NeuroImage 198 (2019) 271–282

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Cerebral microbleed detection using Susceptibility Weighted Imaging and deep learning

Saifeng Liu [a,*], David Utriainen [a,b], Chao Chai [c], Yongsheng Chen [a,d], Lin Wang [a], Sean K. Sethi [a,b], Shuang Xia [c], E. Mark Haacke [a,b,d]

[a] The MRI Institute for Biomedical Research, Bingham Farms, MI, United States
[b] Magnetic Resonance Innovations, Bingham Farms, MI, United States
[c] Department of Radiology, Tianjin First Central Hospital, Tianjin, China
[d] Department of Radiology, Wayne State University, Detroit, MI, United States

ARTICLE INFO

Keywords:
Cerebral microbleeds
Computer aided detection
Deep learning
Convolutional neural networks
Susceptibility weighted imaging
Quantitative susceptibility mapping

ABSTRACT

Detecting cerebral microbleeds (CMBs) is important in diagnosing a variety of diseases including dementia, stroke and traumatic brain injury. However, manual detection of CMBs can be time-consuming and prone to errors, whereas the current automatic algorithms for CMB detection are usually limited by large number of false positives. In this study, we present a two-stage CMB detection framework which contains a candidate detection stage based on a 3D fast radial symmetry transform of the composite images from Susceptibility Weighted Imaging (SWI), and a false positive reduction stage based on deep residual neural networks using both the SWI and the high-pass filtered phase images. While the SWI images provide exquisite sensitivity to the presence of blood products, the high-pass filtered phase images enable the differentiation of diamagnetic calcifications from paramagnetic microbleeds. The deep learning model was trained using 154 data sets, and the best models were selected using 25 validation data sets. Finally, the models were tested using 41 cases, including 13 hemodialysis cases, 9 traumatic brain injury cases, 9 stroke cases and 10 healthy controls. Using 3D SWI and high-pass filtered phase images as input, the best model led to a sensitivity of 95.8%, a precision of 70.9%, and 1.6 false positives per case. This model achieved similar performance to the most experienced human rater and outperformed recently reported CMB detection methods. This study demonstrates the potential of applying deep learning techniques to medical imaging for improving efficiency and accuracy in diagnosis.

1. Introduction

Cerebral microbleeds (CMBs) are small foci of chronic blood products that are frequently observed in patients affected by Alzheimer's disease, stroke and traumatic brain injury (TBI) (Greenberg et al., 2009; Yates et al., 2014). While the location of CMBs reflects the etiology, the number of CMBs could indicate the risk of future intracerebral hemorrhage (ICH) and cognitive impairment (Fan et al., 2003; Poels et al., 2012). Deep CMBs in the basal ganglia or thalamus are typically related to hypertension, whereas the presence of lobar CMBs may suggest cerebral amyloid angiopathy (Poels Mariëlle et al., 2010). Detecting CMBs can be clinically significant to weigh the benefits and risks in antithrombotic treatment planning for stroke patients (Greenberg et al., 2009). Finally, CMB detection is valuable for TBI diagnosis and prognosis. It is estimated that 50 million people are affected by TBI worldwide every year (Maas et al., 2017), and over 75% of these cases are mild TBI cases (McMahon et al., 2014). The high prevalence of CMBs in TBI makes it a valuable marker for monitoring the progress of the disease (Trifan et al., 2017). Thus, accurate and reliable detection of CMBs is indispensable.

The detection of CMBs is usually performed by using magnetic resonance imaging (MRI) techniques. Susceptibility weighted imaging (SWI) is a widely used MRI data post-processing technique for detecting CMBs thanks to its exquisite sensitivity to the susceptibility effects induced by blood products (Haacke et al., 2004). In SWI, susceptibility weighting masks are generated from high-pass filtered phase images and multiplied with the magnitude images to form composite images (denoted as “SWI images” in this paper). CMBs appear as small spherical or oval regions with low intensity on SWI images (Liu et al., 2017a). However, there are various types of CMB mimics, such as veins, iron deposition in the basal ganglia, calcifications and signal void due to poor flow compensation or cusp artifacts caused by failures in coil combination (Liu et al., 2017a;

* Corresponding author. 30200 Telegraph Road, Ste. 104, Bingham Farms, MI, 48025, United States.
E-mail address: liusaifeng@gmail.com (S. Liu).

https://doi.org/10.1016/j.neuroimage.2019.05.046
Received 3 January 2019; Received in revised form 15 May 2019; Accepted 17 May 2019
Available online 20 May 2019

Abbreviations:

AdamWR	adaptive moment estimation with warm restart
ANTs	advanced normalization tools
AUC	area under the receiver operating characteristic curve
AUC-PR	area under the precision-recall curve
BatchNorm	batch normalization
BET	brain extraction tool
CMB	cerebral microbleed
CNN	convolutional neural network
FPavg	average number of false positives per case
FP/cmb	average number of false positives per CMB
FRST	fast radial symmetry transform
ICH	intracerebral hemorrhage
MEDI	morphology enabled dipole inversion
MRI	magnetic resonance imaging
QSM	quantitative susceptibility mapping
ReLU	rectified linear unit
SHARP	sophisticated harmonic artifact reduction for phase data
SWI	susceptibility weighted imaging
TBI	traumatic brain injury
TTA	test time augmentation

Haacke et al., 2015). As a result, manual detection of CMBs can be time-consuming and prone to errors. This can be alleviated by using automatic CMB detection algorithms.

Usually a two-stage process is used for automatic CMB detection, with a candidate detection stage and another classification stage for false positive reduction (Barnes et al., 2011; Bian et al., 2013; Fazlollahi et al., 2014; van den Heuvel et al., 2016; Dou et al., 2016; Chen et al., 2018). Bian et al. applied a 2D radial symmetry transform on the minimum intensity projections of SWI images and used shape features for false positive reduction, which led to a sensitivity of 86.5% with 44.9 false positives per subject (Bian et al., 2013); Fazlollahi et al. developed a random forest model and reported a sensitivity of 92.0% with 16.8 false positives per subject (Fazlollahi et al., 2014); van den Heuvel also developed a random forest model with shape features for detecting bleeds in TBI and achieved a sensitivity of 89.1% with 25.9 false positives per subject (van den Heuvel et al., 2016). Although high sensitivity can be easily achieved in the first stage, the performance of the second stage is usually poor, leading to many false positives and low precision. For example, Barnes et al. developed a machine learning model based on the support vector machine algorithm with shape and intensity features, but had to perform a manual false positive removal (Barnes et al., 2011). The main drawback of conventional machine learning models is related to the requirement of feature engineering. Because of the variation of the shape and intensities of CMBs on SWI images, designing effective and robust features can be quite challenging. This can be overcome by using convolutional neural network (CNN) based deep learning techniques, which have significantly improved the performance of object detection and revolutionized the computer vision field (Krizhevsky et al., 2012; LeCun et al., 2015; He et al., 2016a). CNN based techniques have been applied to medical imaging, such as cancer lesion classification (Liu et al., 2017b). It has also been used for CMB detection. Dou et al. developed a CNN based model using SWI images and achieved a sensitivity of 93.2%, a precision of 44.3% with 2.7 false positives per subject (Dou et al., 2016). More recently, Chen et al. designed a more complex residual network using SWI images acquired at 7T and reported a sensitivity of 94.7%, a precision of 72.0% with 11.6 false positives per subject (Chen et al., 2018).

The low precision or the large number of false positives of these methods is partly caused by relying on SWI images alone. Particularly, differentiating blood products from calcifications is difficult, since both of them appear dark on SWI images (Haacke et al., 2004; Liu et al., 2017a). In addition, the contrast in the SWI data can be dependent on imaging parameters, especially on main field strength, echo time and image resolution. These problems can be solved by using both magnitude and phase images with normalization (Wu et al., 2009; Fatemi-Ardekani et al., 2009). Using phase images, or quantitative susceptibility maps (QSM) reconstructed from phase images, diamagnetic materials such as calcification can be separated from paramagnetic blood products based on the sign of the phase or susceptibility. Moreover, because of the much higher susceptibility of the CMBs than the surrounding tissue or veins, phase and QSM may provide valuable features for separating CMBs from their mimics.

In this study, we present a CNN based deep learning model for automatic CMB detection, using both SWI and phase images. CMB candidates are selected by using a 3D fast radial symmetry transform (Loy et al., 2002) and then the false positives are reduced by using a deep residual neural network (He et al., 2016a). We demonstrate that with data pre-processing and augmentation, the model provided high sensitivity with a small number of false positives, outperforming both human raters and the single-channel models reported in earlier studies.

2. Materials and methods

2.1. Data acquisition and data splitting

The data included in this retrospective study are from previous studies which were approved by the appropriate institutional review boards, and informed consent was obtained prior to the scan of each subject (Pacurar et al., 2016). Specifically, a total of 220 data sets were used in this study, including 100 hemodialysis cases, 97 TBI cases, 13 stroke cases, and 10 healthy controls. Data were acquired on 1.5T Essenza, 3T Verio or 3T Trio Siemens scanners (Siemens Healthineers, Erlangen, Germany), using single-echo gradient-echo sequences with full flow compensation (Haacke et al., 2004, 2009; Liu et al., 2017a). The imaging parameters of the 1.5 T data were: echo time (TE): 40 ms; repetition time (TR): 49 or 50 ms; bandwidth/pixel (BW/pixel): 80 Hz/pixel; flip angle (FA): 15°; in-plane resolution: 0.45×0.57 to $0.53 \times 1.05\ mm^2$; slice thickness: 2–2.65 mm; matrix size: 512×304 to 512×448; and number of slices: 56 or 60. The imaging parameters of the 3T data were: TE: 17.5–20 ms; TR: 27–34 ms; BW/pixel: 100–425 Hz/pixel; FA: 12° or 15°; in-plane resolution: 0.50×0.50 to $0.54 \times 1.07\ mm^2$; slice thickness: 1.2 or 2 mm; matrix size: 448×322 to 512×416; and number of slices: 56 to 128.

