BioMedical Engineering OnLine

RESEARCH

Open Access

Deep learning based on susceptibility-weighted MR sequence for detecting cerebral microbleeds and classifying cerebral small vessel disease

Ruizhen Wu[1], Huaqing Liu[2], Hao Li[3], Lifen Chen[4], Lei Wei[1], Xuehong Huang[1], Xu Liu[1], Xuejiao Men[1], Xidan Li[2], Lanqing Han[2], Zhengqi Lu[1] and Bing Qin[1*]

*Correspondence:
qinb5@mail.sysu.edu.cn

[1] Department of Neurology, The Third Affiliated Hospital of Sun Yat-Sen University, Sun Yat-Sen University, No. 600 Tianhe Road, Guangzhou 510630, People's Republic of China
[2] Center for Artificial Intelligence in Medicine, Research Institute of Tsinghua, Pearl River Delta, Tsinghua University, No. 98 Xiangxue 8Th Road, Guangzhou 510700, People's Republic of China
[3] Department of Neurology, Maoming People's Hospital, No.101 Weimin Road, Maoming 525000, People's Republic of China
[4] Department of Neurology, the First Affiliated Hospital of SHANTOU University Medical College, Shantou University, No. 57 of Changping Road, Shantou 515041, People's Republic of China

Abstract

Background: Cerebral microbleeds (CMBs) serve as neuroimaging biomarkers to assess risk of intracerebral hemorrhage and diagnose cerebral small vessel disease (CSVD). Therefore, detecting CMBs can evaluate the risk of intracerebral hemorrhage and use its presence to support CSVD classification, both are conducive to optimizing CSVD management. This study aimed to develop and test a deep learning (DL) model based on susceptibility-weighted MR sequence (SWS) to detect CMBs and classify CSVD to assist neurologists in optimizing CSVD management. Patients with arteriolosclerosis (aSVD), cerebral amyloid angiopathy (CAA), and cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) treated at three centers were enrolled between January 2017 and May 2022 in this retrospective study. The SWSs of patients from two centers were used as the development set, and the SWSs of patients from the remaining center were used as the external test set. The DL model contains a Mask R-CNN for detecting CMBs and a multi-instance learning (MIL) network for classifying CSVD. The metrics for model performance included intersection over union (IoU), Dice score, recall, confusion matrices, receiver operating characteristic curve (ROC) analysis, accuracy, precision, and F1-score.

Results: A total of 364 SWS were recruited, including 336 in the development set and 28 in the external test set. IoU for the model was 0.523 ± 0.319, Dice score 0.627 ± 0.296, and recall 0.706 ± 0.365 for CMBs detection in the external test set. For CSVD classification, the model achieved a weighted-average AUC of 0.908 (95% CI 0.895–0.921), accuracy of 0.819 (95% CI 0.768–0.870), weighted-average precision of 0.864 (95% CI 0.831–0.897), and weighted-average F1-score of 0.829 (95% CI 0.782–0.876) in the external set, outperforming the performance of the neurologist group.

Conclusion: The DL model based on SWS can detect CMBs and classify CSVD, thereby assisting neurologists in optimizing CSVD management.

Keywords: Cerebral small vessel disease, Cerebral microbleeds, Susceptibility-weighted MR Sequence, Deep learning

Background

Cerebral small vessel disease (CSVD) is a chronic and progressive vascular disease affecting capillaries, arterioles, and small veins supplying brain deep structures, often leading to cerebral hemorrhage, dementia, and stroke [1, 2]. CSVD comprises arteriolosclerosis (aSVD) related to age or vascular risk factors, cerebral amyloid angiopathy (CAA) caused by vascular deposition of β-amyloid, and cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) caused by *NOTCH3* gene mutations which encode a vascular smooth muscle transmembrane protein involved in vascular development and smooth muscle cell differentiation [3, 4]. The current treatment strategy for CSVD typically involves secondary stroke prevention [5, 6], focusing on prevalent types, such as aSVD and CAA. The role of CADASIL, a typically inherited small vessel disease, in stroke prevention should not be ignored [7]. However, there is controversy regarding whether the secondary prevention measures for stroke are suitable for these types of CSVD; there is no uniform treatment standard [3]. For example, antithrombotic therapy is generally not recommended for treating CAA because amyloid angiopathy is an independent risk factor for hemorrhage [8, 9]. Patients with CADASIL should avoid using antiplatelet drugs to prevent an increased risk of hemorrhage [10, 11], antiplatelet therapy should not be withheld for patients with aSVD, and may even be beneficial for the prevention of lacunar stroke [6, 12]. Optimizing the management of CSVD depends on the individual patient's intracerebral hemorrhage risk and the type of CSVD. Cerebral microbleeds (CMBs), which manifest as small, oval, hypointense lesions on brain susceptibility-weighted MR sequences [SWS, including T2-star-weighted angiography (SWAN) and susceptibility-weighted imaging (SWI)], serve as indicators for evaluating the risk of intracerebral hemorrhage [13, 14]. CMBs also serve as neuroimaging biomarkers for CSVD and play a critical role in the diagnosis of CSVD, especially CAA [13, 15–17]. Hence, accurate identification of CMBs is useful for evaluating the risk of intracerebral hemorrhage; facilitating classification of CSVD based on the presence of CMBs, both of which are beneficial for optimizing CSVD management.

Recently, there has been a keen interest in the application of deep learning (DL) to medical images. Several studies indicate that DL can reliably detect lesions and diagnose diseases, such as detecting primary bone tumors on radiographs and diagnosing muscular dystrophies using MRI [18, 19]. The clinical application of these models may improve the reliability and accuracy of lesion assessment or disease diagnosis, potentially leading to improved diagnostics and better treatment. Therefore, in this study, we used SWS data from patients with three types of CSVD collected from three independent centers to develop and test an end-to-end, two-task DL model that can detect CMBs and classify CSVD.

