 2 voxels of each rCMB were considered to be a unique positive. Therefore, FP samples were identified as a detection farther than 2 voxels from any rCMB. In addition, smoothing with a Gaussian filter (std 2.5 voxels) was applied and local maxima were counted to yield the final number of FP.

¹ See supplementary Fig. 1

² See supplementary Fig. 2

³ See supplementary Fig. 3

⁴ See supplementary Fig. 4

⁵ See supplementary Fig. 5

Table 2

Results of CMB patch classification. Specificity is reported when sensitivity was set to 95%.

Training model	reported results (8,000 samples)		reported results (16,000 samples)	
	AUC	Spe	AUC	Spe
M1	0.984 ± 0.0039	0.931 ± 0.0142	–	–
M2	0.995 ± 0.0008	0.978 ± 0.0048	0.997	0.994
M3	0.993 ± 0.0013	0.974 ± 0.0065	–	–
M4	0.995 ± 0.0014	0.979 ± 0.0035	0.998	0.994

AUC: area under ROC curve; Spe: specificity.

Table 3

Result of whole SWI CMB detection for 95% sensitivity.

Training model	results for 8,000 samples			results for 16,000 samples		
	AUC	Spe	#FP/ scan	AUC	Spe	#FP/scan
M1	0.995 ± 0.0003	0.981 ± 0.0047	208.4 ± 45	–	–	–
M2	0.997 ± 0.0007	0.992 ± 0.0007	86.7 ± 7	0.997	0.991	93.2
M3	0.997 ± 0.0007	0.990 ± 0.0013	110.8 ± 14	–	–	–
M4	0.998 ± 0.0008	0.992 ± 0.0004	85.6 ± 6	0.998	0.992	85.8

AUC: area under ROC curve; Spe: specificity.

Although performance was different, the ranking of the performance for the 4 methods was similar to the patch classification experiment: the highest AUC was found for M4 and M2, then M3, while M1 was lower (details in Table 3 and Fig. 10)⁶. Using a pre-set sensitivity of 95%, the specificity was close to or above 99%. The number of false positive per scan for 95% sensitivity (Fig. 11) were 85.62 ± 6 and 86.73 ± 7 for M4 and M2, whereas M3 and M1 had 110.86 ± 14 and 208.44 ± 45 FP per scan respectively⁷. Adding more samples on whole SWI had no specific improvement for both M2 and M4 Table 3 summarizes our results⁸.

Re-training existing random forest classifier

We investigated whether training using synthetic CMB could improve a more sophisticated classifier. The binary random forest classifier proposed by Fazlollahi et al [13] was re-trained using 4,000 synthetic lesions (sCMB) and tested on real lesions (rCMB) by the original author. Because we could generate sCMB in larger number than the rCMB available, the new training set was imbalanced with a ratio of 1:5 (#Pos:#Neg) whereas the original method had an imbalanced ratio of 1:260. The classifier used 3 stages with increasing specificity. The final results for the same originally reported sensitivity of 92.6% were 9.2 FP per scan with 24.5% precision, an improvement over the original published results of 10.2 FP [13].

Comparing 10-fold vs 2-fold cross validation

In order to investigate the performance degradation due to smaller datasets, we compared 2-fold and 10-fold cross validation design. We used 8,000 samples and compared M2 and M4 with a 2-fold cross validation repeated 25 times with an ensemble of 10 ANN; results were averaged. ROC curves are shown in supplementary materials⁹ and Table 4 compares performance metrics between 2 and 10-fold for the same 95% sensitivity. As hypothesized, the performances using M4 (SMOTE) degraded when using 2-fold since less data could be used for training, whereas performances when using M2 (synthetic lesions) remained the same. The AUC and specificity were higher for M2 than for M4 when 2-fold was

used. The result was statistically significant for AUC (p-value <1e-06) and borderline for specificity (p-value ~0.09) using one tailed T-test.

Comparison with reported results from the literature

We compared our results using ANN trained on synthetic data to existing published methods, as summarized in Table 5. However, because all the published methods use different datasets, it is very difficult to show a fair comparison. Up to the section "Re-training existing random forest classifier", all our results were achieved on our dataset. Since we provide the ROC, one can easily read any operating point (sensitivity and specificity) and in Table 2 we report the specificity for a sensitivity of 95% (suitable for clinical application).

