Detection of Cerebral Microbleeds in MR Images Using a Single-Stage Triplanar Ensemble Detection Network (TPE-Det)

Haejoon Lee, BSc,[1,2] Jun-Ho Kim, BSc,[1] Seul Lee, MSc,[1] Kyu-Jin Jung, MSc,[1]
Woo-Ram Kim, ME,[3] Young Noh, MD, PhD,[3,4] Eung Yeop Kim, MD, PhD,[5] ⓘⒹ
Koung Mi Kang, MD, PhD,[6,7] Chul-Ho Sohn, MD, PhD,[6,7] Dong Young Lee, MD, PhD,[8,9,10]
Mohammed A. Al-masni, PhD,[11]* ⓘⒹ and Dong-Hyun Kim, PhD[1]* ⓘⒹ

Background: Cerebral microbleeds (CMBs) are microscopic brain hemorrhages with implications for various diseases. Automated detection of CMBs is a challenging task due to their wide distribution throughout the brain, small size, and visual similarity to their mimics. For this reason, most of the previously proposed methods have been accomplished through two distinct stages, which may lead to difficulties in integrating them into clinical workflows.
Purpose: To develop a clinically feasible end-to-end CMBs detection network with a single-stage structure utilizing 3D information. This study proposes triplanar ensemble detection network (TPE-Det), ensembling 2D convolutional neural networks (CNNs) based detection networks on axial, sagittal, and coronal planes.
Study Type: Retrospective.
Subjects: Two datasets (DS1 and DS2) were used: 1) 116 patients with 367 CMBs and 12 patients without CMBs for training, validation, and testing (70.39 ± 9.30 years, 68 women, 60 men, DS1); 2) 58 subjects with 148 microbleeds and 21 subjects without CMBs only for testing (76.13 ± 7.89 years, 47 women, 32 men, DS2).
Field Strength/Sequence: A 3 T field strength and 3D GRE sequence scan for SWI reconstructions.
Assessment: The sensitivity, FP_{avg} (false-positive per subject), and precision measures were computed and analyzed with statistical analysis.
Statistical Tests: A paired *t*-test was performed to investigate the improvement of detection performance by the suggested ensembling technique in this study. A *P* value < 0.05 was considered significant.
Results: The proposed TPE-Det detected CMBs on the DS1 testing set with a sensitivity of 96.05% and an FP_{avg} of 0.88, presenting statistically significant improvement. Even when the testing on DS2 was performed without retraining, the proposed model provided a sensitivity of 85.03% and an FP_{avg} of 0.55. The precision was significantly higher than the other models.
Data Conclusion: The ensembling of multidimensional networks significantly improves precision, suggesting that this new approach could increase the benefits of detecting lesions in the clinic.
Evidence Level: 1
Technical Efficacy: Stage 2

Cerebral microbleeds (CMBs) are microscopic brain hemorrhages that can be caused by structural distortions of the small blood vessels.[1] These lesions have clinical implications for various diseases, such as cerebral amyloid angiopathy, cerebrovascular disease, traumatic brain damage, and Alzheimer's disease.[2] Pathological observations revealed that CMBs cause damage to surrounding brain tissues, leading to dementia, dysfunction, and cognitive impairment.[3] According to a related study, CMBs are regularly identified not only in people with stroke but also in healthy elderly individuals.[1] Therefore, for proper diagnosis and treatment, it is very important to accurately distinguish CMBs from their mimics, such as iron deposits, calcifications, and veins seen in neuro-MRI.

MRI is widely used for detecting CMBs.[4] In particular, susceptibility-weighted imaging (SWI), a technique for reconstructing images with improved visibility of susceptibility sources, is preferred in clinical situations.[5] According to the theory of susceptibility in brain tissue, CMBs generate inhomogeneity in the magnetic field surrounding them, resulting in small spherical regions with hypointensities on the reconstructed SWI images. Due to the high sensitivity of SWI to hemosiderin, it can reveal even minute CMBs.[6] Nevertheless, it is still difficult to detect CMBs because they are highly variable in position and very small in size, and numerous mimics, such as flow voids from blood vessels or air–bone interfaces, exist.[5]

Many of the recently published works on the automatic detection of CMBs consist of two separate stages: 1) the CMB candidate detection stage, which employs conventional detection algorithms such as fast radial symmetry transform (FRST) or the CNN-based networks (eg YOLO detector), and 2) the false-positive (FP) reduction stage via CNN, which classifies the potential candidates of the first stage into CMB or non-CMB.[7–11] For example, a two-stage framework[8,9] consisting of FRST as a candidate detection network and 3D-CNN as an FP reduction network was developed. The usage of FRST for candidate detection could help to simplify model training procedures and prevent the risk of over-fitting. However, the first stage with FRST took about 80% of the whole computation time.[8]

To boost the efficiency of the first stage, some works[7,11–13] employed deep learning networks in their candidate detection stage. Since an immense number of FPs are generated from the candidate detection stages, most of the studies mentioned above exhibited improvements in performance by employing 3D networks in their FP reduction stages. Nevertheless, there are still limitations to these approaches. First, the extracted features from the first stage could not be passed to the second stage since the two stages are not concatenated. This could lead to losses in information and efficiency. Second, the network of each stage was trained independently and sequentially. Thus, the second stage

needed the outputs from the first stage for training, and it was not possible to train the two-stage frameworks end to end. Additionally, each stage of some of the two-stage frameworks was implemented using different platforms. For example, Liu et al,[8] Chen et al,[9] and Al-masni et al[12] performed their first stage through MATLAB or C++ and the second stage via Python. Therefore, there will be difficulties in integrating them into clinical workflows since each stage was implemented using a different software environment.[10]

To overcome the above limitations, several studies have proposed a framework consisting of only one stage.[14–16] Wang et al[14] and Hong et al[15] applied a 2D sliding window technique to the entire MR data and performed classification on each sliding window for all subjects. Both works achieved high sensitivity. However, applying the 2D sliding window technique across the entire 3D MR volume involves an enormous computation cost and requires a large inference time. For example, if $512 \times 448 \times 72$ voxel data set is considered, more than 16 million 2D patches need to be extracted from a pixel-by-pixel sliding window. Consequently, it is burdensome to employ 3D networks that are much heavier than 2D networks in such one-stage frameworks. In addition, most recently, Kim et al[16] proposed a single-stage CMB detector employing a 3D feature fusion module. Due to the computational cost, they divided each 3D subject datum into multiple patches and performed detection, providing a precision of 25.64%.

Since 3D information is advantageous when seeking to extract such unique features of CMBs, most frameworks with outstanding performances have employed 3D networks,[7–9,12] whereas 2D networks detecting CMBs only on a single plane (i.e. usually the axial plane) may lead to generating a massive amount of FPs. However, employing 3D networks could require trading off huge computation costs against improved accuracies compared to just employing 2D networks.[17] The huge amount of parameters of a 3D network could lead to a high memory requirement for training models,[18] and the network could massively overfit on a small dataset, leading to poor generalization.[19] Note that most of the above approaches exploited 3D networks only in their FP reduction stages (i.e. the second stage). To resolve the computational limitations of 3D networks, applying 2D models individually on each of the three planes has been proposed for white matter hyperintensities segmentation.[20] This method segmented the lesions on each plane using 2D U-Net, and then stacked the predicted slices to generate 3D data by averaging results from all planes. It was reported that the ensemble network reduced discontinuities across slices and provided better lesion boundary delineation, using fewer parameters compared to a 3D CNN.