Results

Patients and datasets characteristics

This study enrolled 364 patients with CSVD, of whom 336 were included in the development set and 28 in the external test set. The development set was comprised of 203 aSVD patients, 99 CAA patients, and 34 CADASIL patients. The mean

age (±standard deviation) of the development set patients was 64.88 years±6.71 [aSVD (65.34 years±6.12); CAA (64.41 years±6.98); CADASIL (63.56 years±8.93)], and 34.82% (n=117) of the development set patients were female [aSVD (76/203, 37.44%); CAA (28/99, 28.29%); CADASIL (13/34, 38.24%)]. The external test set consisted of 14 aSVD patients, 7 CAA patients, and 7 CADASIL patients The mean age (±standard deviation) of the external test set patients was 64.18 years±6.83 [aSVD (64.36 years±6.05); CAA (66.57 years±10.16); CADASIL (61.42 years±3.41)], and 32.14% (n=9) of the external test set patients were female [aSVD (3/14, 21.43%); CAA (3/7, 42.86%); CADASIL (3/7, 42.86%)] (Table 1). The SWS in the development set came from 6 different MRI scanners, and the SWS in the external test set came from another MRI scanner. The MRI scanner in the external test set differed from any scanner in the development set. MRI scanners and SWS parameters are summarized in supplementary materials (Additional file 1: Table S1).

CMBs detection

The reproducibility of the manual annotation of CMBs was calculated with an IoU of 0.820±0.039 for bounding box placement, and a Dice score of 0.900±0.024 for segmentation.

On the internal test set of the development dataset, the performance of the model for detecting CMBs presented an IoU of 0.594±0.258 for bounding box placement, and a Dice score of 0.709±0.230 for segmentations. The proportion of the true CMBs predicted by the model had a recall of 0.879±0.228 in the internal test set of the development set. On the external test set, the performance of the model in detecting CMBs presented an IoU of 0.523±0.319 for bounding box placement, and a Dice score of 0.627±0.296 for segmentations. The proportion of the true CMBs predicted by the model had a recall of 0.706±0.365 in the external test set. Pearson analysis to evaluate the correlations of CMBs areas predicted by the model and the ground truth indicated correlations of 0.832 and 0.797 in the internal test set of the development set and external test set, respectively (Fig. 1). Figure 2 shows an example of CMB detection.

Table 1 Patients and datasets characteristics

Characteristic	Overall	Development set	External testing set
Number of patients	364	336	28
Age (y±SD)*	64.83±6.72	64.88±6.71	64.18±6.83
Female sex (%)	126 (34.62%)	117 (34.82%)	9 (32.14%)
CSVD type			
aSVD (%)	217 (59.62%)	203 (61.42%)	14 (50.00%)
CAA (%)	106 (29.12%)	99 (29.46%)	7 (25.00%)
CADASIL (%)	41 (11.26%)	34 (10.12%)	7 (25.00%)

* mean±standard. The numbers in parentheses are the percentages. Unless otherwise specified the data indicate the number of participants. The susceptibility-weighted MR sequences (SWS) of patients from the Third Affiliated Hospital, Sun Yat-sen University (SYSUTH) and Maoming People's Hospital (MMPH) served as the development set, and the SWS of patients from the First Affiliated Hospital of SHANTOU University Medical College (STUMFH) served as the external test set for testing the geographic performance of the model. The tenfold cross-validation was used to evaluate the stability of the deep learning (DL) model

y year, *SD* standard deviation, *CSVD* cerebral small vessel disease, *aSVD* arteriolosclerosis, *CAA* cerebral amyloid angiopathy, *CADASIL* cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy

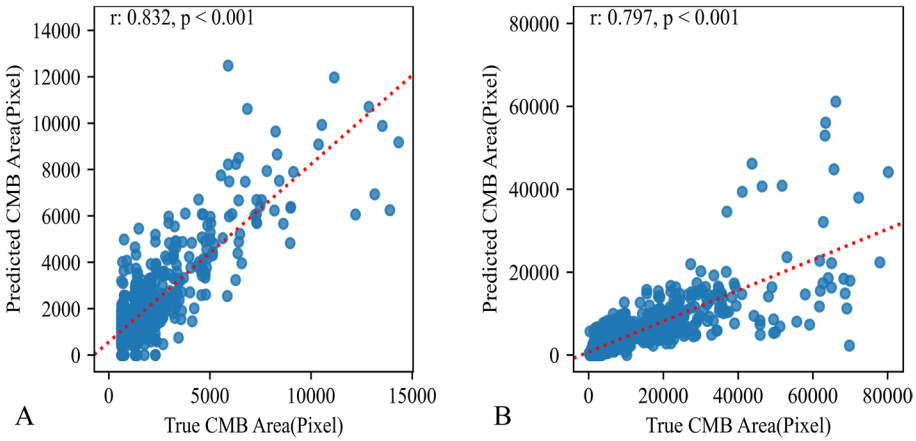

Fig. 1 Pearson analysis for evaluating the model performance in detecting CMB. **A** The analysis yielded a correlation coefficient of 0.832 for the internal test set of the development set. **B** The analysis yielded a correlation coefficient of 0.797 for the external test set. *CMBs* cerebral microbleed

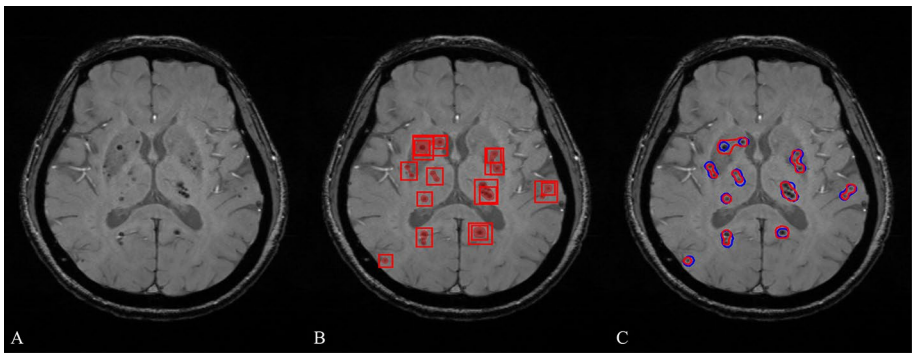

Fig. 2 **An** example of CMBs detection. **A** Original SWS image. **B** Bounding box placement of CMBs predicted by the Mask R-CNN with a probability above 0.6. **C** Comparison between the CMBs region identified by the model and the groundtruth. The probability mask generated by the Mask-RCNN was binarized with a threshold of 0.6. The contours of binarized mask and ground truth were visualized in red and blue, respectively. *CMBs* cerebral microbleed; *SWS* susceptibility-weighted MR sequences

CSVD classification

The inter-rater agreement among committee members was presented with a Fleiss Kappa coefficient of 0.735, ensuring the consistency and accuracy of the annotations.