The data were split into training, validation and test sets, as shown in Table 1. For the training set, 100 data sets were acquired at 3T, 54 at 1.5T; for the validation set, 15 data sets were acquired at 3T, 10 at 1.5T; for the test set, 28 data sets were acquired at 3T, 13 at 1.5T. The test set was initially designed to have 10 subjects in each category. However, two cases had to be removed, due to the large variation in raters' labels caused by severe motion artifacts in a TBI case and macro-bleeding in a stroke case. The hemodialysis cases were acquired on 10 different subjects, with 3 subjects scanned twice. For those 3 subjects, the follow-up scans were more than a month apart from the initial scans, and hence

Table 1
Subjects information and data splitting. The total numbers of CMBs in different categories are shown in the parenthesis.

Subject Category	Hemodialysis	TBI	Stroke	Normal Control	Total
Age (years)	45 to 91, mean 68	4 to 73, mean 21	46 to 72, mean 59	22 to 63, mean 30	4 to 91, mean 42.6
Training	72 (756)	80 (522)	2 (2)	0	154 (1280)
Validation	15 (160)	8 (33)	2 (0)	0	25 (193)
Test	13 (95)	9 (14)	9 (59)	10 (0)	41 (168)

were treated as independent test cases. Additionally, the majority of stroke cases were included in the test set, to check the generalization ability of the model. Normal controls were included in the test set only. The remaining data were split into training and validation sets, with the proportion of TBI cases in the validation set being around 30%, to approximate the composition of the disease types in the test set. For training the deep learning model, CMBs with elongated shape (e.g. damaged veins in TBI cases) were randomly dropped to match the number of CMBs with spherical shape, while false samples (with label 0) were selected from the following categories: 1) CMB mimics with low intensities on the original magnitude and/or SWI, or high intensities on the QSM images; 2) false positives selected by the 3D-FRST algorithm; and 3) randomly selected background regions. In the end, a total of 3072 samples were included in the training data with 1280 true samples, and 480 samples in the validation data with 193 true samples.

2.2. Data pre-processing and normalization

Data pre-processing and normalization was performed to reduce the variation caused by the imaging parameters. First, the bias-field effects in magnitude images were corrected using the N4 algorithm in Advanced Normalization Tools (ANTs) (Tustison et al., 2010). Then the corrected magnitude images were registered to the MNI-152 template (1 mm isotropic resolution) using a non-rigid registration algorithm in ANTs (Mazziotta et al., 1995; Avants et al., 2008). The use of the template is for locating the globus pallidus to avoid detecting the iron deposition, as will be described in section 2.4. The phase images were processed using homodyne high-pass filtering with a k-space window size of 96 × 96. For generating SWI, the phase images were first scaled by (20 ms/TE) ·(3T/B_0), where TE and B_0 are the echo time and main field strength, respectively. Then SWI data were generated using the normalized phase images following the conventional procedure (Haacke et al., 2004, 2009; Liu et al., 2017a). This led to the same effective susceptibility weighting for data collected at different echo times and field strengths. The susceptibility maps were reconstructed using the following steps. First, brain masks were generated using the Brain Extraction Tool (BET) in FSL (Smith, 2002). Next, the remnant background field and aliasing artifacts in the high-pass filtered phase images were reduced using Laplacian unwrapping (Haacke et al., 2015) and the Sophisticated Harmonic Artifact Reduction for Phase data (SHARP) algorithm (Schweser et al., 2011). Finally, susceptibility maps were created using the morphology enabled dipole inversion (MEDI) algorithm with the regularization parameter λ being 300 (Liu et al., 2012).

The magnitude, phase, SWI and QSM images were interpolated to 0.5 mm isotropic resolution, the highest in-plane resolution of the data included in this study. For both the magnitude and SWI images, the baseline intensity of the images (M_b) was estimated from the regions inside the brain where the susceptibility weighting mask equals 1. Then the intensities of the magnitude and SWI images (M) were normalized via:

$$\tilde{M} = \frac{M - M_b}{M_b} \quad [1]$$

The high-pass filtered phase images (P) were normalized via:

$$\tilde{P} = \frac{P \cdot 3T \cdot 20ms}{\pi \cdot B_0 \cdot TE} \quad [2]$$

The QSM images (Q) were normalized via:

$$\tilde{Q} = \frac{Q}{0.4ppm} \quad [3]$$

In the end, the value range of the pixel intensities on the normalized images was truncated to [-1, 1]. After normalization, the intensity of the background white matter region was close to 0 on all types of images, while the absolute value of the intensity of a CMB was close to 1. All the data pre-processing steps were performed using MATLAB v2016b (MathWorks, Natick, MA).

2.3. Gold standard

The CMB labeling was performed on the pre-processed magnitude, phase, SWI and QSM images with SPIN software (SpinTech, Bingham Farms, USA), according to the criteria listed in Greenberg et al. (2009). The hemorrhagic lesions in TBI cases were identified based on their susceptibility effects, shape and anatomical location (Tong et al., 2003). For the training and validation data, the CMBs were labeled by a radiologist and a SWI data processor with eight years of experience. For the test data, the CMBs were labeled by a radiologist and two SWI data processors. The subjects included in the test data were randomly shuffled, and then provided to the three raters without showing the patient type. In addition to the conventional bleeds, the medullary vein damage seen in TBI cases was labeled and treated as a CMB. For a damaged vein, the raters labeled one end of the vein. Finally, the labels were further reviewed by two experts in SWI, using the automatic CMB detection results as a reference to avoid missing any CMB. For any discrepancies in the labels, the raters reviewed the images together, and the gold standard was obtained based on the consensus of all the raters.

2.4. CMB candidate detection with a 3D fast radial symmetry transform (3D-FRST)

Recognizing that CMBs are sparsely distributed in the SWI images and the shape of most CMBs is close to spherical in nature, due to the blooming pattern induced by the magnetic field dipole effects from the deposited iron, the radial symmetry transform appears to be the ideal algorithm for detecting CMB candidates. In fact, the effectiveness of this algorithm has already been demonstrated in earlier studies on CMB detection (Bian et al., 2013; Chen et al., 2018). Here we adopted the 3D version of the FRST algorithm (Kuijf et al., 2012), with the scale parameter N set to a range of 1–4 pixels, and the symmetry strictness parameter α set to 2, as suggested in Loy and Zelinsky (Loy et al., 2002).

The CMB candidates were generated using the following steps. First, an initial candidate map was generated by thresholding the 3D-FRST transformed SWI images. A threshold of th_{rst} was used in regions outside the globus pallidus, and $2th_{rst}$ inside the globus pallidus to avoid detecting increased levels of iron deposition. The localization of the globus pallidus is based on the Harvard-Oxford atlas provided in FSL (Desikan et al., 2006). The sensitivity of CMB candidate detection is dependent on th_{rst}. In order to determine the optimal value, th_{rst} was varied from 0.08 to 0.15 with a step size of 0.01. The optimal th_{rst} was chosen based on the sensitivity and the average number of false positives in the training and validation data.

Because of the noise and variation in the shape of CMBs, there can be lots of isolated pixels which may lead to unnecessary calculations in the false positive reduction stage. Hence, two isolated pixels with a distance less than 4 pixels (~2 mm) were connected, if no edge was detected on the straight line connecting them. Here, an edge is defined as regions with gradient greater than 80% of the gradients of all the regions within the brain. Next, for each connected region in the candidate map, either one or two candidates were generated, depending on the shape of the region. For a spherical region, such as a regular CMB, only one candidate was generated, located at the center of that region; for an elongated region, such as a damaged vein, two candidates were generated, located at the two ends along the major axis of that region. This is to provide sufficient representation of CMBs with different shapes. Hence, a CMB candidate in this study may correspond to either the center or one end of a potential CMB. For shape analysis, we calculated the "fractional anisotropy" (FA) of each connected region, using the same formula as the one in diffusion tensor imaging (Bihan et al., 2001):

$$FA = \sqrt{\frac{3}{2}}\sqrt{\frac{(\lambda_x - \bar{\lambda})^2 + (\lambda_y - \bar{\lambda})^2 + (\lambda_z - \bar{\lambda})^2}{\lambda_x^2 + \lambda_y^2 + \lambda_z^2}} \quad [4]$$

where λ_x, λ_y and λ_z are the eigenvalues of the connected region in x, y, and z directions, and $\bar{\lambda}$ is the mean eigenvalue. These eigenvalues represent the shape of the connected region not the usual diffusion properties. FA is 0 for a sphere and 1 for a line. In this study, an elongated region was defined empirically as a region with major axis length greater than 6 mm and FA greater than 0.85 (such that a cuboid with aspect ratio 3:2:1 was considered to be sufficiently spherical). Finally, 3D volumes with matrix size $32 \times 32 \times 32$ were cropped from magnitude, phase, SWI and QSM images, with the CMB candidates located at the center. These 3D volumes are to be classified as true CMBs or false positives by the deep learning model.