To address the above limitations and issues, this study aimed to develop a clinically feasible end-to-end CMBs detection network with a single-stage structure utilizing 3D information.

Materials and Methods

This study was approved by the Institutional Review Board (IRB) of the institutions where the datasets were acquired, and all subjects provided written informed consent.

Background Hypothesis

As shown in Fig. 1, most CMBs were observed as circular shapes in the axial plane while elongated elliptic shapes in the sagittal and coronal planes in the anisotropic data of this study that are typically used in the clinic.[21,22] The elliptical shapes have not been observed in isotropic resolution data[23] or postmortem MR images of the lesions.[24] The mimics have different shapes with CMBs in at least one plane; mainly, veins, which are the prevailing mimics of CMBs in MR images, show up as long and curvy lines in the other planes. Accordingly, this study hypothesized that most of the mimics could be distinguished from CMBs by observing the other planes with these differences, which is how detection is typically done in the clinic.

General Overview

The overview of the proposed framework is shown in Fig. 2. The CMBs from the axial, sagittal, and coronal images were automatically and independently detected via the proposed TPE-Det. From the detection results on each plane, the 3D coordinates of all the detected bounding boxes in the volume data were calculated from their x, y, and z coordinates. Only objects that were simultaneously detected on all three planes were selected as the final detections. The source code of the proposed TPE-Det is available at https://github.com/anonymzedname/TPE-Det.

The main contributions of this study are as follows. First, we eliminated the need for the two-stage approach by developing an end-to-end method capable of efficiently detecting CMBs. Second, we leveraged 3D information by assembling 2D networks on three different planes.

Data Acquisition

Two datasets (DS1 and DS2) were used in this study (Table 1). DS1 was collected from hospital 1. The SWI data were acquired using 3.0 T Verio and Skyra Siemens MRI scanners (Siemens Healthineers, Erlangen/Germany) with the following imaging parameters: a resolution of $0.5 \times 0.5 \times 2.0 \text{ mm}^3$, a repetition time (TR) of 27 msec, an echo time (TE) of 20 msec, a flip angle (FA) of $15°$, and a field of view (FOV) of $256 \times 224 \times 144 \text{ mm}^3$. A total of 116 subjects with 367 microbleeds and 12 subjects without
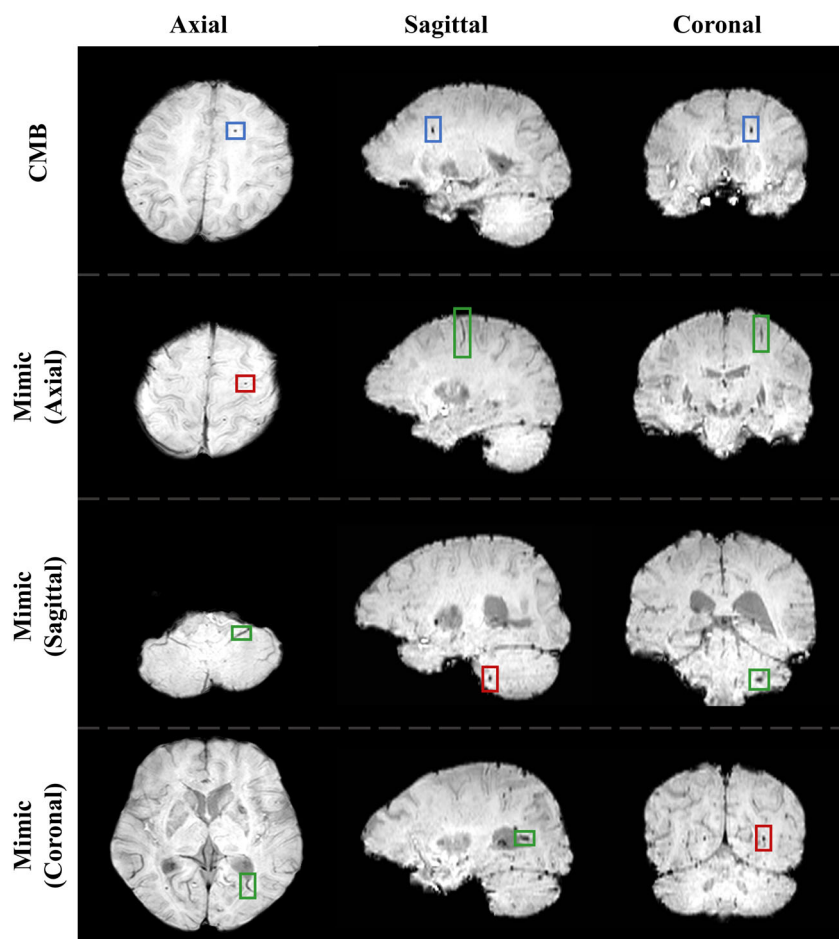
FIGURE 1: Examples of CMB and its mimics on the axial, sagittal, and coronal planes (in SWI images). Each row illustrates each sample on the three plane views from different patients. Blue boxes indicate the CMB, red boxes point to each mimic on each plane, and green boxes indicate the mimic on the other views. The imaging acquisition protocol for this example was 0.50 × 0.50 mm² in the axial and 0.50 × 2.00 mm² in the other planes.