In Table 5, we report what performance our method could achieve, using our data, for the sensitivity that was reported in each of the individual publication (i.e. fixing the sensitivity to the one published by each method and reading the specificity from our ROC shown in Fig 10). The data and methods are different, but at least we could compare all the methods using the same operating point.

In addition, in Table 5 row 8, we re-trained Fazlollahi's method using our data and we show the new specificity for the same sensitivity that was reported in the original paper (to make the comparison easier). That way we could compare the effect of using synthetic data using the same method and the same data.

For all CMB detection algorithms based on image processing techniques [12],[26],[25], our method had higher specificity (99.7, 99.8, 99.9), and fewer FP (28.76 vs 44.9, 12.8 vs 49, and 9.7 vs 17). When comparing with machine learning techniques [15, 14], our proposed CMB detection also resulted in fewer FP (17.3 vs 107.5, 60.5 vs 513) and higher specificity (99.8 vs 95.9, 99.4 vs 93). For methods based on convolutional neural network [9],[16],[24], our results had slightly more FP and less specificity for the same reported sensitivity, although the retrained random forest based method using sCMB came close.

One of the studies [24] had a very small number of subjects and it remains to be seen whether the performance would generalize on larger databases. The best performing study [9] used complex SWI and QSM images, in addition to using a more complex Convolutional Neural Network, resulting in the lowest number of FP reported: 1.6 per scan. However, every method had specific MRI acquisition parameters and disease application (patient population), hampering fair comparison.

⁶ See supplementary Fig. 6⁷ See supplementary Fig. 7⁸ See supplementary Fig. 8⁹ See supplementary Fig. 9

Table 4

Table below compares 2-fold vs 10-fold cross validation CMB patch detection for 95% sensitivity.

Training model	10-fold cross validation (8,000 samples)		2-fold cross validation (8,000 samples)	
	AUC	Spe	AUC	Spe
M2	0.995 ± 0.0008	0.978 ± 0.0048	0.994 ± 0.0009	0.9769 ± 0.0046
M4	0.995 ± 0.0014	0.979 ± 0.0035	0.992 ± 0.0014	0.967 ± 0.0047
p-value	0.7205	0.6541	<1e-06	0.0929

AUC: area under ROC curve; Spe: specificity.

Table 5

CMB detection performance of recent state-of-the-art publications and our proposed ANN trained using synthetic data.

Method	MRI mod/Mag. field	#subjects/ #CMB	Reported results from previous works					Our results with the same matched sensitivity published by each paper			
			Sen %	Acc %	Precision%	Spe %	#FP per scan	Acc %	Precision %	Spe %	# FP per scan
S. Liu et al. [9]			95.8	N/A	70.9	N/A	1.6	99.1	2.3	99.1	91
J. Hong et al. [24]	SWI+QSM/1.5T	1230/1641	99.7	98.3	N/A	96.8	69*	90.2	0.5	90.7	522.1
Wang et al. [18]	SWI/3T	10/NA	96.9	97.1	N/A	97.1	N/A	98.7	1.9	98.9	135.1
Zhang et al. [14]	SWI/3T	20/NA	93	93	N/A	93	513 *	99.4	3.5	99.4	60.5
Zhang et al. [23]	SWI/3T	20/NA	95.1	94.2	N/A	93.3	492 *	99.2	2.6	99.2	88.1
Q. Dou et al. [16]	SWI/3T	320/1149	93.1	N/A	44.3	N/A	2.7	99.4	3.5	99.4	61.9
Fazlollahi et al. [13]	SWI/3T	66/108	92.6	98.3	N/A	99	10.2	99.4	3.6	99.4	59.1
Fazlollahi's method retrained with synthetic data		92.6	96.8	18.6	96.8	9.2	99.4	3.6	99.4	59.1	
Bian et al. [12]	T2*/3T	15/304	86.5	N/A	N/A	N/A	44.9	99.7	6.9	99.7	28.7
Kuijf et al. [26]	T2*/3T	38/148	78	N/A	N/A	N/A	49	99.8	12.5	99.8	12.8
Kuijf et al. [25]	T2*/7T	18/66	71.2	N/A	N/A	N/A	17	99.9	15.1	99.9	9.7
Barnes et al. [15]	SWI/1.5T	6/126	81.7	N/A	N/A	95.9	107.5	99.1	10.3	99.8	17.3

Sen: sensitivity; Spe: specificity; Acc: accuracy; * inferred from other reported results. Bold numbers highlight better performance.