2.5. False positive reduction through deep learning

The architecture of the deep learning model is illustrated in Fig. 1. The residual blocks were implemented in the full pre-activation fashion, with the rectified linear unit (ReLU) and batch normalization (BatchNorm) layer placed before the 3D convolution layers, as proposed by He et al. (2016b). The kernel size of all the 3D convolution layers was $3 \times 3 \times 3$. A $1 \times 1 \times 1$ convolution was used to form the residual connections, when the output size of the layer was changed. A stride of 2 was used for those convolutions (including the $1 \times 1 \times 1$ convolutions) when the spatial dimension was reduced, and a stride of 1 was used for the other convolutions. After the residual blocks, global average pooling was performed, followed by two fully-connected (FC) layers and a softmax layer. We tested different combinations of input channels, including single channel: magnitude (M), phase (P), SWI (S), QSM (Q); two-channel: MP, MQ, PS, SQ; three-channel: MPS, MPQ, MSQ, PSQ; and four-channel: MPSQ. This can also be viewed as a feature selection step. Except for the input layer, the sizes of all the layers were the same for different types of input, so were the numbers of trainable parameters. In addition to the models with input scale $32 \times 32 \times 32$ shown in Fig. 1, we trained models using phase and SWI images with an input scale of $16 \times 16 \times 16$. For these models, the output size of the first convolution layer was $16^3 \times 16$, the first average pooling layer was dropped, while the output sizes of all the other layers were the same as those shown in Fig. 1. This makes it possible to investigate the contribution of individual channel, the effectiveness of channel combinations and the impact of input scales.

2.6. Training the deep learning model

The training of the deep learning model was performed using Keras v2.1.3, with Tensorflow v1.8.0 as the backend, on a desktop equipped with 24 GB RAM and a GTX 1080ti GPU (Nvidia, Santa Clara, CA). The performance of the model was monitored using the area under the receiver operating characteristic curve (AUC) evaluated on the validation data. For each type of input, the training was performed twice and four models were selected based on their validation AUCs. In each training process, the model weights were initialized randomly, and then updated using the Adam with warm restart (AdamWR) optimizer (Loshchilov and Hutter, 2017). The learning rate was set to 0.002, weight decay 0.003, and batch size 256. Cosine annealing was used, with an initial warm-up cycles of 4 epochs, followed by two full training cycles with 16 epochs each. Both the learning rate and weight decay were reduced to 0.8 of the original values in the last cycle. In order to avoid over-fitting, online data augmentation was performed through 3D rotation, translation and random left-to-right flipping. The degrees of rotation about the three axes were sampled independently with zero-mean Gaussian distribution. The ranges of rotation were [-30°, 30°], [-30°, 30°], and [-180°, 180°], for rotations about the row, column and slice directions, respectively. The distances of translation were generated similarly, and the maximum translation along each direction was 1 pixel. An initial matrix size of $48 \times 48 \times 48$ was used for the training samples to avoid introducing too many zeros during 3D rotation. After the transformation, the samples were cropped to $32 \times 32 \times 32$ and input to the deep learning model.

2.7. Testing the two-stage CMB detection

During testing, we performed test time augmentation (TTA) in the false positive reduction step and compared the performance of the models with and without TTA. Specifically, each CMB candidate was rotated about the z-axis by 0° (i.e., the original sample), 90°, 180°, and 270°. For each rotation, the 3D volume was also flipped in the z direction. Using each model, a total of 8 predictions were made for each CMB candidate, and the mean of these predictions was used as the final prediction of that model. Furthermore, when two candidates were generated for the same elongated region (see section 2.4), the two candidates were counted as one potential CMB and their probabilities were averaged.

The performance of the CMB candidate detection and false positive reduction was evaluated using a few metrics, including sensitivity, precision, average number of false positives per case (FPavg), average number of false positives per CMB (FP/cmb), and the area under the precision-recall curve (AUC-PR). For AUC-PRs, the 95% confidence bounds were estimated through a bootstrapping approach (Fawcett, 2006). The permutation test was used to compare the AUC-PRs of the models, with family-wise error rate controlled using a maximum statistic technique (Blair and Karniski, 1993; Groppe et al., 2011), and a p value < 0.05 was considered significant. To check the model performance, two types of probability thresholds were selected: 1) high

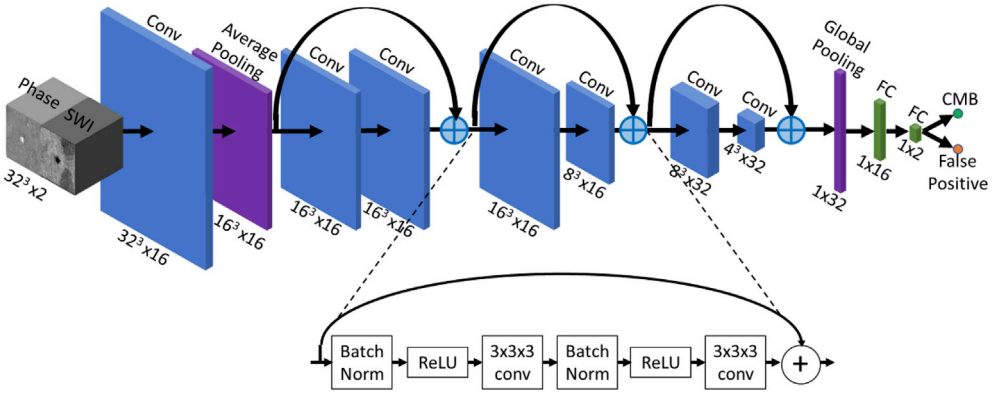

Fig. 1. Architecture of the deep learning model for CMB false positive reduction. Each convolution unit in the residual block is composed of a batch-normalization layer, a ReLU layer and a $3 \times 3 \times 3$ convolution layer. The numbers below each layer correspond to the output size of that layer. For models with different numbers of input channels, the output sizes of all the layers remain unchanged. The total number of parameters was 72082 (71794 trainable) for the model shown here.

sensitivity threshold, which led to the closest performance to the most experienced SWI data processor (rater 3); and 2) high specificity threshold, which led to no false positives in healthy controls. The SWI model was used as a reference, for which only a high sensitivity threshold was selected. For each type of input, the model with the largest AUC-PR was denoted as the "best single model". Additionally, the predictions made by models with the same type of input were averaged and denoted as the "model average".

3. Results

3.1. CMB candidate detection

Fig. 2 shows the impact of th_{rst} on the sensitivity and FPavg in the CMB candidate detection step, evaluated on the training and validation data. The optimal threshold was selected as 0.09, based on both the sensitivity and the FPavg. With this threshold, the sensitivity was 97.9%, and FPavg 312.5. On the test data, the sensitivity of CMB candidate detection was 99.4%, with FPavg 276.8.

3.2. CMB false positive reduction

Fig. 3 demonstrates the performance of the best single models with different types of input. Note that the sensitivity in the CMB candidate detection stage has already been taken into account, and the metrics reported in this section reflects the performance of the whole two-stage CMB detection framework. Among the models with three or four input channels, the one using magnitude, phase, and QSM (denoted as MPQ model, blue lines in Fig. 3a and b) had the largest AUC-PR: 0.91 [0.85, 0.94], with the lower and upper confidence bounds provided in the brackets. The same convention is used in later sections when AUC-PRs are reported. In comparison, the AUC-PR of the model using all four types of images (denoted as MPSQ model, cyan lines in Fig. 3a and b) was 0.87 [0.82, 0.91]. As shown in Fig. 3c and d, the model using the combination of phase and SWI images (denoted as PS model, solid green lines, AUC-PR: 0.92 [0.88, 0.95]) outperformed all the other types of models and achieved similar performance to the most experienced SWI data processor (red dots). Without test time augmentation (TTA), the performance of the PS model (dashed green lines in Fig. 3c and d, AUC-PR: 0.91 [0.87, 0.94]) was reduced significantly ($p = 0.013$), but was still similar to the performance of the radiologist (green dots) and the SWI data processor (blue dots). Among the models with single channel input shown in Fig. 3e and f, the QSM model performed the best (AUC-PR: 0.88 [0.82, 0.92]), followed by SWI (AUC-PR: 0.87 [0.83, 0.91]), phase (AUC-PR: 0.85 [0.79, 0.90]), and magnitude (AUC-PR: 0.83 [0.78, 0.88]), reflecting the usefulness of different types of images. However, no significant difference was found among the AUC-PRs of single-channel models. By comparing the AUC-PRs of the reference SWI model and the best single models with different numbers of input channels (i.e. the MPSQ, MPQ, PS and Q), only the PS model was significantly different from the reference SWI model ($p = 0.024$). The PS model was also significantly different from the MPSQ model ($p = 0.015$). No significant difference was observed between other pairs of models.