a

b

	Backbone	BiFPN		Box/class
	Network	# of channels	# of layers	# of layers
D3	EfficientNet-B3	160	6	4
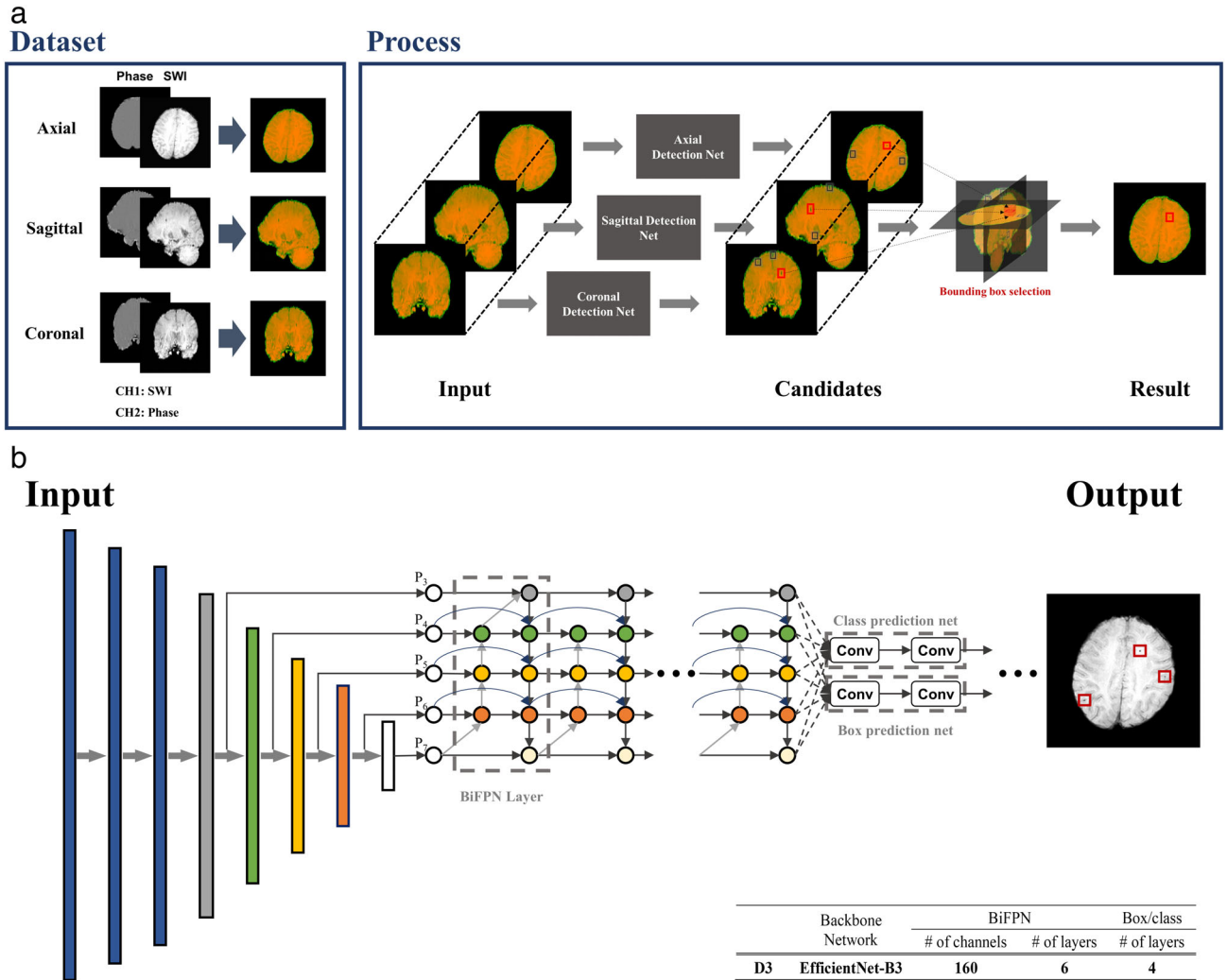
FIGURE 2: Overview diagram of the proposed Triplanar Ensemble Detection network (TPE-Det). (a) The left side illustrates the generation of input images via the combination of SWI and phase images. The right side refers to the ensemble process of the proposed TPE-Det. (b) Architecture of the EfficientDet D3. The table in the right-hand corner displays the configurations of the detection network.

TABLE 1. Details of Datasets

	DS1				DS2		
	CMB Present			CMB Absent	CMB Present	CMB Absent	
	Training	Validation	Testing	Testing	Testing	Testing	Total
Patients	85	6	25	12	58	21	207
CMBs	274	17	76	0	148	0	515

CMBs were collected (70.39 ± 9.30 years, 68 women, 60 men). In the experiments of this study, the whole dataset was randomly divided into three portions on the subject level: 85 patients with 274 CMBs for training, 6 patients with 17 CMBs for validation, and 25 patients with 76 CMBs for testing (Table 1). In this study, the proposed model was trained using only DS1 and them on DS1.

Another dataset was collected (DS2) from hospital 2, with different acquisition parameters. The SWI data were acquired using a 3.0 T Biograph mMR Siemens MRI scanner (Siemens Healthineers) with the following protocol: a resolution of $0.5 \times 0.5 \times 3.0$ mm^3, a TR of 28 msec, a TE of 20 msec, an FA of 15°, and an FOV of $192 \times 219 \times 156$ mm^3. A total of 58 subjects with 148 microbleeds

and 21 subjects absent from CMBs were collected (76.13 ± 7.89 years, 47 women, 32 women).

Labeling

The labels of DS1 were obtained according to the consensus of three raters (E.Y.K., a neuroradiologist with 26 years of experience, W.R.K., a medical imaging researcher with eight years of experience, and Y.N., a neurologist with 17 years of experience). W.R.K. and Y.N. assigned the labels first and then E.Y.K. verified them lastly. The labels of DS2 were annotated by K.M.K., a neuroradiologist with 13 years of experience. The readers utilized in-house software developed using MATLAB (MathWorks Inc., Natick, MA, USA) for labeling. The readers examined the SWI and phase brain images of the same subject. The x, y, and z coordinates of a CMB's center were generated by manually clicking CMBs in the software. Bounding boxes with a size of 17×17 voxels on the axial plane and with a size of 35×20 voxels on the sagittal and coronal planes were then positioned around each labeled CMB as a requirement of the detection network. All labeling procedures were completed by following the criteria proposed by Greenberg et al[25]: 1) black on T2*-weighted MRI, 2) round or ovoid (rather than linear), 3) blooming on T2*-weighted MRI, 4) devoid of signal hyperintensity on T1- or T2-weighted sequences, 5) at least half surrounded by brain parenchyma, 6) distinct from other potential mimics such as iron/calcium deposits, bone, or vessel flow voids, and 7) clinical history excluding traumatic diffuse axonal injury. Using VINCI software (Max Planck Institute for Metabolism Research, Cologne, Germany), the size of bleeding was measured, and only those with diameters smaller than 10 mm were defined as CMBs. Calcifications were distinguished by using phase images.

Data Preprocessing

Brain region extraction, data normalization, phase image inversion, averaging of adjacent slices, and contrast enhancement techniques were applied to both datasets (DS1 and DS2) as data preprocessing procedures. First, a Brain Extraction Tool (BET) was obtained to obtain binary masks of the brain tissue from MR magnitude images[26] and applied them to SWI and phase images. Each volumetric datum of DS2 was interpolated so that it had the same slice thickness as DS1 for the uniform size of the brain regions. Then, the volumetric data for each subject were cropped from the size of $512 \times 448 \times 72$ to $360 \times 360 \times 72$ by discarding the empty background from the top, bottom, right, and left sides and then resized into $360 \times 360 \times 360$ to ensure that the sagittal and coronal images are square, coping the requirement of the employed detection EfficientDet network.[27] Each 3D datum was normalized in a range of 0–1 throughout the whole volume to ensure the consistency of input intensities. Voxel values of the phase images were reversed $(1 - \text{phase})$ to intensify the CMB's signal intensities in the combined images. The two-dimensional images of axial, sagittal, and coronal planes were obtained by slicing the volumetric SWI and phase data. For all three orientations, each slice and its adjacent slices were then averaged to reflect the 3D contextual information of the CMBs, as applied by Al-masni et al.[12] Lastly, the contrast of SWI images was stretched by mapping the pixel intensity values to new wider range such that the image's histogram was linearly scaled, and the values were evenly distributed from maximum to minimum (eg 255 to 0 on an 8-Bit image).