On the internal test set of the development set, the performance of the model for classifying CSVD into aSVD, CAA, and CADASIL across tenfold cross-validation presented a mean weighted-average AUC of 0.865 (95% CI 0.841–0.884), a mean accuracy of 0.732 (95% CI 0.718–0.746), a mean weighted-average precision of 0.755 (95% CI 0.743–0.768), and a mean weighted-average F1-score of 0.733 (95% CI 0.719–0.747) (Table 2). ROC analysis showed a mean AUC for categorizing CSVD into aSVD, CAA, and CADASIL of 0.867 (95% CI 0.805–0.929), 0.863 (95% CI 0.802–0.924), and 0.834 (95% CI 0.773–0.895), respectively, in the internal test set of the development set (Fig. 3A). On the external test set, the performance of the model at classifying CSVD into aSVD, CAA, and CADASIL across tenfold cross-validation presented a mean weighted-average AUC of 0.899 (95% CI 0.884–0.914), a mean accuracy of 0.717 (95% CI 0.704–0.731), a mean

Table 2 Metrics in CSVD classification for the model and the neurologist group

Metrics	Internal test set	External test set		
	Mean of 10-told cross-validation	Mean of 10-told cross-validation	Ensemble model	Neurologist group
Weighted-average AUC	0.865 (0.841–0.884)	0.899 (0.884–0.914)	0.908 (0.895–0.921)	–
Accuracy	0.732 (0.718–0.746)	0.717 (0.704–0.731)	0.819 (0.768–0.870)	0.759 (0.740–0.779)
Weighted-average precision	0.755 (0.743–0.768)	0.717 (0.702–0.733)	0.864 (0.831–0.897)	0.767 (0.747–0.786)
Weighted-average F1-score	0.733 (0.719–0.747)	0.705 (0.691–0.719)	0.829 (0.782–0.876)	0.762 (0.742–0.781)

The numbers in parentheses are the 95% confidence interval

CSVD cerebral small vessel disease, *AUC* area under the receiver operating characteristic curves

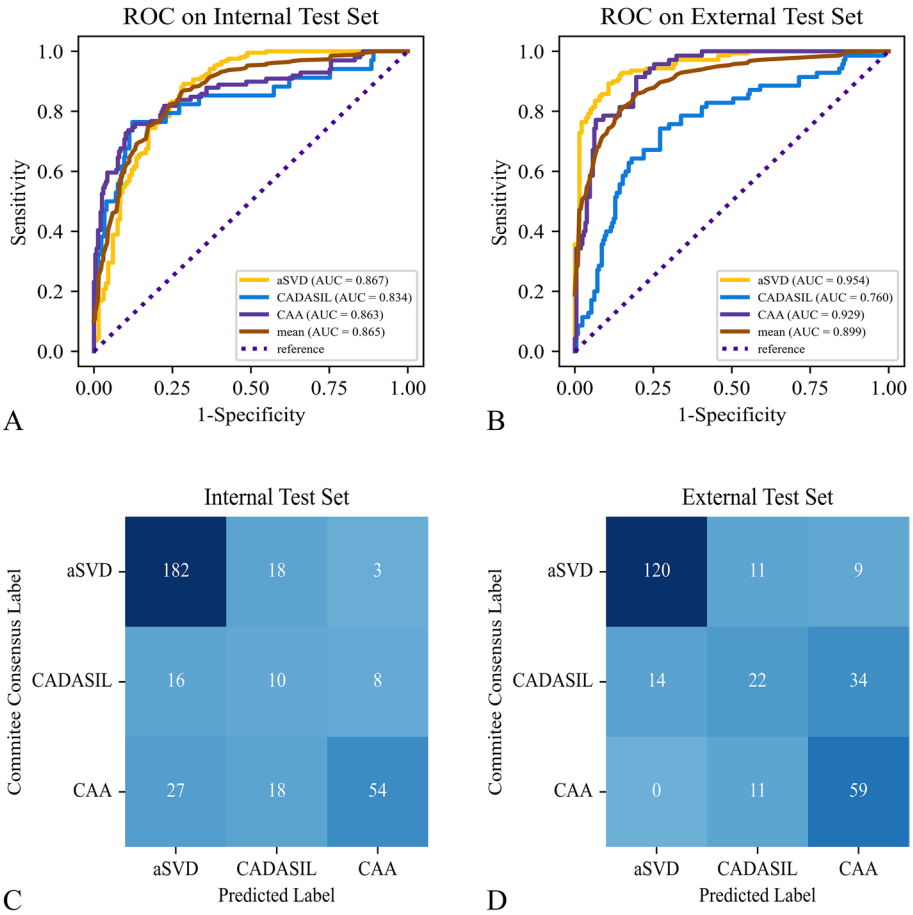

Fig. 3 ROC and confusion matrices for evaluating the model performance in tenfold cross-validation in classifying CSVD. **A**, **B** ROC for model performance in tenfold cross-validation on the internal test set of the development set and the external test set, respectively. **C**, **D** Confusion matrices for model performance in tenfold cross-validation on the internal test set of the development set and the external test set, respectively. *ROC* receiver operating characteristic, *CSVD* cerebral small vessel disease

weighted-average precision of 0.717 (95% CI 0.702–0.733), and a mean weighted-average F1-score of 0.705 (95% CI 0.691–0.719) (Table 2). ROC analysis showed a mean AUC for categorizing CSVD to aSVD, CAA, and CADASIL of 0.954 (95% CI 0.946–0.970), 0.929

(95% CI 0.916–0.940), and 0.760 (95% CI 0.734–0.788), respectively, in the external test set (Fig. 3B). The results for each run of tenfold cross-validation of the model for CSVD classification are summarized in supplementary materials (Additional file 1: Table S2).