Fig. 4a and b show that the PS model with $32 \times 32 \times 32$ input scale performed significantly better than the one with $16 \times 16 \times 16$ input scale ($p = 0.0002$). The performance of the models was not improved much by simple averaging, as indicated in Fig. 4c and d. The relatively lower sensitivity of rater 1, compared to the other raters, was partly due to the missing of a few small lesions, as reflected in Fig. 4e and f. When small CMBs with volume $\leqq 2$ pixels were excluded (15 CMBs in total, with the size of the CMBs estimated from the thresholded 3D-FRST image), the sensitivity of rater 1 increased to 84.3%, with precision 84.3%. Meanwhile, the PS model and the MPQ model outperformed both raters 1 and 2. The other models were still worse than the raters. The performance of the raters, the PS model and the reference SWI model are provided in Table 2.

Using the SWI model, the false positive caused by calcification was not removed, as indicated by the red arrows in Fig. 5, but was eliminated successfully with the PS model. The probability for that calcification to be a true CMB was 60.0%, as predicted by the SWI model, but was reduced to 19.3% by the PS model. In the test data, a total of 3 calcifications were misclassified as CMBs by the SWI model, but were all removed by the PS model. Fig. 6 demonstrates the detection of CMBs in hemodialysis and stroke patients using the PS model. Most of the CMBs detected in these two subjects were located in the basal ganglia structures and the thalamus, indicating the possible relation to hypertension. There were 18 CMBs in the stroke patient, with one small bleed missed by all the raters but detected by the PS model, as indicated by the yellow arrows in Fig. 6e–h. This was the only lesion missed by all three raters in the test set.

Using the PS model, the sensitivity and the number of false positives as functions of the number of CMBs for each subject in the test data are shown in Fig. 7. For the stroke and hemodialysis cases, no significant difference was found between the model performance on the 1.5T and 3T data using a *t*-test. With the high precision threshold (Fig. 7a and c), the average sensitivities (mean ± standard error) were $99.1 \pm 0.9\%$ and $91.8 \pm 5.4\%$ for the 1.5T and 3T data, respectively; while the average numbers of false positives were 2.2 ± 0.4 and 1.8 ± 0.9 for the 1.5T and 3T data, respectively. Except for one hemodialysis case and another TBI case, the sensitivity was 100% when there were fewer than 5 CMBs in any given case, but reduced when there were more CMBs. While the lower sensitivity for the hemodialysis case was mainly due to the small size of the bleeds (Fig. 8a and d), the much lower sensitivity for the TBI case was due to missing two damaged veins with predicted probabilities lower

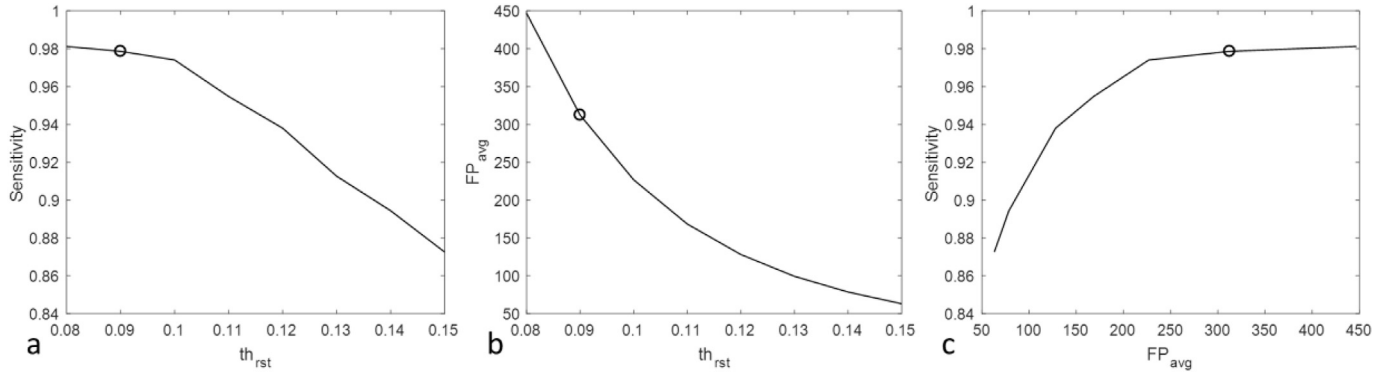

Fig. 2. Impact of the 3D-FRST threshold th_{rst} on the sensitivity and the average number of false positives per case (FPavg) of CMB candidate detection. These results were obtained using the training and validation data. (a) Sensitivity vs. th_{rst}. (b) FPavg vs. th_{rst}. (c) Sensitivity vs. FPavg. The optimal 3D-FRST threshold was chosen to be 0.09, as indicated by the circles.

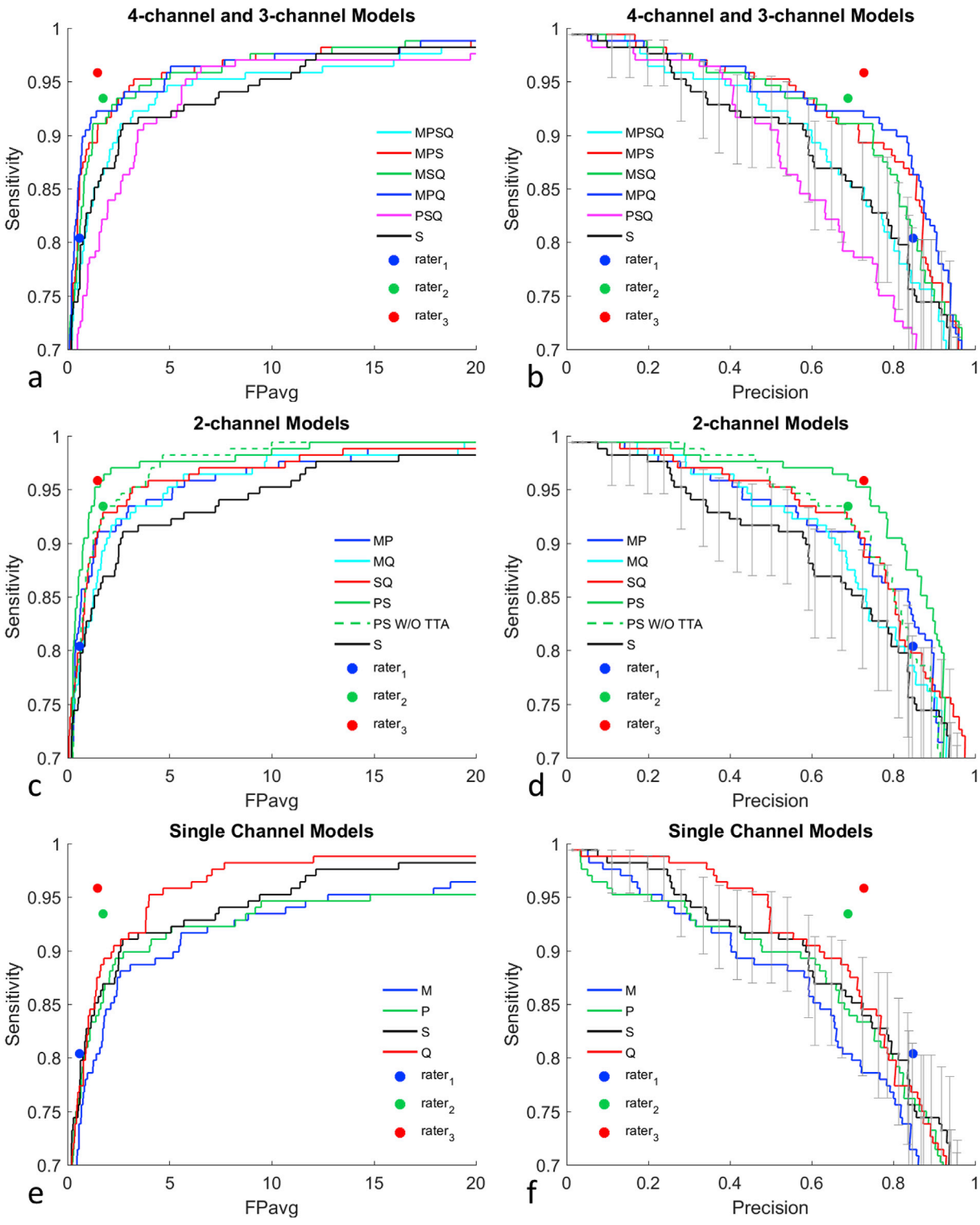

Fig. 3. Performance of the best single models on the test data. (a) and (b): models with three or four input channels. (c) and (d): models with two input channels. Without test time augmentation, the performance of the PS model (dashed green lines in c and d) was reduced. (e) and (f): models with single input channel. The plots in the first column (a, c and e) are free operating characteristic curves, while the plots in the second column (b, d and f) show the relationships between sensitivity and precision, which are essentially the transposed version of the conventional precision-recall curves. The performance of the SWI model is shown as a reference. For better visualization, the 95% confidence bounds were plotted for the SWI model only, in (b), (d) and (f). The letters M, P, S and Q represent the type of input channels. M: magnitude; P: phase; S: SWI; and Q: QSM.

than the threshold (Fig. 8b and e). For the high precision threshold, when no false positives were detected in the healthy controls, still a few false positives were detected for patients with no CMBs. The PS model detected the largest number of false positives in one stroke case (the red dot in the upper right corner in Fig. 7a–d). It was found that the false positives detected in this stroke case were mostly associated with veins with abnormally high susceptibilities, caused by reduced venous oxygen saturation (green circles in Fig. 8c and f). In fact, these asymmetric