This study utilized the phase data together with the SWI.[8] The generated SWI and phase images were combined into two-channel inputs. This process generated 72 images from the axial, 360 images from the sagittal, and 360 images from the coronal plane from one subject's volume datum.

Proposed Detection Network

Each detection network of the TPE-Det is based on EfficientDet.[27] The EfficientDet can simultaneously detect the location of multiple CMBs, providing coordinates of bounding boxes with confidence scores as output. The architecture of the model is shown in Fig. 2b. The configurations of the developed models from compound scaling are depicted in the table in Fig. 2b. The three EfficientDet D3s were independently trained using the preprocessed axial, sagittal, and coronal images.

Bounding Box Selection

Initially, the proposed method independently detected the CMBs using the axial, sagittal, and coronal images via the proposed TPE-Det. As the detection was performed slice by slice using 2D detection networks, the detected candidates were counted as a single CMB if they were cross sections of a single object which was appearing on sequential slices. Then, the three-dimensional coordinates of each detected candidate were calculated from the x and y coordinates and the slice number of the image. As the datasets in this study have an anisotropic resolution ($0.5 \times 0.5 \times 2.0$ and $0.5 \times 0.5 \times 3.0$ mm^3), the number of axial images (N_{ax}) was smaller than the number of the sagittal and coronal images ($N_{sag,cor}$). Thus, the z values of the candidates detected from the sagittal and coronal slices were rounded to the same number of discrete integers as the number of axial slices to match the range of coordinates. This method can be expressed as follows:

$$z'_{sag,cor} = \frac{N_{sag,cor}}{N_{ax}} \times \text{Roundup}\left(\frac{N_{ax}}{N_{sag,cor}} \times z_{sag,cor}, 0\right) \quad (1)$$

where Roundup($X, 0$) rounds X up to zero decimal places. Afterward, the Euclidean distances were calculated between the candidates detected from each plane and deemed the objects to be single CMB if they were closer than a preset distance threshold value. Based on this method, the candidates detected on all three planes were considered to be the final detections. As a result, only those candidates who had morphological characteristics of CMBs in every plane were left, despite not employing an additional classification network.

Implementation

The proposed TPE-Det was implemented through PyTorch 3.8.10. NVIDIA RTX A5000 with 24 GB memory was utilized for accelerated computation. The network on each plane was trained on the training set with a batch size of two and different hyperparameters. This study implemented the following data augmentation techniques during the training to improve the generalizability of the networks: 1) horizontal flip, 2) vertical flip, and 3) random rotation (only axial).

Statistical Analysis

Sensitivity, precision, and FP_{avg} values were evaluated in the following experiments. They were defined as follows:

$$\text{Sensitivity}(TPR) = \frac{TP}{TP + FN}, \quad (2)$$

$$\text{Precision}(PPV) = \frac{TP}{TP + FP}, \quad (3)$$

$$FP_{avg} = \frac{FP}{N}, \quad (4)$$

where TP (true positive): the number of CMBs that were detected correctly, FN (false negative): the number of CMBs that were not detected, FP: the number of false detections, and N: the number of subjects.

Statistical analysis was performed to investigate the improvement of precision by the suggested ensembling technique in this study. A paired *t*-test was employed for comparing performances from each combination of ensembling detectors and other detection models implemented on the same testing set. Analyses were performed using IBM SPSS Statistics version 25.0 (IBM corp., Armonk, NY). A *P* value of less than 0.05 was considered statistically significant.

The ensemble detection network had multiple confidence score thresholds ($CS_{ths_{ax}}$, $CS_{ths_{sag}}$, and $CS_{ths_{cor}}$) for each detector on each plane. Thus, this study jointly scaled the threshold values by employing a common scale factor.

Comparison Against Other Detection Models

For direct comparison, this study implemented several CMB detection models using the same dataset and compared the results. First, this article built a 3D Faster region-based CNN (R-CNN) with a two-stage end-to-end structure and evaluated the detection performances of the network using the same dataset (DS1). The 3D faster R-CNN was built based on the original 2D faster R-CNN framework.[28] To exploit the high-resolution features and 3D information that are highly important in CMBs detection, this study used a 3D U-Net[29] as a backbone network of the region proposal network (RPN) for candidate detection. The outputs were then passed from the RPN into the 3D CMB classification network of Al-masni et al[12] for FP reduction. It is of note that both components of the 3D Faster R-CNN (RPN and 3D CNN) were trained together in an end-to-end manner.

Furthermore, this study implemented the most recently proposed fully deep learning CMB detection methods for each two-stage[10] and single-stage structure[16] utilizing 3D information. All three models were trained and evaluated on the same dataset (DS1) with the proposed TPE-Det.

Testing on the Other Dataset With Different Acquisition Parameters (DS2)

This article also analyzed the testing results on the DS2 subjects that were acquired from different scanning parameters using the proposed trained ensemble models. Two groups were additionally tested: 1) 21 subjects without CMBs from DS2 and 2) 58 subjects with 148 microbleeds from DS2. This study performed the testing without retraining the models to evaluate the proposed approach's generalizability. The testing was performed with the same confidence score threshold values as the testing on DS1.

Results

Detection on Each Plane

The results for CMB detection on DS1's individual axial, sagittal, and coronal images are shown in Fig. 3. Sensitivity and FP_{avg} affected by varying confidence score thresholds on each plane are presented in Fig. 3a,b. By combining these two metrics, this article presents free response operating characteristic (FROC) curves in Fig. 3c. Based on these results, threshold values were chosen that provided higher sensitivities than previously reported values (Table 5) for each axial, sagittal, and coronal plane detection network (circles in Fig. 3). Due to the lower resolution (0.5×2.0 mm^2 vs. 0.5×0.5 mm^2), the detection networks on the sagittal and coronal planes detected CMBs with lower confidence scores. Thus, the smaller thresholds were selected for the sagittal and coronal detectors than the axial detector to ensure high sensitivities. The axial detector detected spherical regions as candidates. Compared to the other two planes, the smaller number of FP in the axial plane was due to the high resolution of the axial images (0.5×0.5 mm^2 vs. 0.5×2.0 mm^2). In the case of the sagittal and coronal detectors, elongated ellipse regions,
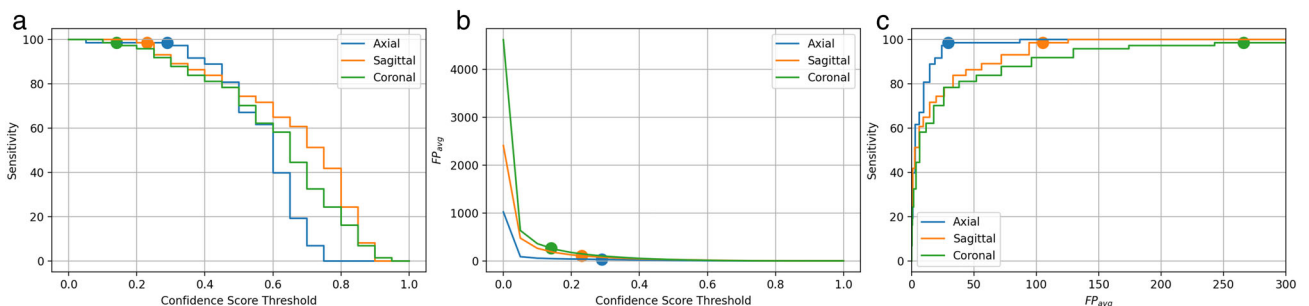

FIGURE 3: Impact of the confidence score threshold on the detection performance of each detection network of each plane. (a) Sensitivity vs. threshold. (b) FP_{avg} vs. threshold. (c) Sensitivity vs. FP_{avg}. The testing was performed on the DS1 testing set.