In the external test set, the performance of the ensemble model for classifying CSVD presented a weighted-average AUC of 0.908 (95% CI 0.895–0.921), an accuracy of 0.819 (95% CI 0.768–0.870), a weighted-average precision of 0.864 (95% CI 0.831–0.897), and a weighted-average F1-score of 0.829 (95% CI 0.782–0.876) (Table 2). ROC analysis showed an AUC for categorizing CSVD to aSVD, CAA, and CADASIL of 0.959 (95% CI 0.948–0.969), 0.946 (95% CI 0.932–0.959), and 0.769 (95% CI 0.731–0.805), respectively, in the external test set (Fig. 4). Comparatively, the performance of the group of neurologists in classifying CSVD into aSVD, CAA, and CADASIL on the external test set was inferior to that of the ensemble model, with an accuracy of 0.759 (95% CI 0.740–0.779), weighted-average precision of 0.767 (95% CI 0.747–0.786), and weighted-average F1-score of 0.762 (95% CI 0.742–0.781). The performance of each neurologist for CSVD classification on the external test set is summarized in supplementary materials (Additional file 1: Table S3 and Additional file 1: Figure S1). Figure 5 shows an example of an aSVD SWS superimposed on a Grad-CAM activation map.

Discussion

Optimizing CSVD management usually involves assessing the risk of intracerebral hemorrhage and determining the type. However, accurately evaluating these factors using MRI, which is the primary tool for examining and diagnosing CSVD, can be inconvenient. Firstly, it is difficult to visualize CMBs, indicators for evaluating the risk of intracerebral hemorrhage [14], on SWS because of their small size and the extensive co-occurrence of similar objects in the brain, such as blood vessels and calcifications, which can be misinterpreted as CMBs. Hence, the visual inspection of CMBs on SWS requires significant expertise and can be time-consuming and prone to errors. Secondly, differentiating CSVD based on MRI results can be challenging since the diagnosis of CSVD primarily relies on neuroimaging biomarkers and clinical presentation, which may not have clear-cut categorical distinctions [3]. In addition, comprehensive diagnosis of CSVD typically requires multiple MR sequences [13], and further biopsy or genetic testing is necessary for confirmation, such as in cases of CAA and CADASIL [10, 15]. Both can not only increase the clinical workload but

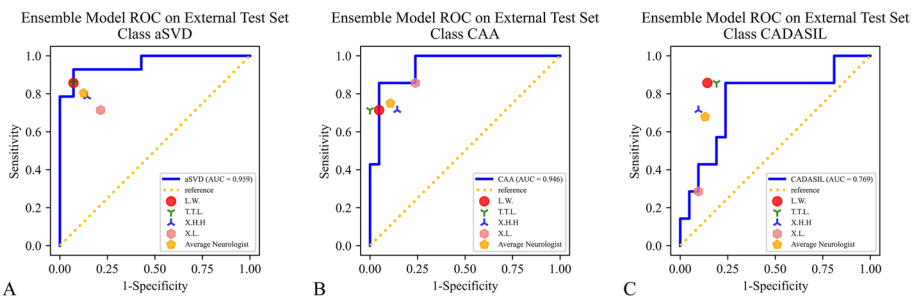

Fig. 4 ROC for evaluating the ensemble model performance in classifying CSVD on the external test set. Dots of different shapes represent individual neurologist performance and average neurologist performance. *ROC* receiver operating characteristic, *CSVD* cerebral small vessel disease

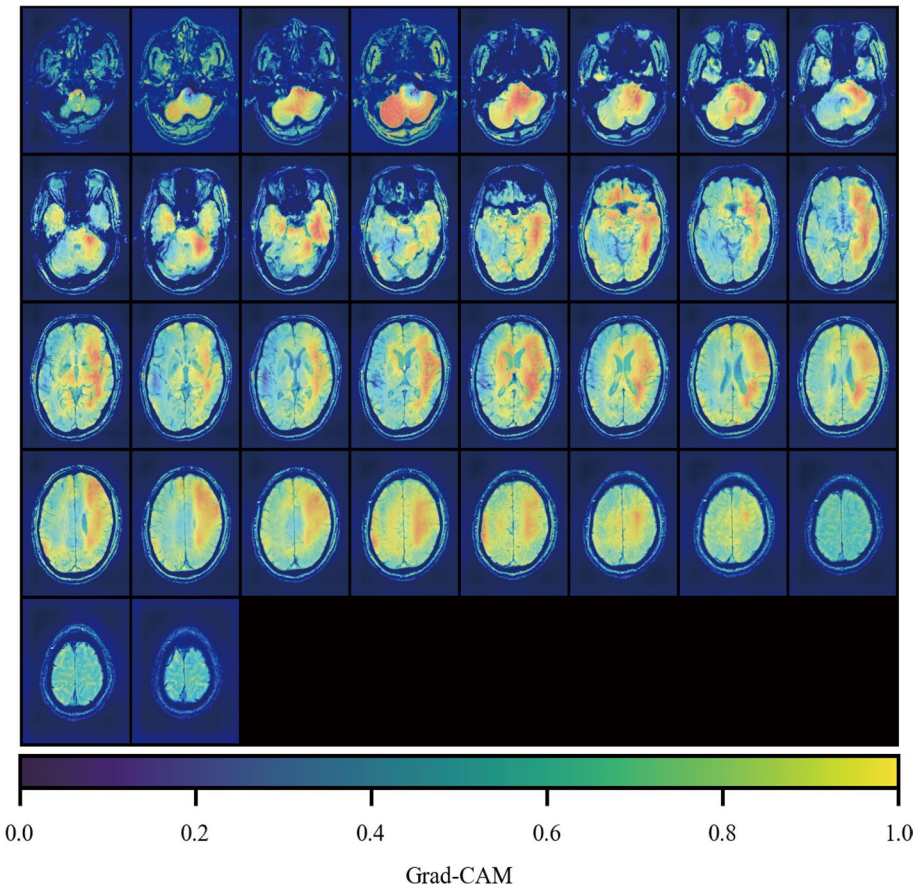

Fig. 5 An example of an aSVD SWS superimposed on the Grad-CAM activation map. *SWS* susceptibility-weighted MR sequences; *aSVD* arteriolosclerosis, *Grad-CAM* Gradient-weighted Class Activation Mapping

also impose a financial burden on patients. Herein, we used SWS to develop and test a DL model that can detect CMBs and classify CSVD into aSVD, CAA, and CADASIL to assist neurologists in optimizing the management of CSVD.