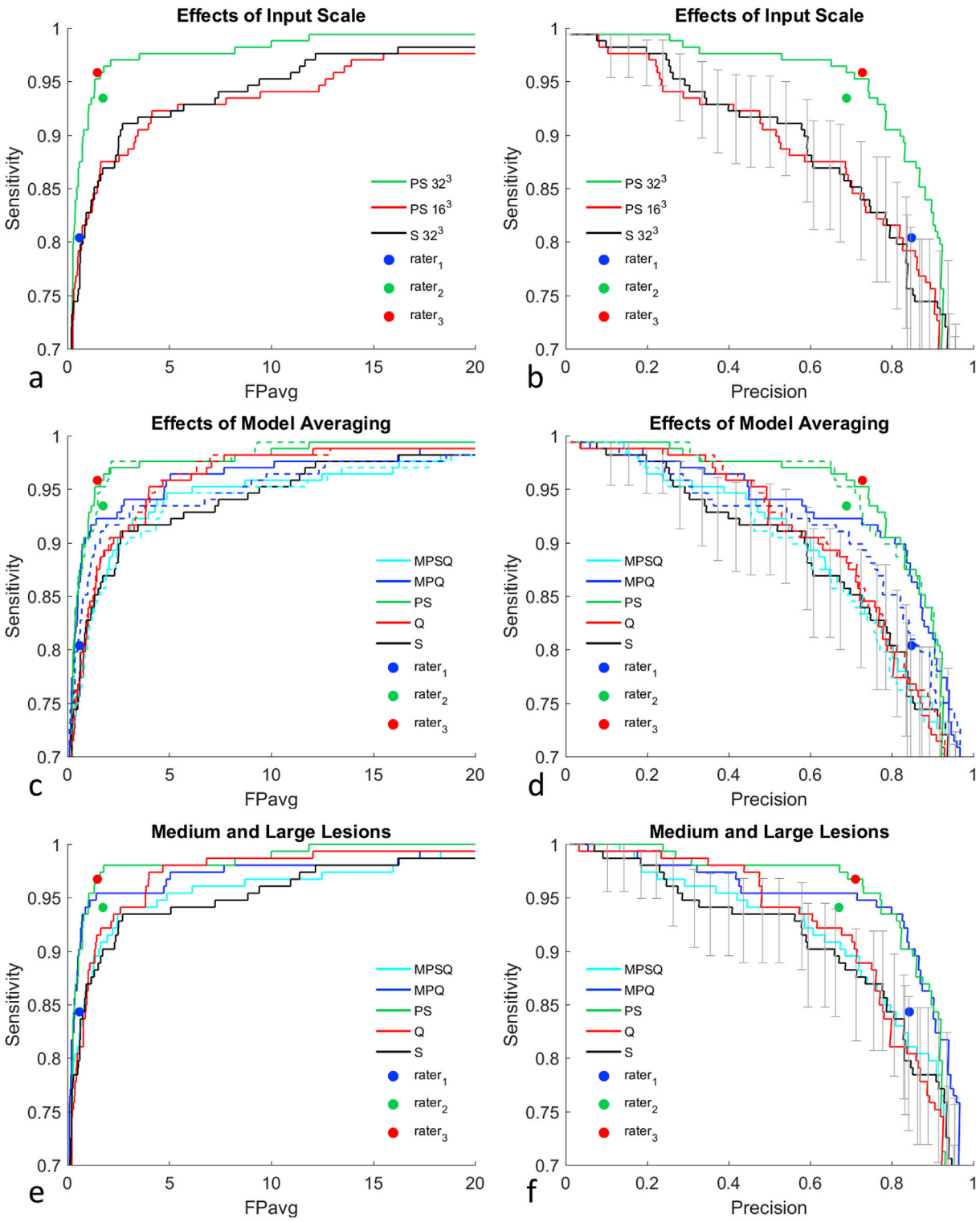

Fig. 4. Effects of input scale, model averaging and CMB size on the performance of best single models. (a) and (b): effects of input scale. The performance of the SWI model with 32^3 input scale (black lines in a and b) is shown as a reference. (c) and (d): effects of model averaging. Solid lines correspond to best single models and dashed lines correspond to model averages. (e) and (f): performance of the best single models when small lesions with size $\leq$ 2pixels were excluded. The plots in the first column (a, c, and e) are the free operating characteristic curves, while the plots in the second column (b, d and f) show the relationships between sensitivity and precision. For better visualization, the 95% confidence bounds were plotted for the SWI model only, in (b), (d) and (f).

prominent cortical veins are typical signs of stroke (Liu et al., 2017a). In the stroke cases, the performance of the PS model with the high precision threshold was: sensitivity: 98.3%, precision: 69.0%, FPavg: 2.9, and FP/cmb: 0.44; when using the high sensitivity threshold, the performance became: sensitivity: 100%, precision: 63.4%, FPavg: 3.8, and FP/cmb: 0.58.

On the test data, the computation time was 48 s per case. The CMB candidate detection step was the most time consuming step, costing 39 s per case. The false positive reduction step cost only 9 s per case, when only the best single model was used for making predictions.

Table 2
Performance of the human raters and the models on the test data.

		Rater 1	Rater 2	Rater 3	SWI	Phase + SWI (High Sensitivity)	Phase + SWI (High Precision)
All Lesions	Sensitivity(%)	80.4	93.5	95.8	91.7	95.8	93.5
	Precision(%)	84.9	68.9	72.9	52.0	70.9	75.5
	FPavg[a]	0.6	1.7	1.5	3.5	1.6	1.2
	FPhc[b]	0	0.1	0	1.2	0.3	0
	FP/cmb[c]	0.1	0.4	0.4	0.8	0.4	0.3
	Probability Threshold(%)	–	–	–	31.1	38.6	42.3
Medium and Large Lesions	Sensitivity(%)	84.3	94.1	96.7	93.5	96.7	95.4
	Precision(%)	84.3	67.0	71.2	56.3	72.9	77.2
	FPavg[a]	0.6	1.7	1.5	2.7	1.3	1.0
	FPhc[b]	0	0.1	0	0.8	0.1	0
	FP/cmb[c]	0.2	0.5	0.4	0.7	0.4	0.3
	Probability Threshold(%)	–	–	–	33.8	40.6	46.8

[a] FPavg: average number of false positives per case.
[b] FPhc: average number of false positives per case, evaluated in the healthy controls.
[c] FP/cmb: average number of false positives per CMB.

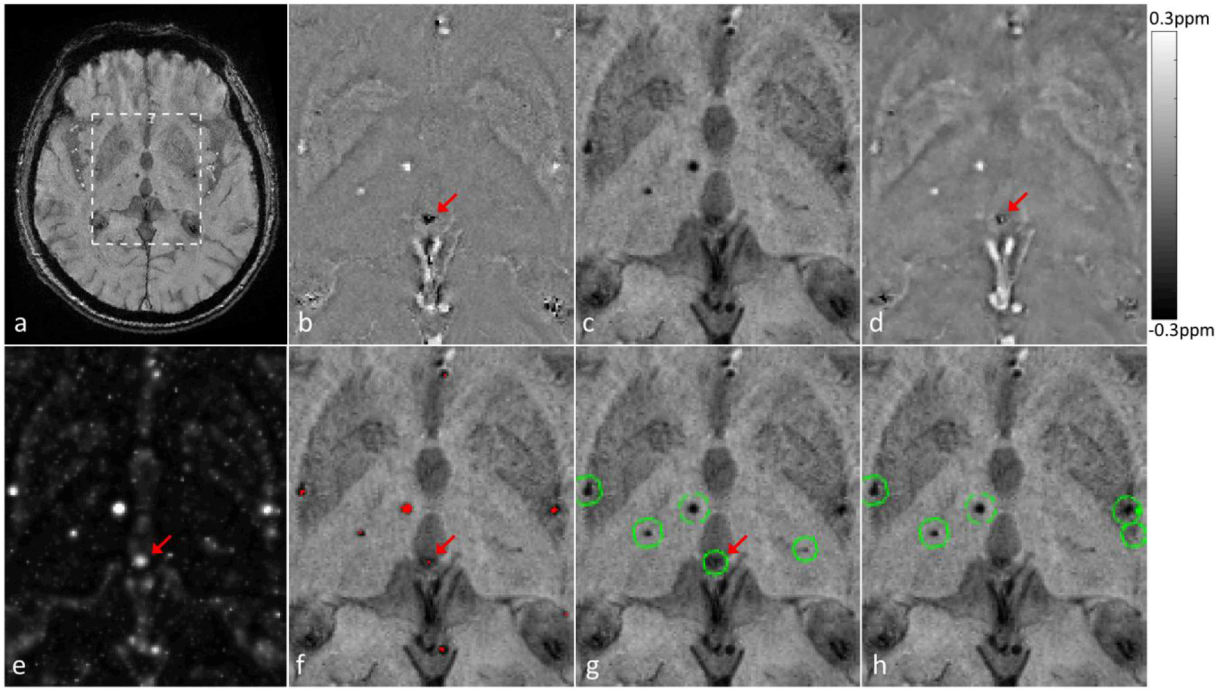

Fig. 5. Differentiation of calcification from CMBs using phase and SWI images. (a) Magnitude image. (b) Phase image. The calcification (red arrow) is negative, while CMBs are positive. (c) SWI image. (d) Quantitative susceptibility map. The scale bar is for (d) only. (e) 3D-FRST transformed SWI image. (f) SWI image with detected CMB candidates (red regions). Both CMBs and calcification (red arrow) were detected. (g) The model using SWI images alone failed to eliminate the false positive caused by the calcification (red arrow). (h) The model using phase and SWI images eliminated the one false positive successfully. (b) to (h) correspond to the dashed box region in (a).