TABLE 2. Performance of Each Ensemble Network With Fixed Confidence Thresholds on the DS1 Dataset

| | DS1 | | | |
| | CMBs Present | | | CMBs Absent |
Ensemble Network	Sensitivity (%)	FP_{avg} (Improvement rate)	Precision (%)	FP_{avg} (Improvement rate)
Axial	97.26	29.43 (baseline)	9.49	30.5 (baseline)
Sagittal	97.26	109.87 (NA)	2.73	117.27 (NA)
Coronal	98.63	261.09 (NA)	1.18	293.08 (NA)
Axial + Coronal	97.26	4.41 (84.88%)	40.49	4.33 (85.80%)
Axial + Sagittal	96.05	2.54 (91.29%)	53.80	2.92 (90.43%)
Axial + Sagittal + Coronal (TPE-Det)	96.05	0.88 (96.99%)	76.76	1.08 (96.45%)
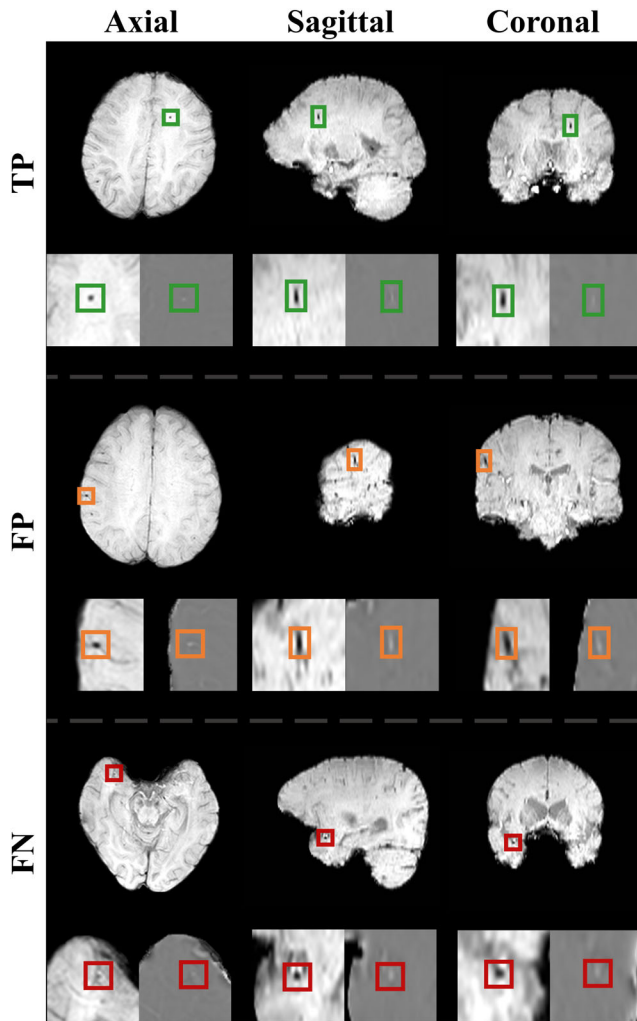
FIGURE 4: Examples of the detected regions via the proposed TPE-Det using DS1 (SWI images). The first, second, and third rows indicate the TP, FP, and FN cases, respectively. The bottom of the brain images in each row shows the ROI on SWI (left) and phase images (right).

displaying analogous forms to ground-truths in Fig. 1, were detected as candidates.

Ensembles of the Axial, Sagittal, and Coronal Detection Networks

Table 2 presents the CMBs detection results and comparisons of various ensembled combinations with the fixed confidence score thresholds that were presented in the previous section. FP_{avg} decreased by up to 96.76%, and precision statistically significantly increased ($P < 0.001$) to 76.76% by ensembling the detection networks of three different perpendicular planes, even with only a 1.21% drop in sensitivity on the DS1 subjects with CMBs.

Qualitatively, this article illustrates the TP, FP, and FN examples via the proposed TPE-Det in Fig. 4. TP CMBs commonly had a small spherical shape in the axial plane but an elongated ellipse shape in the sagittal and coronal planes. The FPs had similar intensities and shapes as CMBs on all three planes. They were usually veins that appeared with mimic shapes in a few slices. FNs were generally too small or vague CMBs to be detected. For example, the FN in Fig. 4 shows up as small spherical shapes even in the sagittal and coronal planes because of its tiny size; the shapes do not have the usual appearances of CMBs on those two planes.

The results of the jointly scaled threshold values of the three detection networks regarding sensitivity and FP_{avg} are shown in Fig. 5a,b. The FROC curves in Fig. 5c prove that the ensemble of multidimensional detection networks leads to evident performance improvement.

TPE-Det vs. Other Detection Models

Table 3 shows the testing results of the TPE-Det vs. other detection models. The threshold values for the TPE-Det were the same as in Table 2. The proposed TPE-Det achieved higher
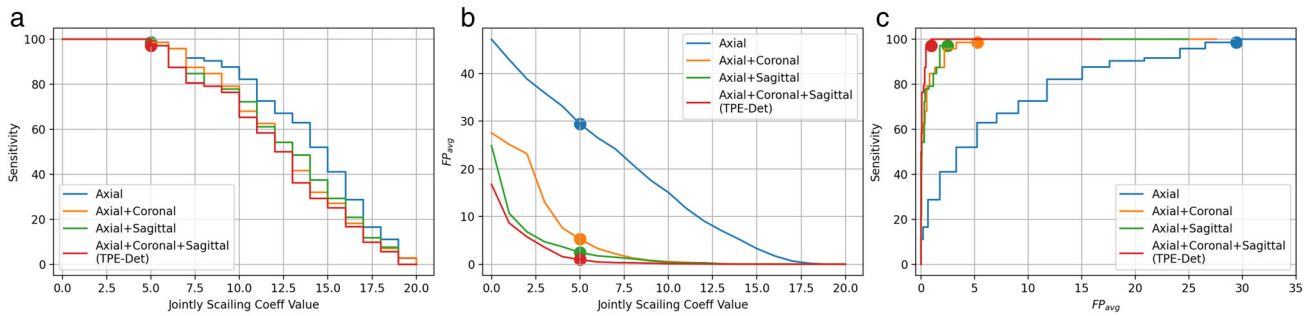
FIGURE 5: The impact of the jointly scaling coefficient value on the detection performance of each ensemble network. (a) Sensitivity vs. threshold scaling coefficients. (b) FP_{avg} vs. threshold scaling coefficients. (c) Sensitivity vs. FP_{avg}. Each circle on the curves means the set of the confidence threshold values in Table 2. The testing was performed on the DS1 testing set.