Some recent studies have used DL to detect CMBs on SWS [20–22], but images were cropped manually around the CMBs, or data were collected from only one center, thus, possibly limiting the clinical applicability and reliability. Moreover, previously proposed models have adopted non-end-to-end algorithms for detecting CMBs, which could potentially result in low efficiency and human errors. In this study, we proposed to use Mask R-CNN, an end-to-end algorithm, to detect CMBs slice-wise, which was developed and tested using SWS collected from three independent centers, regardless of the MRI system platforms or parameters. Using heterogeneous data to develop and test the model indicates that our model has stronger applicability and robustness, and using an end-to-end algorithm for CMBs detection indicates that our model has higher efficiency and fewer human errors. Additionally, another significance of detecting CMBs in this work is the use of the CMBs segmentation masks generated by the model to assist with the second task, i.e., CSVD classification. Performing two tasks to capture more detailed information can improve the model stability [23].

There are several studies on the application of DL in CSVD [24–26], but the focus is on the biomarkers rather than aetiopathogenic classification, limiting the benefit of optimizing CSVD management. Herein, we proposed an SWS-based MIL network that can classify sequence-wise CSVD into aSVD, CAA, and CADASIL from the perspective of etiopathogenesis. The MIL network must be able to extract SWS representations relative to the above types of CSVD to accomplish this task. SWS is presented in a three-dimensional (3D) form with a large volume, making it reasonable to use a 3D convolutional neural network (3D-CNN) to extract representations. However, using a 3D-CNN may be time-consuming and costly, leading to low calculation efficiency. Reslicing can be used to decrease the resolution of the SWS, thereby improving its efficiency. However, high resolution is required to recognize CSVD image biomarkers such as CMBs; therefore, we abandoned the method of lessening SWS volume by reslicing. Here, we used a two-dimensional CNN (2D-CNN) to extract SWS representations instead of 3D-CNN. The main limitation of 2D-CNN is that it is applicable only to 2D but not 3D images composed of multiple ordered slices, such as SWS [27]. To overcome this limitation, we used the attention module to aggregate representations of each SWS slice into the whole SWS representation, inspired by the interpretable weakly supervised DL method for data-efficient whole-slide image processing and learning proposed by *Ming* et al. [28]. Besides, to effectively train the model to classify CSVD based on SWS, we used CMBs segmentation masks generated by the Mask R-CNN to induce the MIL network to concentrate on the representations of the above types of CSVD. In tenfold cross-validation, the model demonstrated consistent performance on both internal and external test set, indicating that our model has excellent generalizability. Based on the Grad-CAM activation map, regions within the brain parenchyma were found to have a high degree of contribution to the model's ability to develop CSVD prediction, indicating that our model relies on brain representations to make differential decisions, further supporting its stability. Additionally, the model achieved better performance than the neurologist group in classifying CSVD into aSVD, CAA, and CADASIL on the external test set, indicating its potential to determine CSVD types using only SWS and providing an effective approach for clinical differential diagnosis of CSVD, especially for generalized medical institutions lacking adequate diagnostic equipment. Furthermore, the model also offers the added advantage of confirming CAA or CADASIL patients without the need for biopsy or genetic testing, thereby decreasing the financial burden on patients.

This study has some limitations and potential areas for future research. First, given the rarity of monogenic CSVD [10, 29], the present study included a relatively small number of patients with CADASIL and did not include other types of CSVD, such as inflammatory vasculitides. To overcome this limitation, future studies should aim to recruit more diverse and extensive cohorts. Second, because the SWS labels used to train and test the model for classifying CSVD are annotated by a committee based on neuroimaging biomarkers and clinical symptoms, there may be random human errors, especially in annotating aSVD and CAA; thus, the performance of our model in classifying CSVD is based on the neurologist-level "gold-standard", rather than a true gold-standard. In future research, molecular and protein-level studies need to be combined to eliminate human errors in the model's performance. Third, we did not include cases with mixed pathologies, such as small vessel disease combined with inflammatory vasculitides, because we

considered that the use of neuroimaging biomarkers and symptoms was insufficient to distinguish these mixed etiologies. In future research, more clinical information, such as biochemical examinations, should be included to train and test the model and achieve better results.

Conclusion

In this study, we developed and tested a DL model based on SWS that can detect CMBs and presented an IoU of 0.523 ± 0.319, a Dice score of 0.627 ± 0.296, and a recall of 0.706 ± 0.365, classification of CVSD into aSVD, CAA, and CADASIL had a weighted-average AUC of 0.908 (95% CI 0.895–0.921), an accuracy of 0.819 (95% CI 0.768–0.870), a weighted-average precision of 0.864 (95% CI 0.831–0.897), and a weighted-average F1-score of 0.829 (95% CI 0.782–0.876), providing a promising approach to assist neurologists in optimizing the management of CSVD.

Methods

Patients selection

In this study, 364 CSVD patients (217 patients with aSVD, 106 patients with CAA, and 41 patients with CADASIL) treated at three independent centers, Third Affiliated Hospital, Sun Yat-sen University (SYSUTH), the Maoming People's Hospital (MMPH), and the First Affiliated Hospital of SHANTOU University Medical College (STUMFH), were enrolled between January 2017 and May 2022 All patients with aSVD had vascular risk factors, such as hypertension, and their MRI neuroimaging met the STandards for ReportIng Vascular changes on nEurouimaging (STRIVE) for CSVD [3, 13, 30]. Patients with CAA met the diagnostic criteria for probable CAA according to the Boston criteria version 2.0 [15, 16]. CADASIL was confirmed by a genetic diagnosis of *NOTCH3* gene mutation or a granular osmiophilic material identified in a skin biopsy [31, 32]. Detailed inclusion criteria can be found in the supplementary materials.

Dataset curation

SWS of enrolled patients were collected to develop and test the model, regardless of SWS parameters, to improve robustness and applicability. The SWS acquisition details are shown in the supplementary materials (Additional file 1: Table S1). SWS of patients from SYSUTH (a hospital in Guangzhou city) and MMPH (a hospital in Maoming city) served as the development set, and SWS of patients from STUMFH (a hospital in Shantou city) served as the external test set for testing the geographic performance of the model [33]. Additionally, the tenfold cross-validation was used to evaluate the stability of the DL model.