Open dataset

Variations in microbleeds characteristics such as intensity, shape, location, can vary depending on diseases, patient populations, and MRI parameters. Therefore, comparing any automated detection methods against each other and investigating their generalization is challenging. We did not have the opportunity to test our method on different studies, but we hope to be able to do so in future work. We believe that sharing datasets would be a step towards designing clinically meaningful methods, and aligned with that goal, we are making our datasets public, including the expert ground truth. To do that, for subjects with definite microbleeds and No CMB, we randomly selected 30 and 100 subjects respectively, and created synthetic microbleeds that were added at random location as explained in our method section¹⁰. All the data can be accessed here: <https://doi.org/10.25919/aeqy-ny12>.

Discussion

We proposed a method to create synthetic microbleeds to train classifiers (ANN) that achieved excellent performance when tested on real lesions on both a database of patches and for lesion detection on whole MRI scans. When our synthetic data were used to train a random forest classifier, it achieved close to state-of-the-art

performance with less than 10 false positives per scans. Our proposed synthetic data generation was compared to other augmentation methods and showed similar or better performance despite not using any ground truth lesions for training.

There are several benefits using synthetic lesions for training machine learning. Data can be generated in very large number with a high diversity of shape, volume, intensity, and location, that would be otherwise difficult to obtain from real lesions. Indeed, in previous works CMB detection have been affected by variation in size, shape, and intensity [12],[15]. For examples, some studies excluded CMB with rare sizes [12],[15], while CMB with irregular shape or near the skull were also excluded from another study [13], or reported as false negatives because of CMB unavailability with this features in their dataset [16]. Using our proposed synthetic data for training, a relatively simple ANN could detect a diversity of lesions ranging from very dark large volume with irregular shape to very pale small lesions, even though very few actual examples existed in the testing data (Fig. 12). In addition, training and validation for any classifier can be done using synthetic data only, thereby keeping the actual data for testing.

Some of our FP might be real lesions. It is challenging, even for expert radiologists, to confidently label a hypointense feature as a microbleed. Using 95% sensitivity, Fig. 13 shows examples of FP for method M2, where blood vessel (red circle) is often confused for microbleeds and where some FP (blue circles) have been labelled as possible microbleeds by our expert. Indeed, when our two ex-

¹⁰ See supplementary Fig. 10

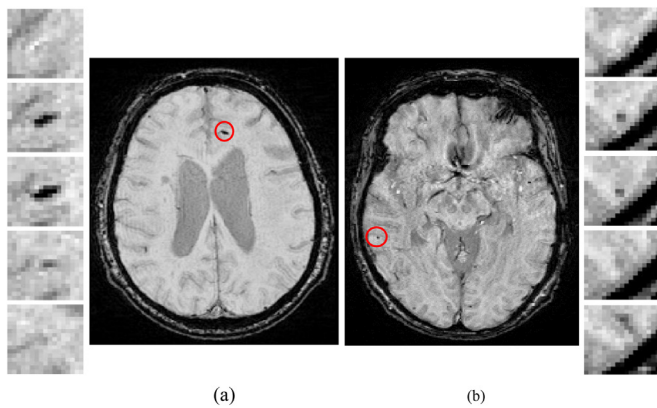


Fig. 12. Examples of extreme case of detected real lesions from our dataset. The top five panels above each scan represent consecutive axial section of a large dark rCMB with irregular shape (a) and small pale rCMB with a spherical shape (b).