TABLE 3. The Comparison of the Proposed TPE-Det Against the Other Detection Models on DS1

Detection Model		Sensitivity (%)	FP_{avg}	Precision (%)
3D Faster R-CNN	First stage	93.33	12.65	19.43
	Second stage	100.00	-	-
	Overall	93.33	1.52	66.28
Al-masni et al[10]	First stage	93.33	53.69	5.36
	Second stage	94.28	-	-
	Overall	88.00	1.75	61.11
Kim et al[16]		94.66	8.82	25.64
TPE-Det		96.05	0.88	76.76

TABLE 4. The Testing Result on the Subjects From DS2

	DS2			
	CMBs Present			CMBs Absent
Ensemble Network	Sensitivity (%)	FP_{avg} (Improvement rate)	Precision (%)	FP_{avg} (Improvement rate)
Axial	92.52	24.36 (baseline)	8.78	23.76 (baseline)
Sagittal	91.22	90.95 (NA)	2.50	89.59 (NA)
Coronal	88.52	216.12 (NA)	1.03	260.43 (NA)
Axial + Coronal	90.00	3.34 (86.29%)	40.58	3.29 (86.15%)
Axial + Sagittal	88.52	1.07 (95.61%)	67.71	1.29 (94.57%)
Axial + Sagittal + Coronal (TPE-Det)	85.03	0.55 (97.74%)	79.67	0.67 (97.18%)

sensitivity with fewer FP_{avg} and significantly higher precision ($P < 0.02$) than the other three models. Even though the 3D Faster R-CNN maintained high sensitivity in the second stage, it missed some TPs in the first candidate detection stage.

Testing on the Other Dataset With Different Acquisition Parameters (DS2)

The results of the subjects with CMBs and subjects without CMBs from DS2 are reported in Table 4.

TABLE 5. The Comparison Between the Proposed Approach With the Latest Studies on CMBs Detection

References	Method	Data Subjects/CMBs	Resolution (mm³)	Performance		
				Sensitivity (%)	FP$_{avg}$	Precision (%)
Indirect comparison (Reported)						
Dou et al[7]	First stage: 3D-FCN / Second stage: 3D-CNN	3.0 T, 320/1149	0.45 × 0.45 × 2.00 (spacing 1)	91.45	2.74	44.31
Liu et al[8]	First stage: 3D-FRST / Second stage: 3D-ResNet	1.5 T and 3.0 T, 220/1641	0.45 × 0.57 × 2.00 and 0.50 × 0.50 × 1.2	95.24	1.60	70.90
Chen et al[9]	First stage: 2D-FRST / Second stage: 3D-ResNet	7.0 T, 73/2835	0.50 × 0.50 × 2.00	94.69	11.58	71.98
Myung et al[11]	First stage: YOLO / Second stage: CSF filtering	3.0 T, 186/1716	0.68 × 0.68 × 2.20 and 0.63 × 0.63 × 2.00	66.90	2.15	72.76
Kuijf et al[30]	3D-FRST	7.0 T, 18/66	0.35 × 0.35 × 0.30	71.20	17.17	13.20
Wang et al[14]	2D-DenseNet	20/68847	-	97.78	11.8	97.65
Hong et al[15]	2D-ResNet-50	10/4287	-	95.71	3.4	99.18
Direct comparison (Implemented)						
Al-masni et al[10]	First stage: YOLO / Second stage: 3D-CNN	3.0 T, 128/367	DS1: 0.50 × 0.50 × 2.00	88.00	1.75	61.11
Kim et al[16]	3D U-Net + RPN			94.66	8.82	25.64
Proposed work	Implemented 3D Faster R-CNN			93.33	1.52	66.28
	Axial			97.26	29.43	9.49
	Axial + Coronal			97.26	4.41	40.49
	Axial + Sagittal			96.05	2.54	53.80
	Axial + Sagittal + Coronal (TPE-Det)			**96.05**	**0.88**	**76.76**
	TPE-Det (only testing)	3.0 T, 79/148	DS2: 0.50 × 0.50 × 3.00	**85.03**	**0.55**	**79.67**

First, the sensitivity, FP_{avg}, and precision values were evaluated on the DS2 subjects with CMBs utilizing the models trained on DS1. Compared to the results on DS1 (Table 2), fewer candidates per subject were detected from each detector due to the different characteristics of the images, leading to lower sensitivities and FPs in the ensemble results. Sensitivity decreased more than in DS1 when the proposed method ensembled coronal and sagittal detection networks with axial networks. These results are likely to be attributable to the lower resolution of the sagittal and coronal images in DS2 than in DS1 ($0.5 \times 0.5 \times 3.0$ mm^3 in DS2 vs. $0.5 \times 0.5 \times 2.0$ mm^3 in DS1). Nevertheless, the precision values were even higher than in Table 2.

Only FP_{avg} values were then evaluated on the DS2 subjects without CMBs. Ensembling the detectors of each plane effectively reduced FPs. All the FP values from the ensemble models were smaller than in Table 2. These results are most likely to be related to the images' different characteristics from their different scan parameters.

Discussion

Reducing FPs while preserving high sensitivity is an important issue since estimating the most accurate number of CMBs in a subject could be decisive for clinicians in making an appropriate diagnosis. The prevalence of CMBs in the general population has been reported from 3% to 27%.[31,32] CMBs are known to be associated with cerebral amyloid angiopathy (CAA) and hypertension.[33] Besides, the presence of multiple CMBs on MR images is known to be associated with an increased risk of hemorrhagic complications in acute ischemic stroke.[34] Therefore, evaluation of even a small number of CMBs is important to make a diagnosis, predict prognosis, and make a decision to prescribe medicine.[35,36] To estimate the number of the lesions as accurately as possible, most relevant works employed additional networks that exploited 3D information in their second stages to reduce FPs, ending up with non-end-to-end structures.[7,9,10,37] This study used a single-stage multiplanar ensemble detection model for FP reduction. The motivation for this approach was based on the fact that clinicians normally use these multiple plane views during manual detection. It is noteworthy that integrating sagittal and coronal detection results significantly increased precision ($P < 0.001$) from when only CMBs were detected on the axial plane. This clearly shows the importance of the information in the sagittal and coronal planes for FP reduction. Compared with relevant studies,[7–11,15,30,37] TPE-Det accurately detected CMBs with a notably small number of FP_{avg}. In addition, this study performed testing on the subjects without CMBs to evaluate the proposed models' FP reduction ability. It was revealed that TPE-Det detected a similar number of FPs on the data from the same scan parameters of the training set, while fewer FPs were generated from the data with different acquisition protocols. The different image characteristics from the training data, especially the lower resolution, are expected to result in fewer detected candidates. Al-masni et al[12] have shown performance degradation in low-resolution data. Thus, performance improvements can be anticipated in the sagittal and coronal planes if the proposed networks detect CMBs on 3D volumetric MRI data with a higher spatial resolution.

