

Development and validation of a deep-learning model for detecting brain metastases on 3D post-contrast MRI: a multi-center multi-reader evaluation study

Shaohan Yin[†], Xiao Luo[†], Yadi Yang[†], Ying Shao, Lidi Ma[°], Cuiping Lin, Qiuxia Yang, Deling Wang, Yingwei Luo, Zhijun Mai, Weixiong Fan, Dechun Zheng, Jianpeng Li, Fengyan Cheng, Yuhui Zhang, Xinwei Zhong, Fangmin Shen, Guohua Shao, Jiahao Wu, Ying Sun[°], Huiyan Luo, Chaofeng Li, Yaozong Gao[°], Dinggang Shen, Rong Zhang[°], and Chuanmiao Xie[°]

State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China (S.Y., X.L., Y.Y., L.M., Cu.L., Q.Y., D.W., Y.L., Z.M., Y.S., H.L., Ch.L., R.Z., C.X.); Department of Radiology, Sun Yat-Sen University Cancer Center, Guangzhou, China (S.Y., X.L., Y.Y., L.M., C.L., Q.Y., D.W., Y.L., Z.M., R.Z., C.X.); R&D Department, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China (Y.S., Y.G., D.S.); Department of Magnetic Resonance, Guangdong Provincial Key Laboratory of Precision Medicine and Clinical Translational Research of Hakka Population, Meizhou People's Hospital, Meizhou, China (W.F., F.C., Y.Z., X.Z.); Department of Radiology, Fujian Cancer Hospital, Fujian Medical University Cancer Hospital, Fuzhou, Fujian Province, China (D.Z., F.S.); Department Of Radiology, Affiliated Dongguan Hospital, Southern Medical University, Dongguan, China (J.L., G.S., J.W.); Department of Radiation Oncology, Sun Yat-Sen University Cancer Center, Guangzhou, China (Y.S.); Department of Medical Oncology, Sun Yat-Sen University Cancer Center, Guangzhou, China (H.L.); Department of Artificial Intelligence Laboratory, Sun Yat-Sen University Cancer Center, Guangzhou, China (C.L.); School of Biomedical Engineering, ShanghaiTech University, Shanghai, China (D.S.)

[†]These authors contributed equally to this work.

Corresponding Authors: Chuanmiao Xie, PhD, The Department of Radiology, 651 Dongfeng Road East, Yuexiu District, Guangzhou 510060, P.R. China (xchuanm@sysucc.org.cn); Rong Zhang, PhD, The Department of Radiology, 651 Dongfeng Road East, Yuexiu District, Guangzhou 510060, P.R. China (zhangr@sysucc.org.cn).

Abstract

Background. Accurate detection is essential for brain metastasis (BM) management, but manual identification is laborious. This study developed, validated, and evaluated a BM detection (BMD) system.

Methods. Five hundred seventy-three consecutive patients (10448 lesions) with newly diagnosed BMs and 377 patients without BMs were retrospectively enrolled to develop a multi-scale cascaded convolutional network using 3D-enhanced T1-weighted MR images. BMD was validated using a prospective validation set comprising an internal set (46 patients with 349 lesions; 44 patients without BMs) and three external sets (102 patients with 717 lesions; 108 patients without BMs). The lesion-based detection sensitivity and the number of false positives (FPs) per patient were analyzed. The detection sensitivity and reading time of three trainees and three experienced radiologists from three hospitals were evaluated using the validation set.

Results. The detection sensitivity and FPs were 95.8% and 0.39 in the test set, 96.0% and 0.27 in the internal validation set, and ranged from 88.9% to 95.5% and 0.29 to 0.66 in the external sets. The BMD system achieved higher detection sensitivity (93.2% [95% CI, 91.6–94.7%]) than all radiologists without BMD (ranging from 68.5% [95% CI, 65.7–71.3%] to 80.4% [95% CI, 78.0–82.8%], all $P < .001$). Radiologist detection sensitivity improved with BMD, reaching 92.7% to 95.0%. The mean reading time was reduced by 47% for trainees and 32% for experienced radiologists assisted by BMD relative to that without BMD.

Conclusions. BMD enables accurate BM detection. Reading with BMD improves radiologists' detection sensitivity and reduces their reading times.

Key Points

- The multi-scale cascaded network enabled accurate brain metastasis detection on MRI.
- The detection system improved radiologists' abilities to detect brain metastases.

Importance of the Study

Early and accurate detection of brain metastases (BMs) is vital for effective treatment. Currently, radiologists manually identify BMs, which is laborious, time-consuming, and particularly challenging when detecting subtle lesions. We developed a multi-scale cascaded convolution network for BM detection and prospectively validated its generality using multi-center datasets. Our method had a superior detection sensitivity of 93.2%, with 0.38 false positives per scan in the multi-center validation. The BM detection

system achieved the highest reported sensitivity, 89%, for metastases sized ≤ 3 mm. Additionally, the multi-reader assessment by six radiologists with various levels of experience demonstrated that using our novel system increased mean sensitivity by 21% and decreased mean reading time by 40%. Therefore, our robust BM detection system will facilitate early and accurate BM diagnosis and improved management by efficiently assisting radiologists in BM detection and reducing radiologists' workloads.

Brain metastases (BMs) are the most common intracranial tumors, affecting an estimated 20–45%¹ of patients with advanced solid tumors. BM incidence is increasing as a result of systemic and brain-directed therapies² and the development of superior neuroimaging strategies.³ BMs can dramatically alter treatment paradigms given the associated neurocognitive deficits, reduced quality of life, and poor prognosis, representing an increasingly prevalent challenge in multiple disciplines. Stereotactic radiosurgery (SRS) is the standard of care for eligible patients with limited BMs^{4,5} because neurocognitive function is better maintained with SRS than with whole-brain radiation therapy.⁶ Additionally, growing evidence supports SRS as a monotherapy for patients with multiple BMs.^{7–10} The development of even small BMs alters the tumor stage, and the number of BMs impacts treatment decisions. Thus, early and accurate detection of the number, size, and location of BMs is essential for decision-making.¹¹

MRI is the preferred modality for BM detection, diagnosis, response evaluation, and surveillance. Three-dimensional contrast-enhanced T1-weighted imaging (CET1WI) has significantly increased BM detection sensitivity.¹² However, manual BM identification by radiologists can be laborious and time-consuming. For patients with multiple metastases, it can be challenging to identify all lesions, especially those sized ≤ 5 –7 mm, for which the detection sensitivity is as low as 50–60%.^{13–16} Thus, effective BM detection is an urgent unmet need.

Several methods have been developed for computer-aided detection (CAD) of BMs on MRI using various algorithms and sequences, and some encouraging preliminary results have been published.^{13,15–25} A recent meta-analysis of 12 studies reported comparable BM detectability between classical machine learning and

deep-learning (DL) approaches, which achieved pooled sensitivities of 88.7% and 90.1%, respectively.²⁴ Despite ongoing CAD tool development, barriers in the widespread clinical translation of these techniques remain, and, as far as we know, no clinically useful system is available. All but one study²⁰ included in the meta-analysis used a single-center design,^{13,15–19,21–23,25} meaning that algorithm robustness was not evaluated. One study¹³ was a prospective design, and all others were retrospective,^{15–23,25} which can lead to selection bias. In addition, most studies had relatively small sample sizes (<200 patients; range 19–1632),^{13–15,17,18,20,21,26,27} and sample representativeness may have been limited. Furthermore, previous studies reported relatively low detection sensitivities (59%–81%) for small lesions (≤ 5 –6 mm)^{16,26} or high number of false positives (FPs; up to 302 per person),^{18,21,24,25} indicating that an optimized model is needed. Finally, most previous studies focused on the standalone performance of systems by testing retrospective data, and only a limited number of studies reported that radiologists' BM detection sensitivity was improved with CAD.^{13,14,22,25} To the best of our knowledge, the impact of incorporating CAD into the clinical workflow on radiologists has not been determined.