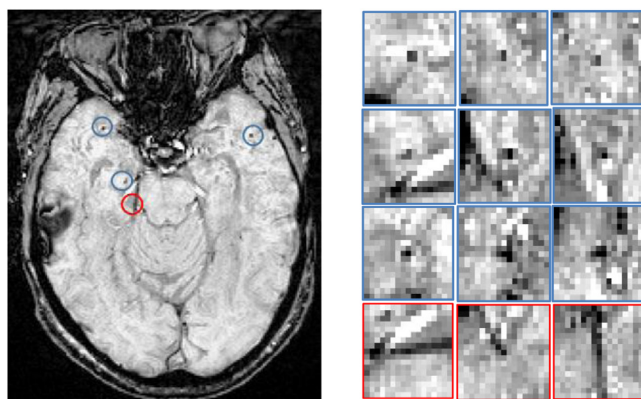


Fig. 13. Some examples of FP for models M2, using a threshold with 95% sensitivity. Blue panels show possible CMB and the red panel represents a vessel cross section along the three views (axial, coronal, and sagittal).

perts labelled lesions, they agreed on 83% of cases, increasing to ~90% after reaching consensus.

An alternative approach could have been to use generative models to create extra data, such as Generative Adversarial Networks (GAN) [35]. We did not investigate those because of the small size of our dataset and the fact that background would have needed to be generated as well. For example, some ranges of lesion volume do not exist in our ground truth data, which would thus have not been generated. In contrast, by controlling our parameters we could generate synthetic data outside the ground truth distribution.

Too small training dataset is often a limitation for standard data augmentation method. By comparing 10-fold and 2-fold cross validation, we showed that with a small dataset, SMOTE does to generate synthetic samples that cover the full range of variation present in the half dataset used for testing. Indeed, if the performance degrades between 2-fold and 10-fold (or even leave one out in the extreme case), one can assume that half the data size is not large enough to properly cover all the cases, and that using 10-fold cross validation provides overly optimistic performance. This issue highlights that any cross-validation design will favour data augmentation methods using existing ground truth, compared to methods like ours that are almost independent on the ground truth. It is thus remarkable that our proposed technique could achieve better performance when compared to data augmentation methods.

Our classifier used only 9,321 parameters and its performance compared favorably with the state-of-art. Only 3 published methods had better results [16],[24],[9]. One method [24] used only 10

subjects with a classifier comprising more than 45,000 parameters, achieving 69 FP per scan. The top two methods [9] and [16] described complex classifiers (both using CNN with around 71,794 and 1,215,536 parameters respectively). It is well known that a small training set for a model with a large number of parameters makes the trained model susceptible to overfitting. It thus remains to be seen whether those methods can generalize.

Another difference with the best method [9] is the use of 2 MRI sequences (QSM and SWI), which take into account the phase information. We have started investigating the generation of synthetic CMB for MRI QSM with promising preliminary results [36]. More complex models including phase and susceptibility point spread function would likely improve the realism of synthetic CMB.

It is also difficult to compare performance because of the different diseases considered: dementia in our case vs stroke [16],[9], hemodialysis, and traumatic brain injury [9]. Ideally, each method should be tested on various diseases and MRI acquisitions. However, when using a re-trained random forest method [13], we could achieve less than 10 FP per scan, a results close to state of the art, despite using a simpler classifier.

Our proposed method has some limitations. Our synthetic lesion model is simple (Gaussian profile). Better models could include more shape and intensity variation, as well as physics-based features (e.g. using a magnetic dipole for simulating phase effect). Some of the parameters that we have used (location, intensity, etc.) might be specific to our dataset, MRI acquisition, or SWI processing, and we have used only one study for which we had access to the ground truth. More work using diverse studies with a more varied population and acquisition range is needed to confirm our findings. We expect that by sharing our data we will encourage other studies to do the same, allowing the community to demonstrate translational potential.

In conclusion, synthetic microbleeds generation is a powerful data augmentation approach and should be considered for training automated lesion detection system from MRI SWI.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.cmpb.2021.106127](https://doi.org/10.1016/j.cmpb.2021.106127).