Image of SWS preprocessing

The SWS images were pre-processed before inputting the DL model. Firstly, the orientation of all SWS was uniformly adjusted to left-posterior-superior, and the brain area was extracted from the irrelevant background. Subsequently, the pixel value in the image was ranked, and those topping $\geq 99.5\%$ were denoted as x_u. Finally, the pixel value of the entire image was normalized and scaled ($X = x/x_u$).

Annotation procedures

To train the model for detecting CMBs, an author (R.Z.W), who was blinded to patients' clinical data, performed segmentations of the CMBs using the software ITK-SNAP (version 3.8.0). Another author (B.Q) reviewed the results. To evaluate the reproducibility of the manual segmentations, 30 SWS were randomly selected from the development dataset, and an additional segmentation was performed 3 months after the initial segmentation.

To train and test the model for classifying CSVD into aSVD, CAA, and CADASIL, all SWS underwent additional annotation procedures to confirm the type of CSVD, i.e., the SWS label. A committee of three senior neurologists (B.Q, Z.Q.L, and H.L) annotated labels based on inclusion criteria, with more resource-intensive neurologist annotations being reserved. In cases where of disagreement regarding a label, the committee discussed and reached a consensus. If consensus could not be achieved, the SWS was excluded. In summary, each SWS label received one committee consensus annotation, which was regarded as the gold-standard for model evaluation. In the external test set, four neurologists (L.W, T.T.L, X.H.H, X.L) that were not part of the committee provided individual annotations for labels, and these labels were utilized to compare the model's performance. To evaluate the consistency and accuracy of the annotations, we employed the Fleiss Kappa coefficient to measure the inter-rater agreement among committee members.

Network architecture

Figure 6 shows detailed information regarding our architecture. Mask R-CNN was used to detect CMBs slice-wise and to obtain a semantic segmentation mask for CMBs [34]. Next, a multi-instance learning (MIL) network was applied to classify CSVD sequence-wise [35]. Semantic segmentation masks of CMBs generated by Mask-RCNN were then spliced with the original corresponding SWS to form a double-channel image as the model's input for classifying CSVD, inducing the model to concentrate on representations of CSVD on SWS. Resnet50 served as the backbone to extract representations of SWS on each slice layer by layer. Representations of each slice were exhibited

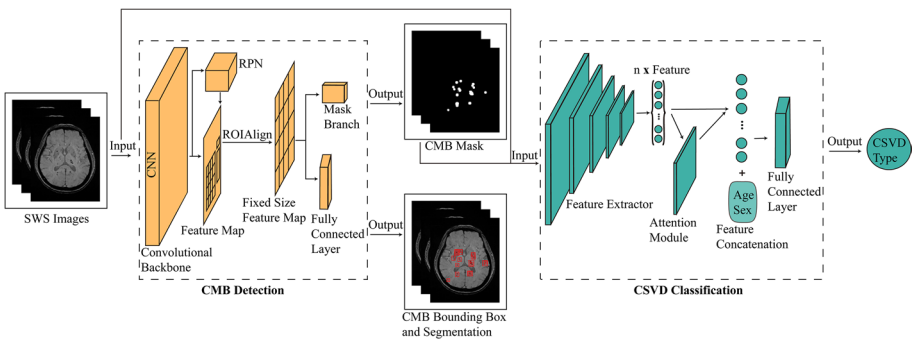

Fig. 6 Deep learning architecture overview. Mask R-CNN was used to detect CMBs slice-wise and to obtain a semantic segmentation mask of CMBs. Next, semantic segmentation masks of CMBs were spliced with the originally corresponding SWS to input the MIL network for classifying CSVD SWS-wise. *CMBs* cerebral microbleed, *SWS* susceptibility-weighted MR sequences, *MIL* multi-instance learning, *CSVD* cerebral small vessel disease

as a 1024-dimension vector. Representations of every slice belonging to the same SWS were aggregated into one feature vector using the attention module, among which those representations of slices relating to the final prediction CSVD type were given higher weights, and vice versa. This feature vector was concatenated with the corresponding patient's age and sex to generate a 1026-dimension vector, serving as the sequence's representation. Finally, a fully connected layer was used to predict CSVD type according to sequence representation. Gradient-weighted Class Activation Mapping (Grad-CAM) [36] was applied to visualize where the MIL network focuses its attention when developing a prediction.

Model training

Adjacent slices in the SWS were combined to form a three-channel two-dimensional image and input into the Mask-RCNN network to identify CMBs. The manual segmentation of CMBs from the middle channel of the two-dimensional image was used as the ground truth. In CMBs detection training, loss functions used for segmentation and boundary box placement included per-pixel sigmoid loss, binary loss, and regression loss, respectively. During the training for CSVD classification, a balanced random sampling of three categories of CSVD samples was realized using weighted sampling, and the batch size was defined as 1, which guaranteed that only one sample was input into the model. The loss function of the model for CSVD classification was the cross-entropy loss function. The parameters of the model, including the feature extraction module, were trained from end to end using the reverse value of the cross-entropy loss function. The Adam optimizer was used to optimize the parameters of the second model, and the initial learning rate was set at 0.001. An early stopping mechanism was used to control the performance of the training process. When the loss function on the internal test set of the development set did not decline for five consecutive epochs, the training of the model was terminated, and the best model was preserved to avoid a large overfitting of the model on the training set. To classify CSVD, we implemented the tenfold cross-validation to obtain ten models on the development set. To test the model's applicability for classifying CSVD, the performance of the ensemble model obtained by aggregating ten models was compared with that of four neurologists (L.W, T.T.L, X.H.H, X.L) on the external test set. The training parameters and source code can be found online (https://github.com/Huatsing-Lau/CSVD-CMBs-Detection-and-Classification.git).

An AMD EPYC 7742, 2.25-GHz CPU, and an A100 GPU (Nvidia) were used and run on a Linux system (Ubuntu, version 7.5.0) with a CUDA version 11.7 platform. Model implementation was performed using open-source software (Python, version 3.8.3; Python Software Foundation), Pytorch, version 1.13.0, and torchvision version 0.14.0.