In this multi-center study, we developed a DL-based system for automatic BM detection (BMD) on 3D CET1WI using a large-scale cohort. We prospectively validated the BMD system using datasets from four institutions and evaluated radiologists' BM detection efficiency with and without assistance from our novel system. We hypothesized that our approach would enable high BM detection sensitivity and generate few FPs, providing a valuable adjunct to radiologists in detecting BMs.

Materials and Methods

Study Design and Participants

This multi-center multi-reader evaluation detection study was conducted in four hospitals of different tiers. Between August 11, 2019 and December 31, 2020, 676 consecutive patients who were newly diagnosed with BMs and underwent 3D-enhanced MRI at Sun Yat-sen University Cancer Center (SYSUCC) were retrospectively included as the training and test set. Between September and December 2020, we also consecutively enrolled patients who had extracranial tumors but no evidence of BMs and had received head MRI, designated as the negative control set for training and testing. To evaluate the robustness of our system, consecutive cohorts, defined as the validation set, were prospectively recruited from SYSUCC and Meizhou People's Hospital (MZPH), Dongguan People's Hospital (DGPH), and Fujian Cancer Hospital (FJCH) between February and May 2021.

For the group with BMs, the inclusion criteria were patients: (1) with extracranial primary tumor(s) confirmed by pathology; (2) with newly developed BMs; (3) who underwent 3D-enhanced brain MRI and at least one follow-up MRI, but only the initial scan with at least one lesion was assessed in this study. For control patients, the inclusion criteria were patients: (1) with extracranial primary tumor(s) confirmed by pathology; (2) who underwent 3D-enhanced brain MRI and showed no evidence of BMs. We excluded patients: (1) with primary intracranial tumor(s); (2) with meningeal metastasis; (3) who had undergone brain surgery; (4) with excessive artifacts. There were no limitations on the number or size of metastases in patients included in the training and test sets, but patients in the validation set with more than 50 BMs were excluded. The detailed distributions of all datasets are provided in [Supplementary Material 1](#).

This multi-center study was approved by the institutional review boards (No. B2021-198-01) and undertaken according to the Declaration of Helsinki. For patients who were retrospectively recruited from SYSUCC for the training set and testing sets, informed consent was exempted by the institutional review board. Informed consent was obtained from patients included in the prospective validation set.

MRI Protocol and Image Quality Control

Details of MRI acquisition are provided in [Supplementary Material 2](#) and [Supplementary Table 1](#). All MRI scans and clinical data were carefully reviewed by two board-certified neuroradiologists (with 12 and 8 years of experience, respectively).

Image Segmentation

No overlapping images were included among the training, test, and validation sets. Ground truths were manually established in the training set by marking a rectangular box around each BM on the axial 3D CET1WI images. This was done by four radiologists (with 3, 3, 8, and 10 years of experience, respectively), using Uii Brain Metastasis

Annotation Software (1.0, United Imaging Intelligence, Shanghai, China). The markers were confirmed by two neuroradiologists (with 12 and 8 years of experience, respectively). For all prospective sets, the BMs were independently confirmed by three experts (with 30, 28, and 11 years of experience, respectively) from three hospitals. Divergences were resolved by consensus in either the training or validation set. The BM size was defined as the largest cross-sectional dimension on the axial image.

BMD Architecture and Implementation

The BMD system consisted of three components, i.e., the multi-scale detection network (feature pyramid network [FPN]),²⁸ the cascade network, and the classification network (a 3D variant of LeNet network).²⁹ Anchor boxes with different shapes were employed in the FPN to improve detection sensitivity for BMs of various sizes and shapes. To reduce FPs while maintaining high sensitivity, we designed a cascade network with multiple serial neural network blocks, where the first detection network aimed to detect lesions with high sensitivity, and subsequent networks were trained using the same positive samples but different negative samples. The negative samples were FPs from the previous network. In this way, the cascaded networks focused on effectively removing FPs in the detection results and improving the overall performance based on the bagging training concept.³⁰ BMD architecture details are provided in [Supplementary Material 3](#) and [Supplementary Figures 1 and 2](#).

Model implementation was based on the Pytorch framework,³¹ an open-source Python DL library. We used focal loss ($\alpha = 2$; $\beta = .999$) as the loss function. The model was trained using the Adam optimization algorithm to dynamically adjust the learning rate. The initial learning rate was 1×10^{-4} . The training used an Intel (R) Xeon (R) CPU E5-2698 v4 @ 2.20 GHz central processing unit (CPU) and an Nvidia Tesla V100-SXM2, 32G \times 8 graphics processing unit (GPU) with CUDA version 10.1. Additionally, data augmentation (e.g., random shifting, rotation) was performed to enrich the training dataset.

Multi-Center Validation of BMD and Multi-Reader Evaluation

We first evaluated BMD performance using an internal test set from SYSUCC. We then assessed the robustness using a validation set from the four participating hospitals.

To assess the clinical value of BMD to radiologists, we designed a multi-center observer performance study. We recruited nine radiologists from three hospitals (SYSUCC, MZPH, DGPH) with varying degrees of experience (trainee [< 3 years], experienced [5–10 years], and expert [> 10 years]). These radiologists, who were not involved in patient selection or image annotation and were blinded to patient information, were asked to independently read the validation images slice by slice using the two reading modes, with an interval of 3 weeks between readings. In the first session, all images were read without BMD. In the second session, the same data were presented again in random order, and the readers used BMD at the start of the assessment. The readers were instructed to mark each lesion. There was no

reading time limit. The reading time per case was recorded in each session for efficiency analysis.

Statistical Analysis

The lesion-based detection sensitivity, free-response receiver operating characteristic (FROC), the number of FPs per scan were evaluated, and the 95% confidence intervals (CIs) were calculated to assess BMD performance. The sensitivities were further categorized based on the BM size. For the two reading sessions of the multi-reader analysis, we compared lesion-based detection sensitivities using McNemar's test, patient-based area under the curve (AUC) using Z Test and reading times using Wilcoxon's signed-rank test. Correlations between lesion size and number and BMD sensitivity were analyzed using binary logistic regression analysis. Because all MRI scans from patients were integrated as one data set, the within-patient correlation for patients with multiple BMs was not considered. Statistical analyses were conducted using SPSS (25.0, IBM Corp., Armonk, NY, USA) and were two-sided with a significance level of 0.05.

Results

Patient Characteristics

A flowchart depicting the patient selection processes is shown in [Figure 1](#). Out of 1,124 patients in the

retrospective cohort from SYSUCC, 174 were excluded. The final data set comprised 950 patients (mean age, 55 ± 12 years; 518 males, 432 females) who were assigned to two sets with a temporal split on October 31, 2020: (a) a training set of 411 patients with 9360 BMs and 269 patients without BMs and (b) a test set of 162 patients with 818 BMs and 108 patients without BMs. For the validation set, data from 191 patients with BMs and 152 patients without BMs were prospectively collected from the four participating hospitals. Forty-three patients were excluded, and the final validation set comprised 148 patients with BMs and 152 patients without BMs as follows: 46 patients with 348 BMs and 44 patients without BMs from SYSUCC; 38 patients with 177 BMs and 32 patients without BMs from MZPH; 34 patients with 315 BMs and 36 patients without BMs from DGPB; 30 patients with 225 BMs and 40 patients without BMs from FJCH. Lung cancer was the leading primary tumor type, with 90%, 82%, and 84% prevalence in the training, test, and validation sets, respectively, followed by breast cancer. The patient characteristics are presented in [Table 1](#). Details of excluded individuals among the participating hospitals are provided in [Supplementary Material 1](#).