4. Discussion

Recent years have witnessed great success of CNN based deep learning techniques in the computer vision field. Novel techniques such as deep residual neural networks have further improved the accuracy of natural image classification. However, several problems must be solved before deep learning can be applied to medical imaging. Unlike the natural images which are 2D with RGB channels, MRI data are usually 3D with multiple types of contrast. Second, depending on the imaging parameters, the intensities of MRI images may vary from scan to scan. In this study, we developed a two-stage CMB detection algorithm where CMB candidates were detected from SWI images using the 3D-FRST algorithm and false positive reduction was based on both phase and SWI images using a deep learning model. Through data pre-processing and normalization, the variations caused by imaging parameters were reduced, the scales of the intensities of different types of images standardized, and the contrast between the CMBs and the background tissue maximized. These procedures are fundamental to fully utilizing the information in the multi-contrast 3D MRI data.

Another major obstacle in applying deep learning to medical imaging is related to the relatively small sample sizes currently available in medical imaging applications, compared to those in natural image classification tasks. This is mainly caused by the high cost in medical data acquisition and labeling. The small sample size may lead to over-fitting and poor generalization. To alleviate this problem, we designed a relatively simple architecture with the number of trainable parameters being significantly smaller than those reported in earlier studies (0.2 million and 1.38 million parameters in the models proposed by Chen et al. (2018) and Dou et al. (Dou et al., 2016), respectively). Additionally, we performed data augmentation through 3D rotation and translation. By controlling the amount of maximum translation, the model was also made to focus on the central region of the 3D volume to reduce

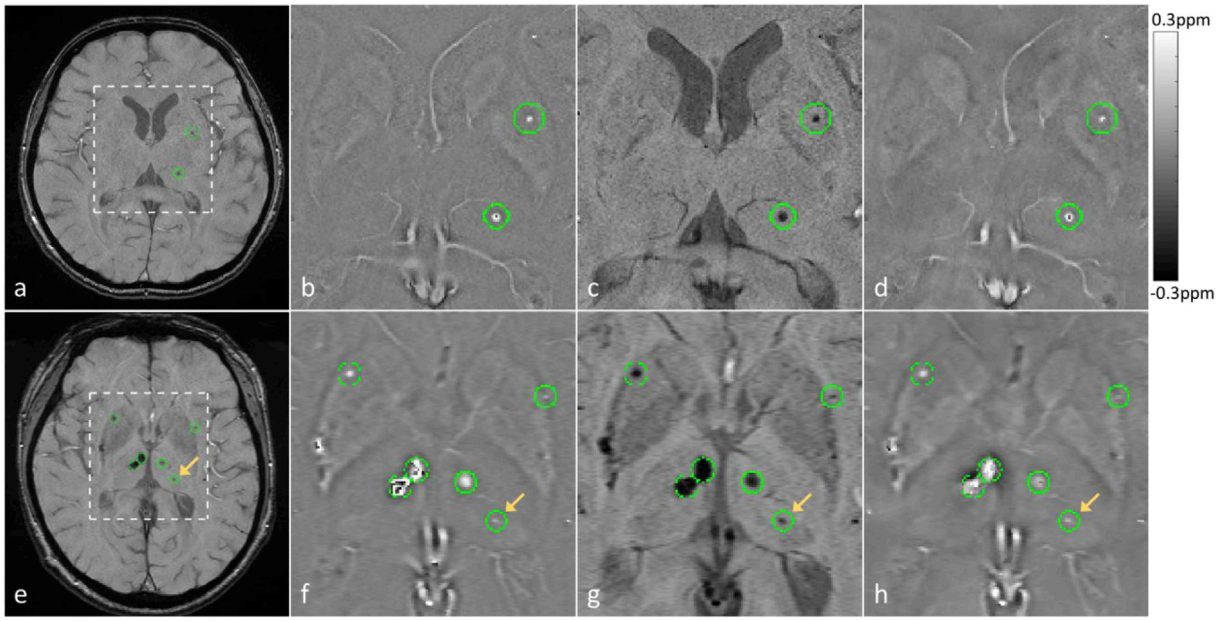

Fig. 6. Detected CMBs in subjects affected by hemodialysis (a to d) and stroke (e to h). (a) and (e): magnitude images. (b) and (f): phase images. (c) and (g): SWI images. (d) and (h): quantitative susceptibility maps. The scale bar is for (d) and (h) only. (b) to (d) and (f) to (h) correspond to the dashed box regions in (a) and (b), respectively. The CMBs are indicated by the green circles. The small lesion indicated by the yellow arrows in (e) to (h) was missed by all three raters.

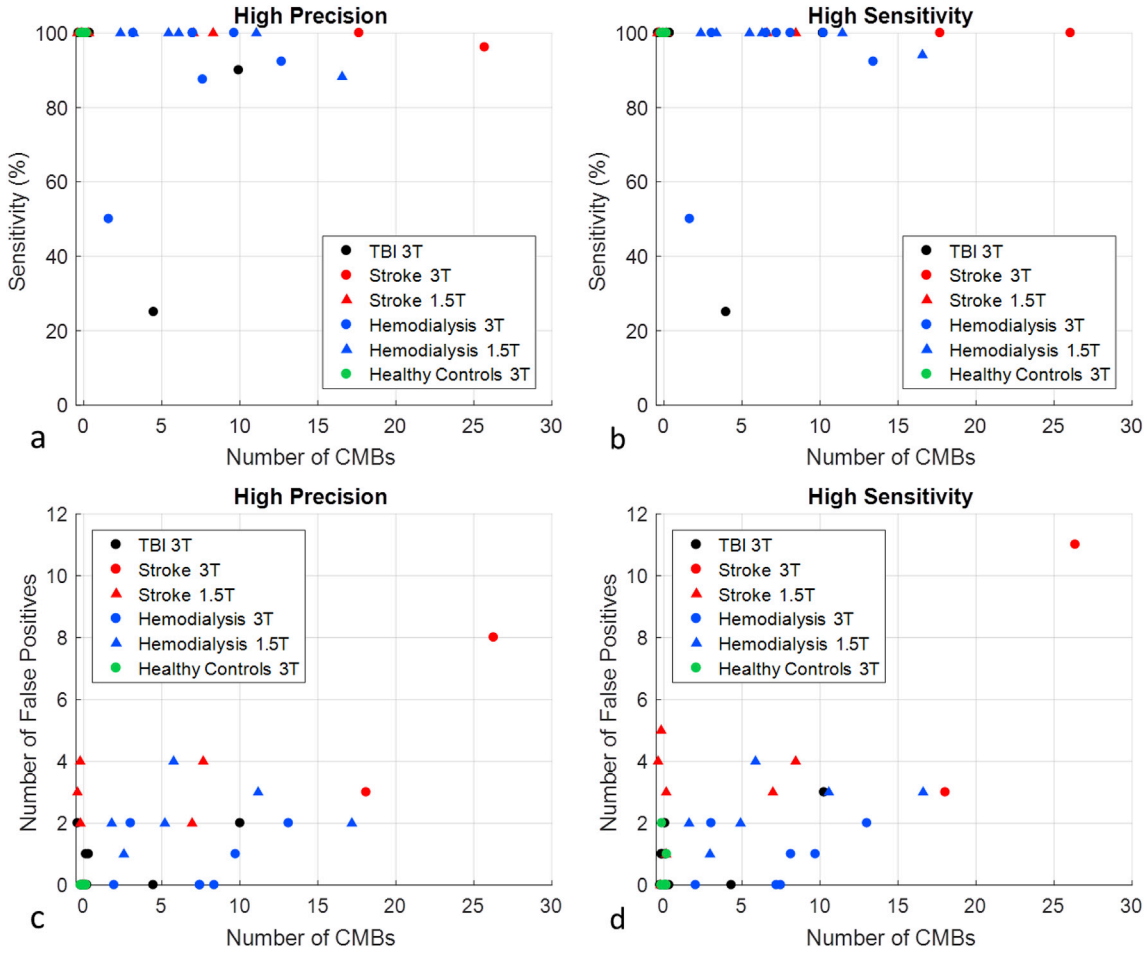

Fig. 7. Sensitivity and precision of the best phase + SWI model measured in each subject. Two probability thresholds were used for calculating sensitivity and precision, with one corresponding to a high precision scenario (a and c), and the other a high sensitivity scenario (b and d). (a) and (b): sensitivity vs. number of CMBs in each subject; (c) and (d): number of false positives vs. number of CMBs in each subject.