Statistical analysis

Mean and standard deviation (SD) were used to describe continuous variables, while percentages were used to describe non-continuous variables. Differences between groups were tested using one-way ANOVA and the Chi-square test, respectively.

The IoU, Dice score, and recall were calculated to evaluate the model's performance in detecting CMBs. The IoU formulas was expressed by Eq. (1) was used to assess the performance of bounding boxes. Bounding box placement was considered correct for an

IoU above 0.5. The Dice score formulas expressed by Eq. (2) was used to assess the performance of the segmentation.

$$IoU = \frac{TP}{FP + TP + FN}, \tag{1}$$

$$Dicescore = \frac{2TP}{FP + 2TP + FN}, \tag{2}$$

where TP, FP, and FN represent true positive pixels, false positive pixels, and false negative pixels, respectively. The recall formulas expressed by Eq. (3) was used to evaluate the proportion of the true CMBs predicted by the model.

$$Recall = \frac{TP}{TP + FN}, \tag{3}$$

where TP and FN represent the number of true positive CMBs detected by the Mask R-CNN (only those predicted CMBs with a Dice score exceeding 0.5 are considered true positive CMBs).

Confusion matrices, ROC curves, precision, and accuracy were used to evaluate the performance of the model in classifying CSVD. Confusion matrices were used to illustrate the label of CSVD classification, where the model prediction or the individual neurologist predictions were discordant with the committee consensus. ROC analysis was used to calculate AUC to assess the discrimination of the model for each label class. With false positive rate (FPR) as the horizontal coordinate and true positive rate (TPR) as the vertical coordinate, ROC can be drawn and AUC can be calculated. The formulas of TPR and FPR were expressed in Eq. (4) and Eq. (5), respectively. Accuracy precision, and F1 score were calculated to provide additional information for ROC analysis [37], providing complementary performance measures to the AUC, especially in the context of multi-class prediction, and reducing sensitivity to class imbalance issues. The weighted-average of the indices was calculated to eliminate the imbalance between the number of categories. The formulas of accuracy, precision, and F1 score were expressed in Eq. (6), Eq. (7), and Eq. (8), respectively.

$$TPR = \frac{TP}{TP + FN}, \tag{4}$$

$$FPR = \frac{FP}{FP + TN}, \tag{5}$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}, \tag{6}$$

$$Precision = \frac{TP}{TP + FP}, \tag{7}$$

$$F1score = \frac{2TP}{2TP + FN + FP}, \tag{8}$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative numbers, respectively.

Statistical significance was set at $P < 0.05$. The 95% confidence interval (CI) was calculated for each index. IBM SPSS Statistics (version 26), scikit-learn (version 0.24.2), and the statsmodels (version 0.13.5) were used to analyze the metrics of the models.

Abbreviations

CMBs	Cerebral microbleeds
CSVD	Cerebral small vessel disease
DL	Deep learning
SWS	Susceptibility-weighted MR sequence
aSVD	Arteriolosclerosis
CAA	Cerebral amyloid angiopathy
CADASIL	Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy
MIL	Multi-instance learning
IoU	Intersection over union
ROC	Receiver operating characteristic curve
AUC	Area under receiver operating characteristic curve
SWAN	T2-star-weighted angiography
SWI	Susceptibility-weighted imaging
STRIVE	STandards for reporting vascular changes on nEurouimaging
SYSUTH	Third Affiliated Hospital, Sun Yat-sen University (SYSUTH)
MMPH	Maoming People's Hospital
STUMFH	First Affiliated Hospital of SHANTOU University Medical College
SD	Standard deviation
Grad-CAM	Gradient-weighted Class Activation Mapping
CI	Confidence interval
3D-CNN	Three-dimensional convolutional neural network
2D-CNN	Two-dimensional convolutional neural network
FPR	False positive rate
TPR	True positive rate

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12938-023-01164-1.

Additional file 1: Table S1. Detailed SWS acquisition. **Table S2.** Results for each run of ten-fold cross-validation of the model for CSVD classification. **Table S3.** Results of four neurologists' performance in classifying CSVD on the external test set. **Figure S1.** Confusion matrices for neurologists' performance in classifying CSVD on the external test set. **A** neurologist W.L, **B** neurologist T.T.L, **C** neurologist X.X.H, **D** neurologist X.L.

Acknowledgements
The authors would like to thank *Tingting Lu* for helping with the CSVD classification on external test set.

Author contributions
Guarantors of the integrity of the entire study, RZW, HQL, ZQL, and LQH; BQ; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscripts drafting or manuscript revision for important intellectual content, all authors; approval of the final version of a submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, RZW and HQL; clinical studies, RZW, HL, LFC, ZQL, and QB; statistical analysis, RZW and HQL; XDL; manuscript editing, RZW, HQL, ZQL, LQH, and QB.

Funding
This work was supported by the National Natural Science Foundation of China (Grant No. 81971110 and 82171307).

Availability of data and materials
The datasets used during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
The Medical Ethics Committee of the Third Affiliated Hospital of Sun Yat-sen University approved this retrospective multicenter study. This study was conducted following the Declaration of Helsinki. Requirements for informed consent were waived because of the study's retrospective nature.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Received: 4 May 2023 Accepted: 10 October 2023
Published online: 17 October 2023