In the training, test, and validation sets, the mean number of BMs per patient, mean metastasis size, and the number of metastases smaller than 10 mm were 17, 5, and $7, 5.5 \pm 4.1$ mm, 7.5 ± 6.3 mm, and 7.1 ± 6.0 mm, and 8895 (92.4%), 674 (82.4%), and 888 (83.3%), respectively. The metastasis size and number distributions are shown in [Figure 2](#).

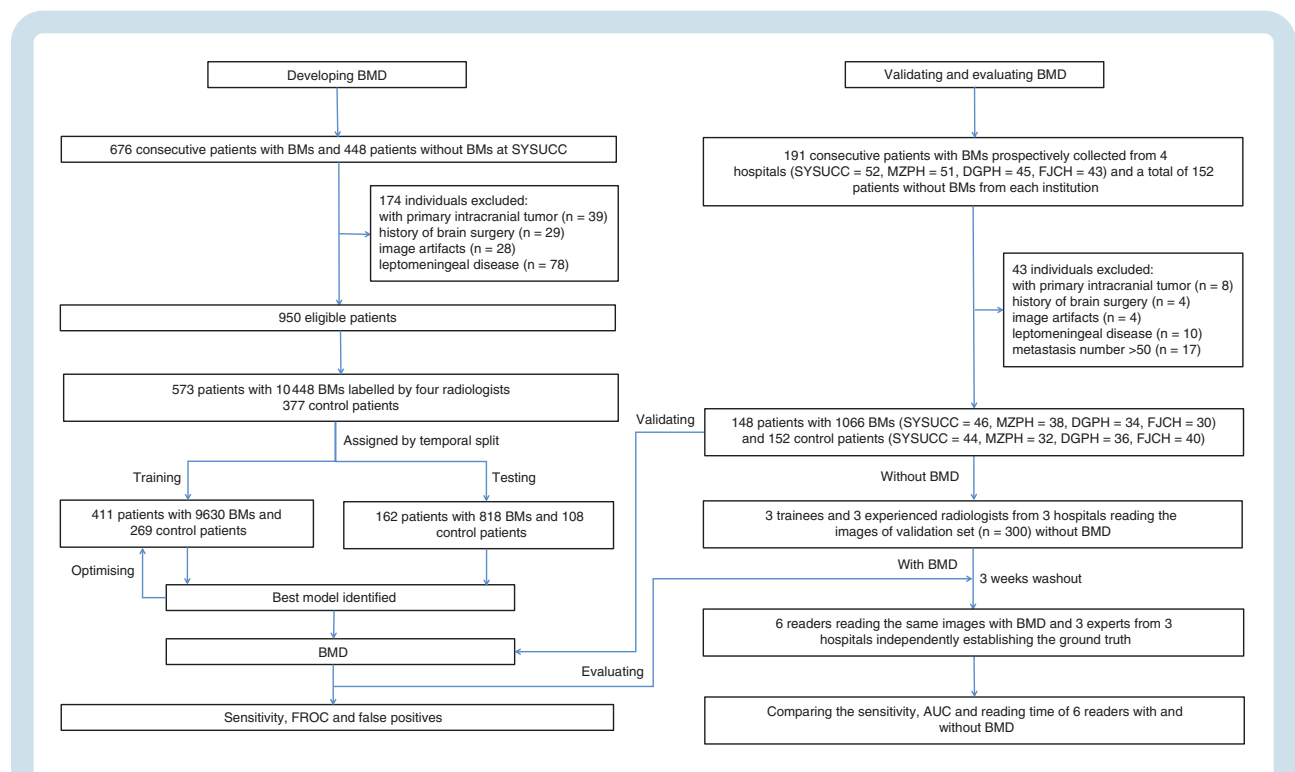


Fig. 1 Flow diagram for development, validation and multi-reader evaluation of BMD. BMD, brain metastases detection; BMs, brain metastases; SYSUCC, Sun Yat-sen University Cancer Center; MZPH, Meizhou People's Hospital; DGPB, Dongguan People's Hospital; FJCH, Fujian Cancer Hospital; FROC, free-response receiver operating characteristic; AUC, area under the curve.

Table 1 Patient Characteristics and Brain Metastases Information

| Variables | Training (n = 680) | | Test (n = 270) | | Validation (n = 300) | | FJCH (n = 70) | | DGPH (n = 70) | | MZPH (n = 70) | | SYSUCC (n = 90) | | Total (n = 1250) | |
|----------------------|--------------------|------------|----------------|------------|----------------------|------------|---------------|------------|---------------|------------|---------------|------------|-----------------|------------|------------------|------------|
| | With BM | Without BM | With BM | Without BM | With BM | Without BM | With BM | Without BM | With BM | Without BM | With BM | Without BM | With BM | Without BM | With BM | Without BM |
| Patients | 411 (60.4) | 269 (39.6) | 162 (60.0) | 108 (40.0) | 46 (51.1) | 44 (48.9) | 38 (54.3) | 32 (45.7) | 34 (48.6) | 36 (51.4) | 30 (42.9) | 40 (57.1) | 148 (49.3) | 152 (50.7) | 721 (57.7) | 529 (42.3) |
| Male | 233 (56.7) | 141 (52.4) | 87 (71.3) | 57 (52.8) | 26 (56.5) | 22 (50.0) | 21 (55.3) | 20 (62.5) | 15 (44.1) | 13 (36.1) | 19 (63.3) | 21 (52.5) | 85 (57.4) | 76 (50.0) | 405 (56.2) | 350 (66.2) |
| Age (years) | 55 ± 12 | 52 ± 14 | 57 ± 10 | 51 ± 13 | 60 ± 12 | 48 ± 11 | 62 ± 10 | 58 ± 13 | 58 ± 13 | 59 ± 10 | 59 ± 9 | 59 ± 10 | 60 ± 11 | 56 ± 13 | 58 ± 11 | 53 ± 13 |
| BM | | | | | | | | | | | | | | | | |
| Number | 9630 | NA | 818 | NA | 349 | NA | 177 | NA | 315 | NA | 225 | NA | 1066 | NA | 11 514 | NA |
| Size (mm) | | | | | | | | | | | | | | | | |
| Mean±SD | 5.5 ± 4.1 | NA | 7.5 ± 6.3 | NA | 7.5 ± 6.2 | NA | 7.9 ± 7.0 | NA | 6.4 ± 5.2 | NA | 6.7 ± 5.8 | NA | 7.1 ± 6.0 | NA | 5.8 ± 4.6 | NA |
| <10 | 8,895 (92.4) | NA | 674 (82.4) | NA | 282 (80.8) | NA | 136 (76.8) | NA | 278 (88.3) | NA | 192 (85.3) | NA | 888 (83.3) | NA | 10,457 (90.8) | NA |
| Magnetic field | | | | | | | | | | | | | | | | |
| 3.0Tesla | 304 (74.0) | 210 (78.1) | 124 (76.5) | 85 (78.7) | 43 (93.5) | 39 (88.7) | 29 (76.3) | 32 (100) | 34 (100) | 36 (100) | 30 (100) | 40 (100) | 136 (91.2) | 147 (96.7) | 564 (78.2) | 442 (83.6) |
| Primary tumor types | | | | | | | | | | | | | | | | |
| Lung | 374 (90.1) | 204 (75.8) | 133 (82.1) | 84 (77.8) | 37 (80.4) | 29 (65.9) | 34 (89.5) | 13 (40.6) | 30 (88.2) | 22 (61.1) | 23 (76.7) | 30 (75.0) | 124 (83.8) | 94 (61.8) | 631 (87.6) | 382 (72.2) |
| Breast | 15 (3.6) | 17 (6.3) | 12 (7.4) | 5 (4.6) | 5 (10.9) | 6 (13.6) | 3 (7.9) | 4 (12.5) | 2 (5.9) | 9 (25.0) | 4 (13.3) | 1 (2.5) | 14 (9.5) | 20 (13.2) | 41 (5.7) | 42 (7.9) |
| Melanoma | 6 (1.5) | 6 (2.2) | 2 (1.2) | 3 (2.8) | 2 (4.3) | 1 (2.3) | 0 | 0 | 0 | 0 | 0 | 0 | 2 (1.4) | 3 (2.0) | 10 (1.4) | 12 (2.3) |
| Renal | 1 (0.2) | 4 (1.5) | 4 (2.5) | 0 | 1 (2.2) | 1 (2.3) | 0 | 0 | 1 (2.9) | 0 | 0 | 0 | 2 (1.4) | 8 (0.7) | 7 (1.0) | 5 (0.9) |
| Colorectal carcinoma | 6 (1.5) | 5 (1.9) | 0 | 1 (0.9) | 0 | 0 | 0 | 1 (3.1) | 0 | 0 | 1 (3.3) | 1 (2.5) | 1 (0.7) | 2 (1.3) | 7 (1.0) | 8 (1.5) |
| Other types | 9 (2.2) | 33 (12.3) | 11 (6.8) | 15 (13.9) | 1 (2.2) | 7 (15.9) | 1 (2.6) | 14 (43.8) | 1 (2.9) | 5 (13.9) | 2 (6.7) | 6 (15.0) | 5 (3.4) | 32 (21.1) | 25 (3.5) | 80 (15.1) |