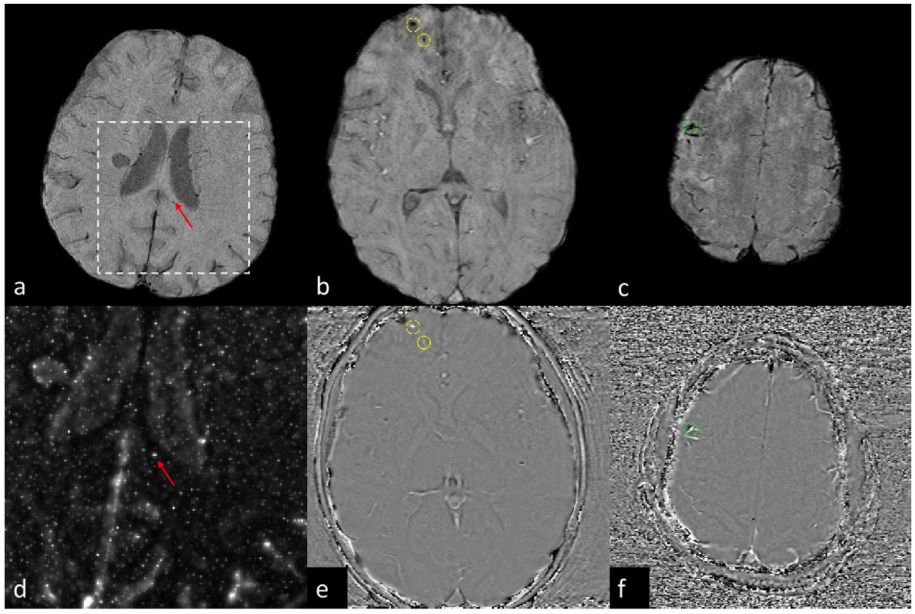

Fig. 8. Examples of false negatives (a, b, d and e) and false positives (c and f) in the test data. (a) and (d): a hemodialysis case in which one small lesion in the corpus callosum (red arrow) was missed during the CMB candidate detection step. (b) and (e): misclassified damaged veins (yellow circles) in the TBI case in which the best phase + SWI model had the lowest sensitivity. (c) and (f): a false positive (green circles) near the asymmetrically prominent cortical veins in the stroke case in which the best phase + SWI model detected the largest number of false positives. (a), (b) and (c) are SWI images; (d) is the 3D-FRST transformed image corresponds to the dashed box region in (a); (e) and (f) are the phase images correspond to (b) and (c), respectively.

interference of structures nearby. The test time augmentation further stabilized the prediction and reduced the uncertainties due to the selection of the centroids of the CMB candidates. A commonly used method for reducing the risk of over-fitting is transfer learning. However, because of the differences between medical data and natural images, borrowing the weights from trained models for natural image classification might not be suitable. Since SWI can be viewed as the output from a network which combines the magnitude and phase images, and QSM the output from a layer with weights being the 3D deconvolution kernel in susceptibility mapping, using the SWI and QSM data can also be viewed as a type of transfer learning.

While the SWI images provide the highest sensitivity to the presence of CMBs, the phase images enable the separation between blood products and calcification. Compared to the SWI model, the better performance of the phase + SWI model is not purely due to the differentiation of calcification, but also due to the removal of false positives associated with vessels, such as arteries with problematic flow compensation or veins. As a type of information derived from phase images, quantitative susceptibility maps are usually more valuable than the original phase, since QSM is not dependent on orientation of the object or imaging parameters (Haacke et al., 2015). Indeed, the QSM model performed the best among the models using only one type of image as input. However, the best model was the one which used a combination of phase and SWI images. This is possibly due to the loss of phase information caused by the high-pass filtering, which limited the accuracy of QSM reconstruction, and the remnant streaking artifacts caused by the ill-posed inverse problem in QSM. When all four types of images were used as input, the performance of the MPSQ model was still worse than that of the PS model. This could be due to the correlation between the input channels, especially between magnitude and SWI images, which may cause over-fitting.

For CMB candidate detection, we applied the 3D-FRST algorithm to the SWI images (Loy et al., 2002; Kuijf et al., 2012). Except for the threshold determined from the training and validation data, no additional parameters were introduced, further reducing the risk of over-fitting. Together with conventional image processing techniques such as dilation and erosion, high sensitivity was achieved, with the average number of false positives being similar to the CNN model used in Dou et al. (Dou et al., 2016) for candidate detection. One drawback of our approach is that multiple connected CMBs might be counted as one CMB. Although a small lesion was missed by the automatic detection in the corpus callosum in a hemodialysis case, it was detected by both the radiologist and the experienced SWI data processor, attributed to their knowledge of the normal appearance of the corpus callosum. This suggests that the sensitivity to CMBs can be further improved by integrating the anatomical position into the detection process.

Nonetheless, 100% sensitivity was achieved in 34 cases out of the total 41 cases, with 1.2 false positives per case (FP/cmb: 0.30) and 0 false positive in the healthy controls. With a lower probability threshold, an overall sensitivity 95.8%, precision 70.9%, with 1.6 false positives per case (FP/cmb: 0.39) was achieved. Practically, the selection of the probability threshold is a trade-off between sensitivity and specificity, depending on the specific aim of the application. The overall performance of the PS model is better than the results reported in earlier studies which also used deep learning techniques (Dou et al., 2016; Chen et al., 2018). In Chen et al., a sensitivity of 94.7%, with FPavg 11.6 (FP/cmb: 0.37) was reported for a model evaluated on 12 patients with 377 radiation induced CMBs (Chen et al., 2018). At a similar sensitivity of 95.2%, the FPavg of the PS model was 1.3, which corresponds to 0.33 FP/cmb. However, these differences in model performance could be attributed to the differences in the data used for training and evaluation in different studies.

For a given population, the optimal model performance is dependent on the distance between the distributions of false positives and true CMBs in the feature space (Cortes and Vapnik, 1995). While false positives can be caused by sources independent of the disease type, such as normal veins, calcification, arteries with poor flow compensation and susceptibility artifacts induced by air-tissue interfaces, they can also be caused by sources correlated with the underlying disease, such as veins with reduced oxygen saturation due to impaired blood perfusion in stroke. On the other hand, the distribution of CMBs in the feature space can also be affected by the disease type. Generally speaking, the difficulty in CMB detection for a given population can be reflected by the number of CMBs per subject. Hence, the FP/cmb is a more informative metric than FPavg, as FP/cmb provides the necessary normalization required in the comparison of models developed for different applications. When many cases with few potential CMBs are included, FPavg may be diluted, whereas FP/cmb will not be much affected.

For the test set used in this study, there were approximately 4 CMBs per subject. This is close to the average number of CMBs in the test set used in Dou et al. (a total of 117 CMBs in 50 stroke and normal aging subjects) (Dou et al., 2016). We found that the performance of our SWI

model was similar to the results obtained by Dou et al., who also developed a CNN based model using SWI images as input and reported a sensitivity of 93.2%, a precision of 44.3%, and FPavg 2.7 (FP/cmb: 1.17). Meanwhile, the performance of our PS model was significantly better, suggesting that the improvement is largely attributed to the use of phase images, rather than the specific design of the model architecture.

On a new data set, the performance of the PS model is mainly dependent on two factors, data acquisition and disease type. So long as the product of field strength and echo time is approximately 60 T*ms, and the disease type is similar to those investigated in this study, it can be expected that the performance of the PS model on the new test data will be similar to the results shown in this study. For the stroke cases which were mostly included in the test set, the PS model led to a sensitivity of 98.3%, with FPavg 2.9 (FP/cmb: 0.44), reflecting the generalization ability of this model.

There are a few limitations in this study. First, the samples used for both the training and testing were still limited in size. Consequently, the performance of the model varied between the different types of diseases, especially for the stroke cases where there were more false positives being detected (likely because only 4 stroke cases were included in the training and validation data). We noticed that most of the "false positives" found in stroke cases were associated with the asymmetrically prominent cortical veins, which are typical signs of stroke (Liu et al., 2017a). Hence, these false positives may be valuable for other applications focusing on this type of vein. For the model ensemble, we simply averaged over the predictions from multiple models. With more data, more sophisticated ensemble techniques could be used, such as weighted averaging. Moreover, we did not train separate models for different field strengths. Because of the products of echo time and field strength in the 1.5T and 3T data were approximately the same, leading to similar contrast between CMBs and surrounding tissue in the phase and SWI data. As a result, no significant difference was observed between the performance of the PS model on 1.5T and 3T test data.

Second, although we included TBI cases in the training and validation data, we did not separate the damaged veins or traumatic bleeds from those conventional bleeds not connected with any visible veins, but treated all the bleeds as one category and performed a binary classification. With more data from different disease types, it should be possible to perform a multi-class classification, which will provide more information on the nature of the lesions.

Third, both the training and testing of the model performance relied heavily on the manual labeling of the CMBs. In this study, the gold standard was obtained based on the consensus of experienced SWI data processors and a radiologist. Although relatively large variance in the performance of these raters was observed, the performance of the experienced SWI data processor was similar to that of the radiologist. In fact, we found that the performance of these raters correlated with their experience levels on SWI data processing. Furthermore, since only the data acquired with gradient-echo sequences were used, the manually created labels might contain other types of lesions with the same appearance to CMBs on SWI data, e.g. a cavernous malformation (Greenberg et al., 2009). This type of error can only be avoided with additional types of image contrasts, such as T2 weighted images that can restore some of the signal loss in T2* weighted images. However, using more input channels will add more parameters to the model, and hence increase the risk of over-fitting. This is a trade-off between bias and variance.

Fourth, a two-stage cascade architecture was adopted, with the CNN based techniques used only in the false positive reduction step. Although the use of the conventional image processing method for candidate detection helps to simplify the model training and reduced the risk of over-fitting, it also takes about 80% of the total computation time. With more training and validation data being available, a single step CMB detection based on fully convolutional networks should be possible and may further improve the time efficiency as inspired by the object detection studies in the computer vision field (Dou et al., 2016; Zhao et al., 2019).