Appendix

Synthetic CMB generation

CMB are mostly spherical hypointense small lesions and were modelled using a 3D Gaussian function in high resolution.

$$G(x, y, z) = Ae^{-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2} - \frac{(z-z_0)^2}{2\sigma_z^2}} \quad (1)$$

where coefficient A is the amplitude, σ_x , σ_y , and σ_z are the standard deviation in the x , y , z directions respectively, and x_0 , y_0 , and z_0 are the center of the lesion within the patch. We assumed that the sCMB can be centered uniformly at low resolution, and therefore the center was randomly located within ± 5 pixels from the middle of the patch in high resolution (resulting in random location inside the center pixel in low resolution). The volume of the sCMB was computed after thresholding at half the Gaussian maximum. To simulate shape variation with a constant desired volume around that of perfect sphere, uniform random variation for two of the standard deviations were applied. The third one was computed so that the product of the 3 (at half maximum) equaled the volume of a sphere of radius r . That way, we could specify a target volume and generate a randomly shaped blob with that volume when thresholded at half maximum. Given a sphere volume V , the

corresponding standard deviation of an isotropic Gaussian is:

$$\sigma_t = 3\sqrt{\frac{3 \times V}{4 \times \pi}} \times \frac{1}{1.175} \quad (2)$$

where σ_t is the standard deviation in mm and V is the volume in mm³.

$$\sigma_x = \sigma_t \times (r_{\max} - r_{\min}) \times \text{rand}(1) + r_{\min} \quad (3)$$

$$\sigma_y = \sigma_t \times (r_{\max} - r_{\min}) \times \text{rand}(1) + r_{\min} \quad (4)$$

The σ_x and σ_y are the spread of Gaussian in x and y directions respectively and r_{\min} and r_{\max} are the defined maximum and minimum range for getting random standard deviation specifying the sphericalness of the Gaussian. To cover a wide range of variation, we chose the r_{\min} and r_{\max} from 0.5 to 0.8 and from 0.8 to 1.2. Once σ_x and σ_y were defined as above, the third standard deviation of the Gaussian was defined as:

$$\sigma_z = \frac{V\sigma_x\sigma_y}{\left(\frac{4}{3}\pi\right)1.175^3} \quad (5)$$

In order to vary orientation, we added a uniformly random rotation around the three axis of G between 20° and 90°. All the parameters discussed above were optimized using exhaustive grid search to maximize the normalized cross correlation (NCC) across average of all the rCMB and sCMB:

$$NCC = \frac{1}{n} \times \left(\frac{1}{\sigma_{P_r} \times \sigma_{P_s}} \sum_{x,y} (P_r - \bar{P}_r) \times (P_s - \bar{P}) \right) \quad (6)$$

where P_r and P_s are the average of all real and synthetic $7 \times 7 \times 7$ patches respectively, n is the number of voxels in patches, σ_{P_r} and σ_{P_s} are the standard deviation of P_r and P_s . The best parameters were obtained with rotation of 30°, r_{\min} and r_{\max} as 0.5 and 0.9 for Gaussian sphericalness.

Performance Metrics

The used performance metrics to evaluate and compare different augmentation models for CMB classification (method section) are defined as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Specificity} = \frac{TN}{FP + TN} \quad (8)$$

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \quad (9)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (10)$$

Where TP, TN, FP, and FN are true positive, true negative, false positive and false negative respectively.