Data are presented as either mean ± standard deviation or numbers of patients (%). BM, brain metastasis; SD, standard deviation; NA, not applicable; SYSUCC, Sun Yat-sen University Cancer Center; MZPH, Meizhou People's Hospital; DGPH, Dongguan People's Hospital; FJCH, Fujian Cancer Hospital.

BMD Performance

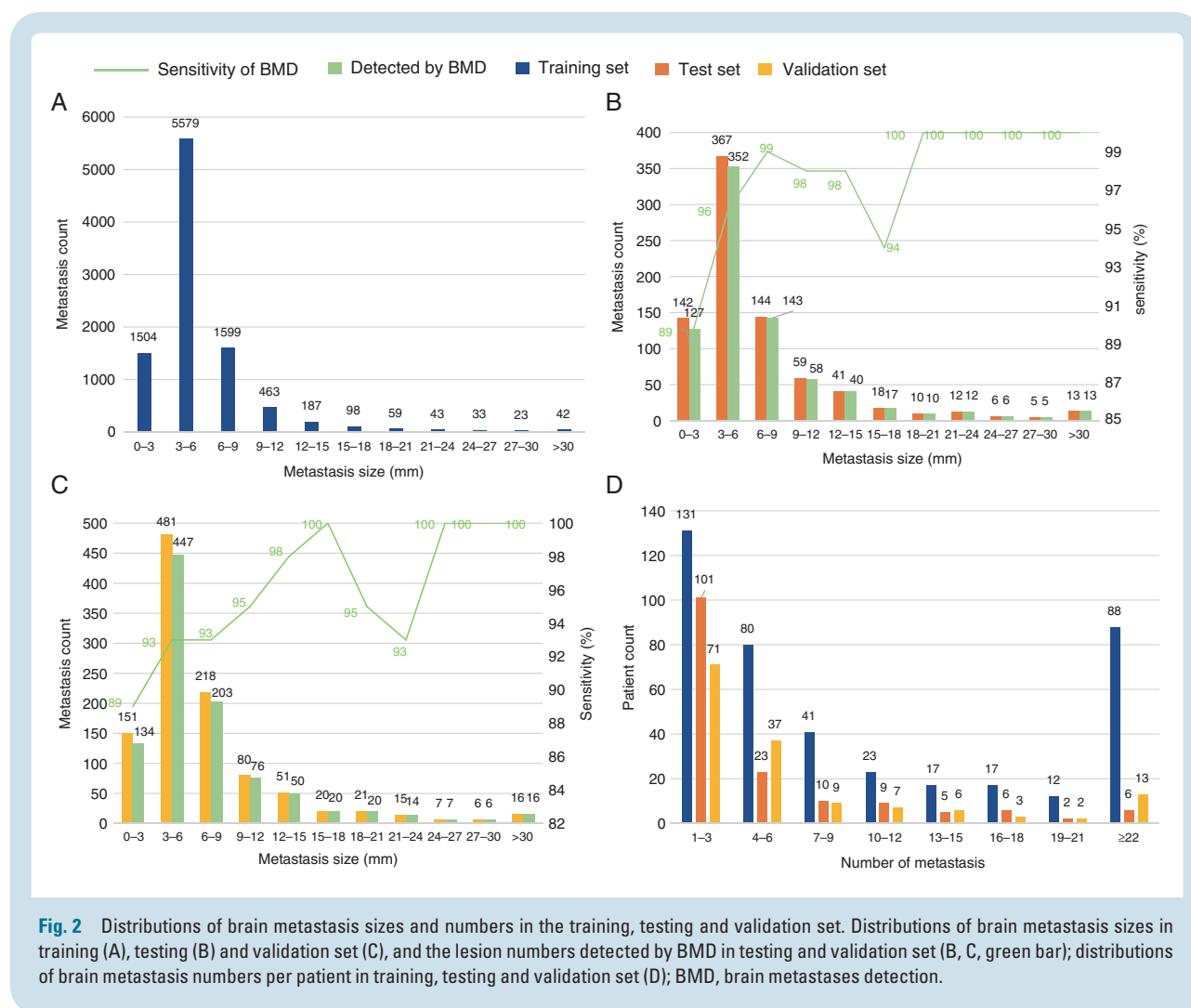
The BMD model performance is summarized in **Figures 2** and **3**. The model achieved detection sensitivities and FPs of 95.8% and 0.39 and 93.2% and 0.38 for the test and validation set, respectively. The detection sensitivities and FPs were 96.0% and 0.27, 95.5% and 0.36, 88.9% and 0.66, and 92.9% and 0.29 for validation sets from SYSUCC, MZPH, DGPB, and FJCH, respectively. Specifically, the sensitivities were 89%, 93%, and 93% for metastases ≤ 3 mm, > 3 mm and ≤ 6 mm, and > 6 mm and ≤ 9 mm in the validation set, respectively. When sets from different centers were analyzed, these sensitivities ranged from 84% to 93%, from 90% to 97%, and from 85% to 97%, respectively. FROC curves showed high and similar detection performance of the BMD for different sets (**Figure 3A**).

Reader Performance

Reader performances and reading times are shown in **Table 2** and **Figures 3** and **4**. For validation set reads without BMD, the mean detection sensitivity was 73% [95% CI: 72–74%], ranging from 69% to 80%. When assisted

by BMD, reader sensitivity significantly improved to 94% [95% CI: 94–95%], $P < .001$. Similar increases were observed among trainees and experienced readers, with mean improvements of 25% and 18%, respectively (70% [95% CI: 69–72%] without BMD vs. 95% [95% CI: 94–96%] with BMD, $P < .001$; 76% [95% CI: 75–78%] without BMD vs. 94% [95% CI: 93–95%] with BMD, $P < .001$). Additionally, the detection sensitivity of each reader improved when using BMD. For the patient-based analysis, the mean AUC of the reading was not significantly different between the two reading modes (0.960 [95% CI: 0.949–0.970] without BMD vs 0.948 [95% CI: 0.937–0.960] with BMD, $P > .05$, **Figure 3B**), even when trainee and experienced reader subgroups were analyzed (0.957 [95% CI: 0.942–0.972] without BMD vs 0.958 [95% CI: 0.943–0.973] with BMD, $P > .05$); 0.963 [95% CI: 0.948–0.977] without BMD vs. 0.939 [95% CI: 0.921–0.957] with BMD, $P > .05$, respectively, **Figure 3C**). One of the trainees' reading AUC was significantly improved with BMD (0.875 [95% CI: 0.831–0.918] without BMD vs. 0.964 [95% CI: 0.939, 0.988] with BMD, $P < .001$; **Figure 3D**).