Additionally, we did not implement all the algorithms proposed in earlier studies, but only compared the PS model with the reference SWI model. A thorough comparison of the CMB detection algorithms will require a test of all the algorithms on the same data set. Nevertheless, the effectiveness of the PS model can be demonstrated by its similar performance to that of the most experienced rater.

Finally, to avoid detecting false positives caused by iron deposition, a higher threshold (th_{rst}) was used for detecting CMB candidates in the globus pallidus region. This may not be optimal for those applications which require high sensitivity to the CMBs in the globus pallidus. In the current study, we only used the co-registered brain template for locating the globus pallidus. In the future, the template can be used for structure dependent candidate detection, and for reporting the anatomical location of the detected CMBs. The proposed method makes it possible to study the association between the presence of CMBs and risk of cognitive impairment or intracerebral hemorrhage. The automatic detection of CMBs has many potential clinical applications, including studies focused on Alzheimer's disease, Parkinson's disease, stroke, hemodialysis and TBI to name a few.

5. Conclusion

In this paper, we presented a two-stage model for CMB detection using deep learning with 3D multi-contrast MRI data. The use of phase images together with SWI images significantly improved the performance over models using SWI images alone. On the test data, the best model achieved similar performance to the most experienced human rater and outperformed the results reported in earlier studies. This study has demonstrated the great potential of applying deep learning to radiologic MR imaging, for improving the diagnostic accuracy and efficiency.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (grant number 81871342 to Shuang Xia).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.05.046.

References

Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C., 2008. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12 (1), 26–41.

Barnes, S.R.S., Haacke, E.M., Ayaz, M., Boikov, A.S., Kirsch, W., Kido, D., 2011. Semiautomated detection of cerebral microbleeds in magnetic resonance images. Magn. Reson. Imaging 29 (6), 844–852.

Bian, W., Hess, C.P., Chang, S.M., Nelson, S.J., Lupo, J.M., 2013. Computer-aided detection of radiation-induced cerebral microbleeds on susceptibility-weighted MR images. NeuroImage Clin. 2, 282–290.

Bihan, D.L., Mangin, J.-F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., et al., 2001. Diffusion tensor imaging: concepts and applications. J. Magn. Reson. Imaging 13 (4), 534–546.

Blair, R.C., Karniski, W., 1993. An alternative method for significance testing of waveform difference potentials. Psychophysiology 30 (5), 518–524.

Chen, Y., Villanueva-Meyer, J.E., Morrison, M.A., Lupo, J.M., 2018. Toward automatic detection of radiation-induced cerebral microbleeds using a 3D deep residual network. J. Digit. Imaging. https://doi.org/10.1007/s10278-018-0146-z.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31 (3), 968–980.

Dou, Qi, Chen, Hao, Yu, Lequan, Zhao, Lei, Qin, Jing, Wang, Defeng, et al., 2016. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans. Med. Imaging 35 (5), 1182–1195.

Fan, Yu Hua, Zhang, Lei, Lam Wynnie, W.M., Mok Vincent, C.T., Wong, Ka Sing, 2003. Cerebral microbleeds as a risk factor for subsequent intracerebral hemorrhages among patients with acute ischemic stroke. Stroke 34 (10), 2459–2462.

Fatemi-Ardekani, A., Boylan, C., Noseworthy, M.D., 2009. Identification of breast calcification using magnetic resonance imaging. Med. Phys. 36 (12), 5429–5436.
Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27 (8), 861–874.
Fazlollahi, A., Meriaudeau, F., Villemagne, V.L., Rowe, C.C., Yates, P., Salvado, O., et al., 2014. Efficient machine learning framework for computer-aided detection of cerebral microbleeds using the Radon transform. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI), 113–6.
Greenberg, S.M., 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 (2), 165–174.
Groppe, D.M., Urbach, T.P., Kutas, M., 2011. Mass univariate analysis of event-related brain potentials/fields I: a critical tutorial review. Psychophysiology 48 (12), 1711–1725.
Haacke, E.M., Xu, Y., Cheng, Y.N., Reichenbach, J.R., 2004. Susceptibility weighted imaging (SWI). Magn. Reson. Med. 52 (3), 612–618.
Haacke, E.M., Mittal, S., Wu, Z., Neelavalli, J., Cheng, Y.-C.N., 2009. Susceptibility-weighted imaging: technical aspects and clinical applications, Part 1. Am. J. Neuroradiol. 30 (1), 19–30.
Haacke, E.M., Liu, S., Buch, S., Zheng, W., Wu, D., Ye, Y., 2015. Quantitative susceptibility mapping: current status and future directions. Magn. Reson. Imaging 33 (1), 1–25.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770–778.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, pp. 630–645.
Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105.
Kuijf, H.J., de Bresser, J., Geerlings, M.I., Conijn, M.M.A., Viergever, M.A., Biessels, G.J., et al., 2012. Efficient detection of cerebral microbleeds on 7.0T MR images using the radial symmetry transform. Neuroimage 59 (3), 2266–2273.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Liu, J., Liu, T., de Rochefort, L., Ledoux, J., Khalidov, I., Chen, W., et al., 2012. Morphology enabled dipole inversion for quantitative susceptibility mapping using structural consistency between the magnitude image and the susceptibility map. Neuroimage 59 (3), 2560–2568.
Liu, S., Buch, S., Chen, Y., Choi, H.-S., Dai, Y., Habib, C., et al., 2017a. Susceptibility-weighted imaging: current status and future directions. NMR Biomed. 30 (4).
Liu, S., Zheng, H., Feng, Y., Li, W., 2017b. Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. In: Medical imaging 2017: computer-aided diagnosis, 10134, p. 1013428.
Loshchilov, I., Hutter, F., 2017. Fixing Weight Decay Regularization in Adam. arXiv preprint. arXiv:1711.05101.
Loy, G., Zelinsky, A., 2002. A fast radial symmetry transform for detecting points of interest. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (Eds.), Computer Vision — ECCV 2002. Springer Berlin Heidelberg, pp. 358–368.
Maas, A.I.R., Menon, D.K., Adelson, P.D., Andelic, N., Bell, M.J., Belli, A., et al., 2017. Traumatic brain injury: integrated approaches to improve prevention, clinical care, and research. Lancet Neurol. 16 (12), 987–1048.
Mazziotta, J.C., Toga, A.W., Evans, A., Fox, P., Lancaster, J., 1995. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). Neuroimage 2 (2), 89–101.
McMahon, P., Hricik, A., Yue, J.K., Puccio, A.M., Inoue, T., Lingsma, H.F., et al., 2014. Symptomatology and functional outcome in mild traumatic brain injury: results from the prospective TRACK-TBI study. J. Neurotrauma 31 (1), 26–33.
Pacurar, E.E., Sethi, S.K., Habib, C., Laze, M.O., Martis-Laze, R., Haacke, E.M., 2016. Database integration of protocol-specific neurological imaging datasets. Neuroimage 124, 1220–1224.
Poels, M.M.F., Ikram, M.A., van der Lugt, A., Hofman, A., Niessen, W.J., Krestin, G.P., et al., 2012. Cerebral microbleeds are associated with worse cognitive function: the Rotterdam Scan Study. Neurology 78 (5), 326–333.
Poels Mariëlle, M.F., Vernooij Meike, W., Arfan, Ikram M., Albert, Hofman, Krestin Gabriel, P., van der Lugt, Aad, et al., 2010. Prevalence and risk factors of cerebral microbleeds. Stroke 41 (10_Suppl. l_1), S103–S106.
Schweser, F., Deistung, A., Lehr, B.W., Reichenbach, J.R., 2011. Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: an approach to in vivo brain iron metabolism? Neuroimage 54 (4), 2789–2807.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17 (3), 143–155.
Tong, K.A., Ashwal, S., Holshouser, B.A., Shutter, L.A., Herigault, G., Haacke, E.M., et al., 2003. Hemorrhagic shearing lesions in children and adolescents with posttraumatic diffuse axonal injury: improved detection and initial results. Radiology 227 (2), 332–339.
Trifan, G., Gattu, R., Haacke, E.M., Kou, Z., Benson, R.R., 2017. MR imaging findings in mild traumatic brain injury with persistent neurological impairment. Magn. Reson. Imaging 37, 243–251.
Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., et al., 2010. N4ITK: improved N3 bias correction. IEEE Trans. Med. Imaging 29 (6), 1310–1320.
van den Heuvel, T.L.A., van der Eerden, A.W., Manniesing, R., Ghafoorian, M., Tan, T., Andriessen, T.M.J.C., et al., 2016. Automated detection of cerebral microbleeds in patients with traumatic brain injury. NeuroImage Clin. 12, 241–251.
Wu, Z., Mittal, S., Kish, K., Yu, Y., Hu, J., Haacke, E.M., 2009. Identification of calcification with MRI using susceptibility-weighted imaging: a case study. J. Magn. Reson. Imaging 29 (1), 177–182.
Yates, P.A., Villemagne, V.L., Ellis, K.A., Desmond, P.M., Masters, C.L., Rowe, C.C., 2014. Cerebral microbleeds: a review of clinical, genetic, and neuroimaging associations. Front. Neurol. 4, 205.
Zhao, Z., Zheng, P., Xu, S., Wu, X., 2019. Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn Syst. 1–21.