References

1. Inzitari D, Pracucci G, Poggesi A, Carlucci G, Barkhof F, Chabriat H, et al. Changes in white matter as determinant of global functional decline in older independent outpatients: 3 year follow-up of LADIS (leukoaraiosis and disability) study cohort. BMJ. 2009;339:b2477.
2. Markus HS, van Der Flier WM, Smith EE, Bath P, Biessels GJ, Briceno E, et al. Framework for clinical trials in cerebral small vessel disease (FINESSE): a review. JAMA Neurol. 2022;79(11):1187–98.
3. Pantoni L. Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges. Lancet Neurol. 2010;9(7):689–701.
4. Domenga V, Fardoux P, Lacombe P, Monet M, Maciazek J, Krebs LT, et al. Notch3 is required for arterial identity and maturation of vascular smooth muscle cells. Genes Dev. 2004;18(22):2730–5.
5. Petty GW, Brown RD Jr, Whisnant JP, Sicks JD, O'Fallon WM, Wiebers DO. Ischemic stroke subtypes: a population-based study of functional outcome, survival, and recurrence. Stroke. 2000;31(5):1062–8.
6. Cannistraro RJ, Badi M, Eidelman BH, Dickson DW, Middlebrooks EH, Meschia JF. CNS small vessel disease: A clinical review. Neurology. 2019;92(24):1146–56.
7. Chabriat H, Joutel A, Dichgans M, Tournier-Lasserve E, Bousser MG. Cadasil. Lancet Neurol. 2009;8(7):643–53.
8. Tapia J. Amyloid angiopathy brain hemorrhage. Rev Med Chil. 2021;149(1):76–87.
9. Kozberg MG, Perosa V, Gurol ME, van Veluw SJ. A practical approach to the management of cerebral amyloid angiopathy. Int J Stroke. 2021;16(4):356–69.
10. Mancuso M, Arnold M, Bersano A, Burlina A, Chabriat H, Debette S, et al. Monogenic cerebral small-vessel diseases: diagnosis and therapy consensus recommendations of the european academy of neurology. Eur J Neurol. 2020;27(6):909–27.
11. Smith EE, Markus HS. New treatment approaches to modify the course of cerebral small vessel diseases. Stroke. 2020;51(1):38–46.
12. Bouasquevisque DS, Benavente OR, Shoamanesh A. Antiplatelet therapy in cerebral small vessel disease. Curr Neurol Neurosci Rep. 2019;19(9):61.
13. Wardlaw JM, Smith EE, Biessels GJ, Cordonnier C, Fazekas F, Frayne R, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. Lancet Neurol. 2013;12(8):822–38.
14. Charidimou A, Kakar P, Fox Z, Werring DJ. Cerebral microbleeds and recurrent stroke risk: systematic review and meta-analysis of prospective ischemic stroke and transient ischemic attack cohorts. Stroke. 2013;44(4):995–1001.
15. Charidimou A, Boulouis G, Frosch MP, Baron JC, Pasi M, Albucher JF, et al. The Boston criteria version 2.0 for cerebral amyloid angiopathy: a multicentre, retrospective MRI-neuropathology diagnostic accuracy study. Lancet Neurol. 2022;21(8):714–25.
16. Charidimou A, Boulouis G. Clinical diagnosis of probable cerebral amyloid angiopathy: diagnostic accuracy meta-analysis of the boston criteria. Stroke. 2022. https://doi.org/10.1161/STROKEAHA.122039501.
17. Rudilosso S, Chui E, Stringer MS, Thrippleton M, Chappell F, Blair G, et al. Prevalence and significance of the vessel-cluster sign on susceptibility-weighted imaging in patients with severe small vessel disease. Neurology. 2022;99(5):e440–52.
18. von Schacky CE, Wilhelm NJ, Schafer VS, Leonhardt Y, Gassert FG, Foreman SC, et al. Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. Radiology. 2021;301(2):398–406.
19. Verdu-Diaz J, Alonso-Perez J, Nunez-Peralta C, Tasca G, Vissing J, Straub V, et al. Accuracy of a machine learning muscle MRI-based tool for the diagnosis of muscular dystrophies. Neurology. 2020;94(10):e1094–102.
20. Suwalska A, Wang Y, Yuan Z, Jiang Y, Zhu D, Chen J, et al. CMB-HUNT: automatic detection of cerebral microbleeds using a deep neural network. Comput Biol Med. 2022;151(Pt A):106233.
21. Liu S, Utriainen D, Chai C, Chen Y, Wang L, Sethi SK, et al. Cerebral microbleed detection using susceptibility weighted imaging and deep learning. Neuroimage. 2019;198:271–82.
22. 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.
23. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. Radiology. 2019;290(3):590–606.

24. Hsieh YZ, Luo YC, Pan C, Su MC, Chen CJ, Hsieh KL. Cerebral small vessel disease biomarkers detection on mri-sensor-based image and deep learning. Sensors. 2019;19(11):2573.
25. Duan Y, Shan W, Liu L, Wang Q, Wu Z, Liu P, et al. Primary categorizing and masking cerebral small vessel disease based on "deep learning system." Front Neuroinform. 2020;14:17.
26. Karel MFA, Roosen M, Tullemans BME, Zhang CE, Staals J, Cosemans J, et al. Characterization of cerebral small vessel disease by neutrophil and platelet activation markers using artificial intelligence. J Neuroimmunol. 2022;367:577863.
27. Li Z, Liu F, Yang W, Peng S, Zhou J. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE Trans Neural Netw Learn Syst. 2022;33(12):6999–7019.
28. Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. Nat Biomed Eng. 2021;5(6):555–70.
29. Bordes C, Sargurupremraj M, Mishra A, Debette S. Genetics of common cerebral small vessel disease. Nat Rev Neurol. 2022;18(2):84–101.
30. Cai W, Chen X, Men X, Ruan H, Hu M, Liu S, et al. Gut microbiota from patients with arteriosclerotic CSVD induces higher IL-17A production in neutrophils via activating RORgammat. Sci Adv. 2021;7(4):eabe4827.
31. Chabriat H, Joutel A, Tournier-Lasserve E, Bousser MG. CADASIL: yesterday, today, tomorrow. Eur J Neurol. 2020;27(8):1588–95.
32. Tikka S, Mykkanen K, Ruchoux MM, Bergholm R, Junna M, Poyhonen M, et al. Congruence between NOTCH3 mutations and GOM in 131 CADASIL patients. Brain. 2009;132(Pt 4):933–9.
33. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. Radiology. 2018;286(3):800–9.
34. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. IEEE Int Conf Comput Vis. 2017. https://doi.org/10.48550/arXiv.1703.06870.
35. Carbonneau M-A, Cheplygina V, Granger E, Gagnon G. Multiple instance learning: a survey of problem characteristics and applications. Pattern Recogn. 2018;77:329–53.
36. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. IEEE Int Conf Comput Vis. 2017. https://doi.org/10.48550/arXiv.1610.02391.
37. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS ONE. 2015;10(3): e0118432.

Publisher's Note