For all readers, the mean reading time per case was $140 \text{ s} \pm 100$ without BMD and $85 \text{ s} \pm 57$ with BMD (**Table 2**), representing a 40% decrease. Reading times for trainees and experienced readers were reduced by 47% and 32%,



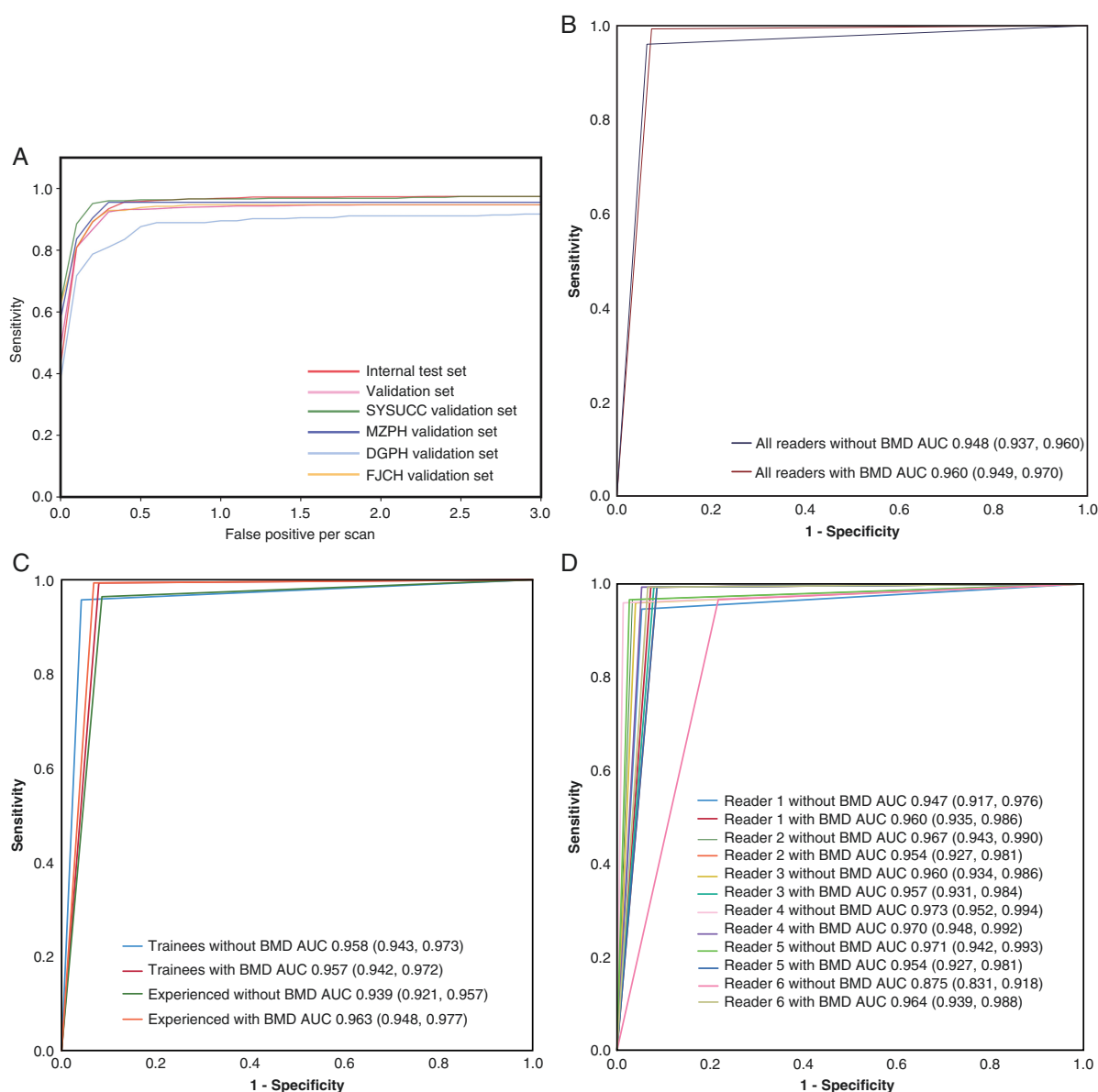


Fig. 3 The FROCs of BMD and ROCs of readers. The lesion-based FROC of BMD on test and validation set (A); the patient-based ROC of readers on validation set (B-D). Reader 1, 2, and 3 were trainees, and reader 4, 5, and 6 were experienced radiologists. FROC, free-response receiver operating characteristic; BMD, brain metastases detection; ROC, receiver operating characteristic; SYSUCC, Sun Yat-sen University Cancer Center; MZPH, Meizhou People's Hospital; DGPH, Dongguan People's Hospital; FJCH, Fujian Cancer Hospital; validation set prospectively included from participating hospitals.

respectively, with BMD. The time-saving benefits were more apparent among the trainees, with a 60% reduction observed for Reader 3. The distributions of reading time per case with and without BMD for each reader are shown in Figure 4. We also analyzed reading time for cases with and without BMs. For cases with BMs, the mean relative time-savings were 42%, 48%, and 35% for all readers, trainees, and experienced readers, respectively, and 29%, 44%, and 28% for cases without BMs. The results indicated that BMD could shorten reading times for patients with and without BMs.

Discussion

In this multi-center study, we developed a cascaded model based on FPN for BM detection on 3D CET1WI MRI across a large number of patients. We assessed robustness using four prospective cohorts, including three external sets. Additionally, we conducted a multi-center multi-reader assessment to evaluate the impact of BMD assistance on reader detection performance and reading time. The BMD system successfully detected 93.2% of BMs with low FPs

Table 2 Detection Sensitivities and Reading Time of Readers on the Validation Set