Most of the earlier works that achieved notable performances adopted two-stage frameworks[7,9–11,37]: 1) the CMB candidate detection stage and 2) the FP reduction stage. The performances were achieved by employing 3D networks in their second FP reduction stages after their first candidate detection stages. Additionally, the impressive performances achieved by the other single-stage frameworks[14,15] could be attributed to the enormous number of microbleeds samples per subject compared to other studies. Note that 68,847 CMBs and over 56 million non-CMBs were derived from only 20 subjects,[14] and 4287 CMBs were extracted from only 10 subjects.[15] These two examples clearly show the impact of data distribution on the performance of this task. When this study adjusted the thresholds of TPE-Det to achieve a higher sensitivity (98.48%) than the 97.78% of Wang et al,[14] the FP_{avg} generated from TPE-Det was 1.58, which is fewer than the 11.8 of Wang et al.[14] TPE-Det could perform well with a smaller dataset by exploiting the ensemble method. Even though EfficientDet was employed to maximize the efficiency in this study, it is expected that any 2D object detection network could serve as a detector for each plane. Therefore, the proposed framework is expected to be employed on any dataset without any difficulties, disregarding specific deep learning network structures.

This article presents a direct comparison with other 3D detection models that were implemented on the same dataset. Even though other works in the literature presented methods of automatic CMBs detection that employed 3D networks,[7–9,12,16] most of the frameworks were not implemented as single networks nor end-to-end structures, which may cause difficulties when seeking to incorporate them into routine clinical use. Therefore, for comparison, this study built the 3D Faster R-CNN as an end-to-end structure for CMBs detection and compared the testing performance. Furthermore, this study directly compared TPE-Det against the most recently presented deep learning detection models with two-stage[10] and single-stage structures[16] on the same dataset. As a result, the proposed TPE-Det achieved significantly higher precision ($P < 0.02$) with higher sensitivity and a lower FP_{avg}. This result may be explained by the main benefit of ensemble learning models: better generalization performance.[38] After network training, the networks that provided the lowest training and validation errors were chosen for inference. In this case, because of the small amount of available data used for training, an undesirable model with poor generalizability could be selected if it gave low training and validation errors on the few training samples.[39]

Therefore, the ensemble of the axial, sagittal, and coronal detection networks may lead to a training algorithm that reduces the risk of being an unreliable detector compared to a single 3D-Faster R-CNN.

In reality, most medical institutions use their own images generated from different acquisition protocols for diagnosis. Note that, as shown in Table 5, every research group employed different datasets from different scan parameters, such as resolution. Therefore, this article assumed that the model's performance on the other unseen data from different acquisitions should be considered in order to develop a feasible solution for automatic CMB detection. This article additionally collected a dataset from another institution (DS2) and tested the proposed models on it without retraining. As shown in the results, lower sensitivities and fewer FP_{avg} were obtained than DS1. However, the precision values were almost equal or even higher. This result may be explained by the fact that even in different resolutions, microbleeds appear as similar shapes within both anisotropic resolution datasets.

The data distribution has a huge impact on the performance of this task. The precision was not higher than 80%, which could seem not enough for clinical routine. The reason is anticipated that the dataset in this study contains only 2.49 CMBs per subject. If TPE-Det achieved an even sensitivity of 100% with only FP_{avg} of 0.77 in the DS1 testing set, the precision would be 79.8%, which is still lower than 80%. The number of CMBs per subject could be very different by dataset. For example, the dataset of Wang et al[8] has 3,422.35 TPs per subject, and Hong et al[15] have 428.7 CMBs from only 10 subjects. On the other hand, the number of FPs was almost the same in both cases of the subjects with CMBs and without CMBs in the presented experiment. This study expects the reason to be that most FPs are veins, which are natural human brain structures. Thus, the precision could dramatically vary by which dataset was utilized for this task. Note that Hong et al[15] achieved a precision of 99.18% with a sensitivity of 95.71% and FP_{avg} of 3.4, while TPE-Det achieved a precision of 76.76% with a sensitivity of 96.05% and FP_{avg} of 0.88. If the proposed method maintained the same sensitivity and FP_{avg} on the dataset of Wang et al,[8] the precision would be 99.97%.

Limitations

First is the limited size of labeled data that this study utilized, which is a general problem in deep learning methods for medical image analysis. The dataset is relatively small compared to other up-to-date object detection tasks based on deep learning approaches. Also, TPE-Det was built on the datasets acquired with single field strength, 3 T, generated from a single vendor (Siemens Healthineers). Accordingly, including additional subjects with more CMBs will improve the model's generalizability in further study. Interestingly, TPE-Det was able to detect several unlabeled tiny CMBs that were missed by the raters in the

labeling process. Thus, the proposed approach could be useful for collecting additional labeled data with clinicians by utilizing active learning.[40]

Second, the proposed method was only constructed on the MR images that were generated by a single set of scan parameters (DS1). Most of the relevant works' datasets have different in-plane resolutions.[7–11,15,30,37] The appearance of CMBs can vary in intensity, shape, and size depending on the resolution. As TPE-Det detects microbleeds not only from the axial plane but also from the sagittal and coronal planes, the slice thickness also plays a crucial role in deciding the CMBs' unique characteristics. In this case, each detection model cannot help being biased toward its dataset's specific set of parameters. As shown in the detection results on the unseen data DS2, the TPE-Det detected CMBs with smaller confidence scores and resulted in lower sensitivity and FP_{avg}. From the results, it could be deduced that the resolution, slice thickness, and type of scanner could play an essential role in CMBs detection since DS2 has almost the same scan parameters such as TR (DS1: 27 msec and DS2: 28 msec), TE (DS1 and DS2: 20 msec), and FA (DS1 and DS2: 15°) as DS1, except for those protocols. It is expected that an entirely different resolution from DS1 would substantially degrade the detection performance of TPE-Det, which was trained on DS1 only. Therefore, in the future, collecting additional data from different scan parameters and utilizing them for model training could lead to building more robust CMB detection models.

Third, the proposed approach requires three different times for training the networks on axial, sagittal, and coronal images. Even though the ensemble technique could lead to a single-stage structure with an appealing performance and improve the model's generalizability, the multiple training procedure may be time-consuming. In addition, calculating the distances between all candidates from the three detectors for selecting the ensemble results could be complicated. Therefore, it will be interesting to handle this issue in future work to merge the three detection networks during the training stage. Each network could learn robust contextual features from each plane.

Conclusion

This study demonstrated that the ensembling of multidimensional networks reduces the FPs better than the case of only detecting CMBs on a particular plane or using a 3D network, suggesting that this new approach could potentially increase the benefits of detecting and evaluating the lesions in the clinic.

REFERENCES

1. Martinez-Ramirez S, Greenberg SM, Viswanathan A. Cerebral microbleeds: Overview and implications in cognitive impairment. Alzheimers Res Ther 2014;6(3):33.

2. Haller S, Vernooij MW, Kuijer JPA, Larsson EM, Jager HR, Barkhof F. Cerebral microbleeds: Imaging and clinical significance. Radiology 2018;287(1):11-28.

3. Koennecke HC. Cerebral microbleeds on MRI - prevalence, associations, and potential clinical implications. Neurology 2006;66(2):165-171.