References

- [1] A. Charidimou, D.J. Werring, Cerebral microbleeds: detection, mechanisms and clinical challenges, *Future Neurol* 6 (5) (Aug. 2011) 587–611, doi:[10.2217/fnl.11.42](#).
- [2] C. Cordonnier, R. Al-Shahi Salman, J. Wardlaw, Spontaneous brain microbleeds: systematic review, subgroup analyses and standards for study design and reporting, *Brain* 130 (8) (Aug. 2007) 1988–2003, doi:[10.1093/brain/awl387](#).
- [3] K.A. Knudsen, J. Rosand, D. Karluk, S.M. Greenberg, Clinical diagnosis of cerebral amyloid angiopathy: validation of the Boston criteria, *Neurology* 56 (4) (Feb. 2001) 537–539.
- [4] J.-H. Heo, D.-G. Im, S.-H. Lee, J.-Y. Ahn, The clinical significance of brain microbleeds in patients with Alzheimer's disease: preliminary study, *Ann. Indian Acad. Neurol.* 19 (4) (2016) 495–498, doi:[10.4103/0972-2327.194426](#).
- [5] E.M. Haacke, S. Mittal, Z. Wu, J. Neelavalli, Y.-C.N. Cheng, Susceptibility-weighted imaging: technical aspects and clinical applications, part 1, *AJNR Am. J. Neuroradiol.* 30 (1) (Jan. 2009) 19–30, doi:[10.3174/ajnr.A1400](#).
- [6] S.M. Gregoire, et al., The Microbleed Anatomical Rating Scale (MARS), *Neurology* 73 (21) (Nov. 2009) 1759, doi:[10.1212/WNL.0b013e3181c34a7d](#).
- [7] C. Cordonnier, et al., Improving Interrater agreement about brain microbleeds: development of the Brain Observer MicroBleed Scale (BOMBS), *Stroke* 40 (1) (Jan. 2009) 94–99, doi:[10.1161/STROKEAHA.108.526996](#).
- [8] S. Haller, M.W. Vernooij, J.P.A. Kuijfer, E.-M. Larsson, H.R. Jäger, F. Barkhof, Cerebral microbleeds: imaging and clinical significance, *Radiology* 287 (1) (Apr. 2018) 11–28, doi:[10.1148/radiol.2018170803](#).
- [9] S. Liu, et al., Cerebral microbleed detection using susceptibility weighted Imaging and deep learning, *NeuroImage* 198 (Sep. 2019) 271–282, doi:[10.1016/j.neuroimage.2019.05.046](#).
- [10] H.J. Kuijfer, J. de Bresser, G.J. Biessels, M.A. Viergever, K.L. Vincken, Detecting cerebral microbleeds in 7.0 T MR images using the radial symmetry transform, in: 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Mar. 2011, pp. 758–761, doi:[10.1109/ISBI.2011.5872516](#).
- [11] G. Loy, A. Zelinsky, Fast radial symmetry for detecting points of interest, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (8) (Aug. 2003) 959–973, doi:[10.1109/TPAMI.2003.1217601](#).
- [12] W. Bian, C.P. Hess, S.M. Chang, S.J. Nelson, J.M. Lupo, Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images, *NeuroImage Clin* 2 (2013) 282–290, doi:[10.1016/j.nicl.2013.01.012](#).
- [13] A. Fazlollahi, et al., Computer-aided detection of cerebral microbleeds in susceptibility-weighted imaging, *Comput. Med. Imaging Graph.* 46 (Part 3) (Dec. 2015) 269–276, doi:[10.1016/j.compmedimag.2015.10.001](#).
- [14] Y.-D. Zhang, et al., Voxelwise detection of cerebral microbleed in CADASIL patients by leaky rectified linear unit and early stopping, *Multimed. Tools Appl.* (Jan. 2017), doi:[10.1007/s11042-017-4383-9](#).
- [15] S.R. Barnes, E.M. Haacke, M. Ayaz, A.S. Boikov, W. Kirsch, D. Kido, Semiautomated detection of cerebral microbleeds in magnetic resonance images., semi-automated detection of cerebral microbleeds in magnetic resonance images, *Magn. Reson. Imaging Magn. Reson. Imaging* 29 (6) (Jul. 2011) 844–852 29, no.6, pp. 844 [10.1016/j.mri.2011.02.028](#) . [10.1016/j.mri.2011.02.028](#).
- [16] Q. Dou, et al., Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, *IEEE Trans. Med. Imaging* 35 (5) (May 2016) 1182–1195, doi:[10.1109/TMI.2016.2528129](#).
- [17] S. Wang, J. Sun, I. Mehmood, C. Pan, Y. Chen, Y.-D. Zhang, Cerebral microbleeding identification based on a nine-layer convolutional neural network with stochastic pooling, *Concurr. Comput. Pract. Exp.* 0 (0) (2019) e5130, doi:[10.1002/cpe.5130](#).