| Variables | Sensitivity (%; 95%CI) | | Reading time (seconds) | | | | Difference between without BMD and with BMD | | | | Time change (%) | | | | P | | | | |
|--------------------|------------------------|------------|------------------------|-----------------|--------------------|-----------|---|--------------------|-----------------|-----------------|--------------------|-----------|-----------------|--------------------|----|-----------------|-----------------|--------------------|--|
| | Without BMD | With BMD | Without BMD | | With BMD | | all cases | | cases with BMDs | | cases without BMDs | | all cases | | | cases with BMDs | | cases without BMDs | |
| | | | all cases | cases with BMDs | cases without BMDs | all cases | cases with BMDs | cases without BMDs | all cases | cases with BMDs | cases without BMDs | all cases | cases with BMDs | cases without BMDs | | all cases | cases with BMDs | cases without BMDs | |
| Trainee | | | | | | | | | | | | | | | | | | | |
| Reader 1 | 69(66, 71) | 95(93, 96) | <.001 | 135 ± 90 | 155 ± 103 | 116 ± 70 | 60 ± 39 | 70 ± 49 | 49 ± 20 | 75 ± 70 | 84 ± 78 | 68 ± 60 | 56 | 54 | 59 | <.001 | <.001 | <.001 | |
| Reader 2 | 70(68, 73) | 95(93, 96) | <.001 | 165 ± 96 | 194 ± 115 | 138 ± 63 | 66 ± 32 | 77 ± 36 | 56 ± 24 | 99 ± 78 | 116 ± 95 | 83 ± 50 | 60 | 60 | 60 | <.001 | <.001 | <.001 | |
| Reader 3 | 72(69, 75) | 94(93, 96) | <.001 | 130 ± 105 | 154 ± 122 | 106 ± 79 | 103 ± 52 | 110 ± 63 | 96 ± 38 | 26 ± 101 | 44 ± 117 | 9 ± 80 | 20 | 29 | 8 | 0.003 | 0.003 | 0.866 | |
| Average | 70(69, 72) | 95(94, 96) | <.001 | 143 ± 98 | 168 ± 115 | 120 ± 72 | 76 ± 46 | 86 ± 54 | 67 ± 35 | 67 ± 89 | 81 ± 102 | 53 ± 72 | 47 | 48 | 44 | <.001 | <.001 | <.001 | |
| Experienced | | | | | | | | | | | | | | | | | | | |
| Reader 4 | 80(78, 83) | 93(91, 94) | <.001 | 115 ± 100 | 142 ± 123 | 88 ± 61 | 76 ± 62 | 90 ± 73 | 63 ± 46 | 38 ± 58 | 52 ± 74 | 25 ± 31 | 33 | 37 | 28 | <.001 | <.001 | <.001 | |
| Reader 5 | 75(72, 78) | 95(94, 96) | <.001 | 164 ± 107 | 193 ± 129 | 136 ± 71 | 93 ± 62 | 105 ± 81 | 82 ± 31 | 71 ± 89 | 88 ± 106 | 54 ± 64 | 43 | 46 | 40 | <.001 | <.001 | <.001 | |
| Reader 6 | 69(66, 72) | 94(92, 95) | <.001 | 133 ± 92 | 151 ± 103 | 114 ± 76 | 109 ± 65 | 123 ± 73 | 96 ± 54 | 23 ± 95 | 29 ± 105 | 18 ± 83 | 17 | 19 | 16 | 0.001 | 0.008 | 0.048 | |
| Average | 76(75, 78) | 94(93, 95) | <.001 | 137 ± 102 | 162 ± 121 | 113 ± 72 | 93 ± 65 | 106 ± 77 | 80 ± 46 | 44 ± 85 | 56 ± 99 | 32 ± 65 | 32 | 35 | 28 | <.001 | <.001 | <.001 | |
| All readers | 73(72, 74) | 94(94, 95) | <.001 | 140 ± 100 | 165 ± 118 | 104 ± 69 | 85 ± 57 | 96 ± 67 | 74 ± 42 | 56 ± 88 | 69 ± 101 | 30 ± 69 | 40 | 42 | 29 | <.001 | <.001 | <.001 | |

^a lesion-based sensitivity.

^b brain metastasis detection.

^c brain metastases.

^d P value for reading time difference between with and without BMD; CI, confidence interval.

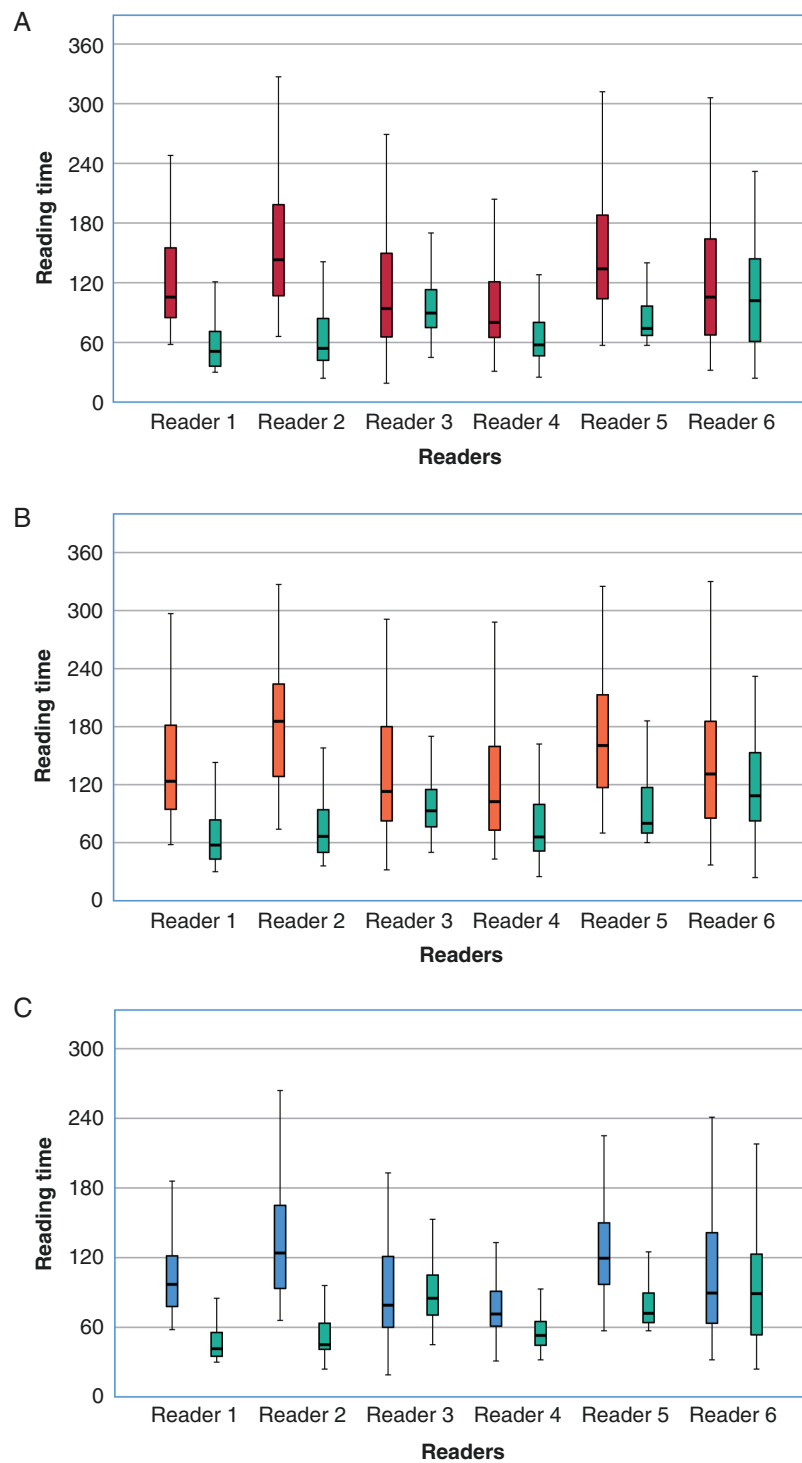


Fig. 4 Reading time of readers with and without assistance of BMD on validation set. Reading time per case of readers for all patients (A), patients with brain metastases (B) and control patients (C), respectively. BMD, brain metastases detection; reader 1, 2, and 3 were trainees, and reader 4, 5, and 6 were experienced radiologists; boxes indicate interquartile range, crosses indicate mean, center lines indicate median, and whiskers indicate range.

of 0.38 per scan in the validation set and yielded superior sensitivities in the three external validation sets, ranging from 88.9% to 95.5%. Furthermore, our model achieved the

highest reported sensitivity of 89% (134/151) for detecting BMs ≤ 3 mm and 97% (173/178) for BMs ≥ 10 mm, the target size criteria for response evaluation recommended

in the Response Assessment in Neuro-Oncology Brain Metastases.³² Additionally, this system assisted radiologists with various experience levels to improve their detection effectiveness, increasing the mean sensitivity by 21% and decreasing the mean reading time by 40%. Hence, the developed network may provide a robust and suitable tool for the early screening of patients with high BM risk,³³ routine follow-ups, and response evaluations.