4. Lee J, Sohn EH, Oh E, Lee AY. Characteristics of cerebral microbleeds. Dement Neurocogn Disord 2018;17(3):73-82.

5. Charidimou A, Werring DJ. Cerebral microbleeds: Detection, mechanisms and clinical challenges. Future Neurol 2011;6(5):587-611.

6. Kargiotis O, Safouris A, Magoufis G, et al. Cerebral microbleeds: Incidence, imaging characteristics, common and uncommon causes. J Neurosonol Neuroimag 2018;10(2):80-94.

7. Dou Q, Chen H, Yu LQ, et al. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. IEEE Trans Med Imaging 2016;35(5):1182-1195.

8. Liu SF, Utriainen D, Chai C, et al. Cerebral microbleed detection using susceptibility weighted imaging and deep learning. Neuroimage 2019;198:271-282.

9. Chen YC, Villanueva-Meyer JE, Morrison MA, Lupo JM. Toward automatic detection of radiation-induced cerebral microbleeds using a 3D deep residual network. J Digit Imaging 2019;32(5):766-772.

10. Al-masni MA, Kim WR, Kim EY, Noh Y, Kim DH. Automated detection of cerebral microbleeds in MR images: A two-stage deep learning approach. Neuroimage-Clin 2020;28:102464.

11. Myung MJ, Lee KM, Kim HG, et al. Novel approaches to detection of cerebral microbleeds: Single deep learning model to achieve a balanced performance. J Stroke Cerebrovasc 2021;30(9):105886.

12. Chung H, Kang KM, Al-Masni MA, et al. Stenosis detection from time-of-flight magnetic resonance angiography via deep learning 3D squeeze and excitation residual networks. IEEE Access 2020;8:43325-43335.

13. Al-masni MA, Kim W-R, Kim EY, Noh Y, Kim D-H. A two cascaded network integrating regional-based YOLO and 3D-CNN for cerebral microbleeds detection. *42nd annual international conference of the IEEE engineering in Medicine & Biology Society (EMBC)*. Montreal, QC, Canada: IEEE; 2020. p 1055-1058.

14. Wang SH, Tang CS, Sun JD, Zhang YD. Cerebral micro-bleeding detection based on densely connected neural network. Front Neurosci-Switz 2019;13:422.

15. Hong J, Cheng H, Zhang YD, Liu J. Detecting cerebral microbleeds with transfer learning. Mach Vision Appl 2019;30(7–8):1123-1133.

16. Kim JH, Al-masni MA, Lee HJ, Choi YS, Kim DH. A single-stage detector of cerebral microbleeds using 3D feature fused region proposal network (FFRP-net). *2022 IEEE 4th international conference on artificial intelligence circuits and systems (AICAS)*. Incheon, Republic of Korea: IEEE; 2022. p 1-4.

17. Yu JZ, Yang BH, Wang J, Leader J, Wilson D, Pu JT. 2D CNN versus 3D CNN for false-positive reduction in lung cancer screening. J Med Imaging 2020;7(5):051202.

18. Milletari F, Navab N, Ahmadi SA. V-net: Fully convolutional neural networks for volumetric medical image segmentation. *Fourth international conference on 3D Vision (3DV)*. Stanford, CA, USA: IEEE; 2016. p 565-571.

19. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929-1958.

20. Sundaresan V, Zamboni G, Rothwell PM, Jenkinson M, Griffanti L. Triplanar ensemble U-net model for white matter hyperintensities segmentation on MR images. Med Image Anal 2021;73:102184.

21. Sudre CH, Wijnen KV, Dubost F, et al. Where is VALDO - vascular lesions detection and segmentation challenge at MICCAI 2021. arXiv 2021;1-44. arXiv preprint arXiv:220807167.

22. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol 2013;12(8):822-838.

23. van den Heuvel TLA, van der Eerden AW, Manniesing R, et al. Automated detection of cerebral microbleeds in patients with traumatic brain injury. Neuroimage-Clin 2016;12:241-251.

24. Tatsumi S, Shinohara M, Yamamoto T. Direct comparison of histology of microbleeds with postmortem MR images. Cerebrovasc Dis 2008; 26(2):142-146.

25. Greenberg SM, Vernooij MW, Cordonnier C, et al. Cerebral microbleeds: A guide to detection and interpretation. Lancet Neurol 2009;8(2):165-174.

26. Smith SM. Fast robust automated brain extraction. Hum Brain Mapp 2002;17(3):143-155.

27. Tan M, Pang R, Le QV. Efficientdet: Scalable and efficient object detection. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. Seattle, WA, USA: IEEE; 2020. p 10781-10790.

28. Ren SQ, He KM, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. Adv Neur In 2015;28:91-99.

29. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: Learning dense volumetric segmentation from sparse annotation. *International conference on medical image computing and computer-assisted intervention (MICCAI), medical image computing and computer-assisted intervention – MICCAI 2016*. Athens, Greece: Springer, Cham; 2016. p 424-432.

30. Kuijf HJ, de Bresser J, Geerlings MI, et al. Efficient detection of cerebral microbleeds on 7.0 T MR images using the radial symmetry transform. Neuroimage 2012;59(3):2266-2273.

31. Poels MM, Vernooij MW, Ikram MA, et al. Prevalence and risk factors of cerebral microbleeds: An update of the Rotterdam scan study. Stroke 2010;41(10 Suppl):S103-S106.

32. Romero JR, Preis SR, Beiser A, et al. Risk factors, stroke prevention treatments, and prevalence of cerebral microbleeds in the Framingham heart study. Stroke 2014;45(5):1492-1494.

33. Fazekas F, Kleinert R, Roob G, et al. Histopathologic analysis of foci of signal loss on gradient-echo T2*-weighted MR images in patients with spontaneous intracerebral hemorrhage: Evidence of microangiopathy-related microbleeds. AJNR Am J Neuroradiol 1999;20(4):637-642.

34. Dannenberg S, Scheitz JF, Rozanski M, et al. Number of cerebral microbleeds and risk of intracerebral hemorrhage after intravenous thrombolysis. Stroke 2014;45(10):2900-2905.

35. Bailey RD, Hart RG, Benavente O, Pearce LA. Recurrent brain hemorrhage is more frequent than ischemic stroke after intracranial hemorrhage. Neurology 2001;56(6):773-777.

36. Sperling RA, Jack CR Jr, Black SE, et al. Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: Recommendations from the Alzheimer's Association research roundtable workgroup. Alzheimers Dement 2011;7(4):367-385.

37. Liu CC, Qi JY. Higher SNR PET image prediction using a deep learning model and MRI image. Phys Med Biol 2019;64(11):115004.

38. Ganaie M, Hu M. Ensemble deep learning: A review. arXiv Preprint: arXiv210402395 2021;1-47.

39. Dietterich TG. Ensemble methods in machine learning. Lect Notes Comput Sc 2000;1857:1-15.

40. Settles B. *Active learning literature survey, Report 1648*. Madison, WI, United States: University of Wisconsin-Madison, Department of Computer Sciences; 2009.