- [18] S. Wang, Y. Jiang, X. Hou, H. Cheng, S. Du, Cerebral micro-bleed detection based on the convolution neural network with rank based average pooling, *IEEE Access* 5 (2017) 16576–16583, doi:[10.1109/ACCESS.2017.2736558](#).
- [19] Y.D. Zhang, X.X. Hou, Y.D. Lv, H. Chen, Y. Zhang, S.H. Wang, Sparse autoencoder based deep neural network for voxelwise detection of cerebral microbleed, in: 2016 IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS), Dec. 2016, pp. 1229–1232, doi:[10.1109/ICPADS.2016.0166](#).
- [20] Y. Chen, J.E. Villanueva-Meyer, M.A. Morrison, J.M. Lupo, Toward automatic detection of radiation-induced cerebral microbleeds using a 3D deep residual network, *J. Digit. Imaging* 32 (5) (Oct. 2019) 766–772, doi:[10.1007/s10278-018-0146-z](#).
- [21] S. Lu, K. Xia, S.-H. Wang, Diagnosis of cerebral microbleed via VGG and extreme learning machine trained by Gaussian map bat algorithm, *J. Ambient Intell. Humaniz. Comput.* (Feb. 2020), doi:[10.1007/s12652-020-01789-3](#).
- [22] J. Hong, H. Cheng, Y.-D. Zhang, J. Liu, Detecting cerebral microbleeds with transfer learning, *Mach. Vis. Appl.* 30 (7–8) (Oct. 2019) 1123–1133, doi:[10.1007/s00138-019-01029-5](#).
- [23] Y.-D. Zhang, Y. Zhang, X.-X. Hou, H. Chen, S.-H. Wang, Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed, *Multimed. Tools Appl.* (Mar. 2017), doi:[10.1007/s11042-017-4554-8](#).
- [24] J. Hong, S.-H. Wang, H. Cheng, J. Liu, Classification of cerebral microbleeds based on fully-optimized convolutional neural network, *Multimed. Tools Appl.* (Nov. 2018), doi:[10.1007/s11042-018-6862-z](#).
- [25] H.J. Kuijfer, et al., Efficient detection of cerebral microbleeds on 7.0T MR images using the radial symmetry transform, *NeuroImage* 59 (3) (Feb. 2012) 2266–2273, doi:[10.1016/j.neuroimage.2011.09.061](#).
- [26] H.J. Kuijfer, et al., Semi-Automated Detection of Cerebral Microbleeds on 3.0 T MR Images, *PLoS ONE* 8 (6) (Jun. 2013) e66610, doi:[10.1371/journal.pone.0066610](#).
- [27] S. Momeni et al., "Data Augmentation using synthetic lesions improves machine learning detection of microbleeds from MRI," in *Simulation and Synthesis in Medical Imaging*, 2018, pp. 12–19.
- [28] K.A. Ellis, et al., The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease, *Int. Psychogeriatr.* 21 (4) (Aug. 2009) 672–687, doi:[10.1017/S1041610209009405](#).
- [29] N.J. Tustison, et al., N4ITK: Improved N3 Bias Correction, *IEEE Trans. Med. Imaging* 29 (6) (Jun. 2010) 1310–1320, doi:[10.1109/TMI.2010.2046908](#).
- [30] V.L. Villemagne, et al., Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study, *Lancet Neurol* 12 (4) (Apr. 2013) 357–367, doi:[10.1016/S1474-4422\(13\)70044-9](#).
- [31] N. Yassi, et al., Influence of comorbidity of cerebrovascular disease and Amyloid- β on Alzheimer's Disease, *J. Alzheimers Dis.* 73 (3) (Jan. 2020) 897–907, doi:[10.3233/JAD-191028](#).
- [32] K. Van Leemput, F. Maes, D. Vandermeulen, P. Suetens, Automated model-based tissue classification of MR images of the brain, *IEEE Trans. Med. Imaging* 18 (10) (Oct. 1999) 897–908, doi:[10.1109/42.811270](#).

- [33] S.M. Smith, Fast robust automated brain extraction, *Hum. Brain Mapp* 17 (3) (Nov. 2002) 143–155, doi:[10.1002/hbm.10062](https://doi.org/10.1002/hbm.10062).
- [34] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, *J. Artif. Intell. Res.* 16 (Jun. 2002) 321–357, doi:[10.1613/jair.953](https://doi.org/10.1613/jair.953).
- [35] I. J. Goodfellow et al., “Generative adversarial networks,” *ArXiv14062661 Cs Stat*, Jun. 2014, Accessed: Jun. 04, 2020. [Online]. Available: <http://arxiv.org/abs/1406.2661>.
- [36] S. Momeni, et al., *Synthetic CMB generation for Training classifiers on QSM images*, ISMRM, Virtual, 2020 August.