In a single-center study, Farjam et al.¹³ prospectively enrolled 29 patients to develop a set of unevenly spaced 3D spherical shell templates to detect BMs in post-Gd T1WI, obtaining a 93.5% sensitivity and 0.024 intracranial FP rate for 186 metastases < 5 mm. However, the ground truth was established by a radiologist with unspecified expertise, and the clinical impact of their system was not evaluated. Additionally, a randomized multi-reader multi-case study of 10 patients with a total of 23 tumors from a tertiary hospital demonstrated that CAD increased the mean detection sensitivity of six physicians from 82.6% to 91.3%.²² However, the study focused on tumor segmentation more than detection and did not analyze the reading time. Our literature search returned only one similar multi-center study by Xue et al.,²⁰ who retrospectively included 1652 patients from three centers to construct a model for detecting and segmenting BMs. The model significantly reduced radiation oncologists' segmentation time, but FPs and reading time were not assessed. Reports on the underlying BM detection efficiency improvement in the clinical application of CAD are limited. A similar CAD-aided BM detection study reported that readers took an additional 72 s per case using CAD after complete interpretation; however, performance was slightly improved, with a figure-of-merit AUC increasing from 0.874 to 0.898 using CAD.

Unlike most previous publications, the current study used a multi-center multi-reader prospective validation design, in which the performance and clinical significance of the model were comprehensively assessed. It should also be noted that cases without BMs were included in the multi-reader interpretation. We believe this scenario better reflects the real-world setting in which radiologists work. We observed a reading time reduction for both metastatic and non-metastatic cases; thus, our BMD system will provide an efficient assistant tool that can be integrated into the clinical workflow as a concurrent-reading mode. At the same time, several issues should not be ignored when using BMD. First, we analyzed the average number of FPs per case in the BMD only and reader with BMD groups. The obtained numbers were 0.38 for AI only and 0.23 for reader with BMD, respectively. This suggests that the results of BMD should be carefully reviewed and identified by radiologists, especially for the trainees. Second, not in all cases, the reading time was reduced with BMD. For cases with limited BMs, if BMD has multiple FPs, readers will spend time distinguishing FPs from real ones, potentially resulting in longer reading time. Third, users should pay close attention to the lesions that were easy to be overlooked by BMD, especially for the small metastases with slight enhancement or non-enhancement ([Supplementary Figure 3B](#)). Nonetheless, we believe that human and AI are complementary. BMD has been integrated into our radiology workstation since May 2021 as a right-hand partner and has been optimized through clinical application.

Previous studies have developed methods for the detection and/or segmentation of BMs across a diverse range of lesion numbers (1–180 per person) and sizes (1–52 mm),^{13–27} utilizing either classical machine learning^{13,14,17,25,26} or DL^{15,16,18–24,27} and diverse sequences (single, two,^{23,26} three,¹⁵ or a combination¹⁸). Among them, the overall sensitivity and FPs per person ranged from 81.1%²⁶ to 100%²⁰ and 0.59²³ to 302,²⁵ respectively. The current work investigated the largest number of metastases (11 514) across the most extensive ranges of metastasis numbers per patient (1–587) and lesion sizes (1.2–72.1 mm), with the highest proportion of small lesions ≤ 5 mm (58.6%, 6754/11 514), covering a diverse set of 32 primary tumor subtypes. We believe our study was more representative of real-world data. Regardless of the notable abilities of multi-sequence mode,^{14,17,22,25} the costs of additional scan time and sequence availability may prevent wide clinical implementation. Our method, which only used CET1WI sequences, produced a high sensitivity of 89% for detecting metastases ≤ 3 mm (example provided in [Supplementary Figure 3A](#)), higher than the previously reported maximum sensitivity of 82.4%²² for metastases of this size. Considering the trade-off between sensitivity and feasibility, our model may be more suitable for broad practical application. Additionally, we found that the sensitivity of our network on validation set was weakly correlated with the lesion size ($P = .023$, odds ratio: 1.079, 95% CI: 1.010–1.151), but was not related to the number of lesions per patient ($P = .984$). FPN was applied to boost sensitivity; however, this should be optimized for a few small lesions with lower contrast than that of normal brain parenchyma.

Compared with existing DL techniques for BM detection, our approach contains several unique aspects. First, we included negative controls to train, test, and validate our system, unlike previous studies designating controls either in testing or validating.^{14,23,25} An intelligent model must learn to consistently distinguish normal anatomy in addition to tumors, much like a human being would. Additionally, these controls were used to evaluate the specificity of the DL model. Another feature that separates this study from previous works is the strength of using 16 types of scanners manufactured by four suppliers, diverse images from four institutions, and the largest number of participating centers and readers. It is more challenging to maintain an overall high performance for BMD considering these factors. The clinical applicability and advancement of existing BMD models have been limited by the retrospective nature, small sample sizes, and/or single institutional research at similar-tier hospitals. Thus, the robustness and general applicability of previously developed networks remain unknown. Our BMD system was developed and validated using a large cohort of over 1,000 patients from different-tier hospitals and exhibited an overall high sensitivity (89–96%) for BM detection in the validation sets, which strongly suggests this BMD is applicable in diverse real-world scenarios. Furthermore, differences in the network architecture set our BMD system apart. A multi-scale network based on FCN with different shapes of anchor boxes at different scales was employed for various BM sizes and shapes, i.e., detecting by lesion sub-classifications. Subsequently, a multiple serial cascade network was designed to reject FPs, and the results

were then judged by the 3D variant of LeNet to further decrease the number of FPs.

FPs and false negatives were inevitable in our study. Both metastases and blood vessels can manifest as a nodule or punctuation of hyperintensity on CET1WI, and the latter may be mislabeled as metastasis. Most FPs in this model derived from blood vessels (77%; examples provided in [Supplementary Figure 3C and 3D](#)), consistent with other studies^{13,14,17,21} that reported blood vessels accounted for as many as 62%–88% of FPs.^{14,17,21} FPs can be removed by means of size, hyperintensity, and sphericity criteria,¹³ the degree of anisotropy method,¹⁴ artificial neural network algorithms,^{25,26} the RUSBoost algorithm,²¹ and 3D contrast-enhanced black-blood imaging²⁶ that can suppress blood vessel signals, enabling clearer delineation and better detection of small BMs.³⁴ It was reported that black-blood imaging could improve detection sensitivity while reducing small-nodule FPs to 0.42–0.59 per person.^{23,26} Different from previous studies, the current model rejected FPs via a cascade network with multiple serial neural network blocks. Our observer study revealed that FPs are easily recognized by radiologists by tracing anatomical structures in multiple planes. Thus, misdirection would probably be avoided in practice and would not require additional reading time, suggesting that our FPs were clinically acceptable. Our method failed to identify 6.8% of BMs, almost all of which were small lesions with low contrast to surrounding parenchyma or annular enhancement, consistent with previous studies.^{13,16,21} Earlier studies found that CAD methods were likely to overlook lesions that were attached to vessels, had low contrast to the background,^{13,21} or were close to the surface or gray matter–white matter interfaces of the cerebrum.¹⁶ Park et al.²³ combined 3D black-blood imaging with 3D gradient echo in a DL model, which increased the sensitivity from 25.5% to 82.4% for nodules ≤ 3 mm relative to the 3D gradient echo model alone. Compared with single-scale detection networks that only perform detection on the finest level, different scales of image features can be utilized in our multi-scale detection network to detect diverse BMs and improve the sensitivity for tiny lesions. In addition, precisely segmenting vessels from brain tissues may be a potential strategy to detect more lesions adjacent to vessels. The detection of small BMs with lower contrast to the background remains a challenge for CAD and also experienced radiologists.

The potential limitations of this study must be discussed. First, the training and test sets were labeled retrospectively, which might have led to a certain level of selection bias; however, the prospective validation suggests that this limitation is not prominent. Second, only post-contrast T1-weighted MR images were used for this study to reflect what is commonly used in practice. Third, the current model was designed for thin-slice images, and the performance may be reduced for thick-slice images, such as 5 mm. We are training a thick-slice model to address this issue. Fourth, the impacts of other algorithms on reading times were not available for comparison, but our network showed high performance and significantly reduced radiologists' reading time relative to their reading times without BMD. Finally, the current method does not automatically segment the mass contour nor identify the primary tumor types, which we will also address in our ongoing work.

Conclusions

We developed and prospectively validated an automatic system employing FPN for BM detection on CET1WI MR images from four institutions. Our multi-center evaluation showed that this system facilitated radiologists with various levels of experience to achieve a high level of detection sensitivity and reduced reading time using the concurrent-reading mode. Our findings provide evidence for CAD of BMs in clinical translation with improved confidence.

Supplementary material

Supplemental material is available at *Neuro-Oncology* online.

Keywords

automatic detection | brain metastases | cascaded convolutional network | MRI

Acknowledgments

The authors acknowledge the AsiaEdit for language editing.

Funding

None.

Conflict of interest. All authors declare no potential conflicts of interest.

Authorship statement. Design: Y.S., L.X., Y.Y., G.Y., S.D., S.Y., Z.R., X.C. Implementation: Y.S., M.L., L.S., Y.Q., W.D., L.Y., M.Z., F.W., Z.D., L.J., C.F., Z.Y., Z.X., S.F., S.G., W.J., S.Y., L.C. Analysis and interpretation of the data: Y.S., L.X., Y.Y., S.Y. Writing of the manuscript: Y.S., L.X., Y.D.

Review and editing: L.H., G.Y., S.D., Z.R., X.C. Approval of the final version: all authors

Data availability

Data supporting the study results and model codes can be accessed via the corresponding author upon reasonable request.

References

- Achrol AS, Rennert RC, Anders C, et al. Brain metastases. *Nat Rev Dis Primers*. 2019;5(1):5.
- Suh, JH, Kotecha, R, Chao ST, et al. Current approaches to the management of brain metastases. *Nat Rev Clin Oncol*. 2020;17(5):279–299.
- Tong E, McCullagh KL, Iv M. Advanced imaging of brain metastases: from augmenting visualization and improving diagnosis to evaluating treatment response. *Front Neurol*. 2020;11:270.
- Soffietti R, Abacioglu U, Baumert B, et al. Diagnosis and treatment of brain metastases from solid tumors: guidelines from the European Association of Neuro-Oncology (EANO). *Neuro Oncol* 2017;19(2):162–174.
- Chao ST, De Salles A, Hayashi M, et al. Stereotactic radiosurgery in the management of limited (1-4) brain metastases: systematic review and international stereotactic radiosurgery society practice guideline. *Neurosurgery* 2018;83(3):345–353.
- Tsao, M, Xu, W, Sahgal A. A meta-analysis evaluating stereotactic radiosurgery, whole-brain radiotherapy, or both for patients presenting with a limited number of brain metastases. *Cancer* 2012;118(9):2486–2493.
- Yamamoto M, Serizawa T, Shuto T, et al. Stereotactic radiosurgery for patients with multiple brain metastases (JLGK0901): a multi-institutional prospective observational study. *Lancet Oncol*. 2014;15(4):387–395.
- Ammirati, M, Nahed, BV, Andrews D, et al. Congress of neurological surgeons systematic review and evidence-based guidelines on treatment options for adults with multiple metastatic brain tumors. *Neurosurgery* 2019;84(3):E180–E182.
- Hartgerink D, Bruynzeel A, Eekers D, et al. A Dutch phase III randomized multicenter trial: whole brain radiotherapy versus stereotactic radiotherapy for 4-10 brain metastases. *Neurooncol Adv* 2021;3(1):vdab021.
- Kavi A, Gurewitz J, Benjamin CG, et al. Hippocampal sparing in patients receiving radiosurgery for ≥ 25 brain metastases. *Radiother Oncol*. 2021;161:65–71.
- Fecci P, Champion C, Hoj J, et al. The evolving modern management of brain metastasis. *Clin Cancer Res*. 2019;25(22):6570–6580.
- Suh, CH, Jung, SC, Kim, KW, et al. The detectability of brain metastases using contrast-enhanced spin-echo or gradient-echo images: a systematic review and meta-analysis. *J Neurooncol*. 2016;129(2):363–371.
- Farjam, R, Parmar, H, Noll, D, et al. An approach for computer-aided detection of brain metastases in post-Gd T1-W MRI. *Magn Reson Imaging*. 2012;30(6):824–836.
- Pérez, R, Arana, E, Moratal D. Brain metastases detection on MR by means of three-dimensional tumor-appearance template matching. *J Magn Reson Imaging*. 2016;44(3):642–652.
- Gr?vik, E, Yi, D, Iv M, et al. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. *J Magn Reson Imaging*. 2020;51(1):175–182.
- Zhou, Z, Sanders, JW, Johnson JM, et al. Computer-aided detection of brain metastases in T1-weighted mri for stereotactic radiosurgery using deep learning single-shot detectors. *Radiology* 2020;295(2):191479.
- Ambrosini, RD, Wang, P, O'Dell WG. et al. Computer-aided detection of metastatic brain tumors using automated three-dimensional template matching. *J Magn Reson Imaging*. 2010;31(1):85–93.
- Charron O, Lallement A, Jarnet D, et al. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Comput Biol Med*. 2018;95:43–54.
- A ZZ, A JWS, B JMJ, et al. MetNet: Computer-aided segmentation of brain metastases in post-contrast T1-weighted magnetic resonance imaging. *Radiother Oncol*. 2020;153:189–196.
- Xue J, Wang B, Ming Y, et al. Deep learning-based detection and segmentation-assisted management of brain metastases. *Neuro Oncol* 2020;22(4):505–514.
- Zhang M, Young GS, Chen H, et al. Deep-learning detection of cancer metastases to the brain on MRI. *J Magn Reson Imaging*. 2020;52(4):1227–1236.
- Lu SL, Xiao FR, Cheng JC, et al. Randomized multi-reader evaluation of automated detection and segmentation of brain tumors in stereotactic radiosurgery with deep neural networks. *Neuro Oncol*. 2021;23(9):1560–1568.
- Park YW, Jun Y, Lee Y, et al. Robust performance of deep learning for automatic detection and segmentation of brain metastases using three-dimensional black-blood and three-dimensional gradient echo imaging. *Eur Radiol*. 2021;31(9):6686–6695.
- Cho, SJ, Sunwoo, L, Baik SH, et al. Brain metastasis detection using machine learning: a systematic review and meta-analysis. *Neuro Oncol* 2021;23(2):214–225.
- Sunwoo L, Kim YJ, Choi SH, et al. Computer-aided detection of brain metastasis on 3D MR imaging: Observer performance study. *PLoS One*. 2017;12(6):e0178265.
- Yang, S, Nam, Y, Kim MO, et al. Computer-aided detection of metastatic brain tumors using magnetic resonance black-blood imaging. *Invest Radiol*. 2013;48(2):113–119.
- Dikici E, Ryu JL, Demirer M, et al. Automated brain metastases detection framework for T1-weighted contrast-enhanced 3D MRI. *IEEE J Biomed Health Inform* 2020;24(10):2883–2893.
- Lin T-Y, Dollar P, Girshick R, et al. Feature Pyramid Networks for Object Detection. Paper presented at: 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2017 Jul 21–26, 2017; Honolulu, HI.
- Lecun, Y, Bottou, L. Gradient-based learning applied to document recognition. *Proc IEEE*. 1998;86(11):2278–2324.
- Breiman, L. Bagging predictors. *Mach Learn*. 1996;24(2):123–140.
- PyTorch. 2021; <https://pytorch.org/>. Accessed October 28, 2021.
- Lin NU, Lee EQ, Aoyama H, et al. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol*. 2015;16:e270–e278.
- Network NCC. Central nervous system cancers (Version 5.2020 -April 15, 2021). 2021; https://www.nccn.org/professionals/physician_gls/pdf/cns.pdf. Accessed June 12, 2021.
- Park, J, Kim, EY. Contrast-enhanced, three-dimensional, whole-brain, black-blood imaging: application to small brain metastases. *Magn Reson Med*. 2010;63(3):553–561.