

Enhancing Brain Metastases Detection and Segmentation in Black-Blood MRI Using Deep Learning and Segment Anything Model (SAM)

Sang Kyun Yoo^{1,2*}, Tae Hyung Kim^{1,3*}, Jin Sung Kim^{1,2,4}, Sung Soo Ahn⁵, Eui Hyun Kim⁶,
Wonmo Sung⁷, Hojin Kim¹, and Hong In Yoon¹

¹Department of Radiation Oncology, Yonsei Cancer Center, Heavy Ion Therapy Research Institute, Yonsei University College of Medicine, Seoul;

²Medical Physics and Biomedical Engineering Lab (MPBEL), Yonsei University College of Medicine, Seoul;

³Department of Radiation Oncology, Nowon Eulji Medical Center, Eulji University School of Medicine, Seoul;

⁴Oncosoft Inc., Seoul;

⁵Department of Radiology, Severance Hospital, Research Institute of Radiological Science and Center for Clinical Image Data Science, Yonsei University College of Medicine, Seoul;

⁶Department of Neurosurgery, Brain Tumor Center, Severance Hospital, Yonsei University College of Medicine, Seoul;

⁷Department of Biomedical Engineering and of Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul, Korea.

Purpose: Black-blood (BB) magnetic resonance images (MRI) offer superior image contrast for the detection and segmentation of brain metastases (BMs). This study investigated the efficacy and accuracy of deep learning (DL) architectures and post-processing for BMs detection and segmentation with BB images.

Materials and Methods: The BB images of 50 patients were collected to train (40) and test (10) the DL model. To ensure consistency, we implemented piecewise linear histogram matching for intensity normalization and resampling. Modified U-Net, including combination with generative adversarial network (GAN), was applied to enhance the segmentation performance. The U-Net-based networks generated bounding boxes indicating regions of interest, which were then processed in a post-processing using the Segment Anything Model (SAM). We quantitatively assessed the three U-Net-based models and their post-processed counterparts in terms of lesion-wise sensitivity (LWS), patient-wise dice similarity coefficient (DSC), and average false-positive rate (FPR).

Results: The modified U-Net with GAN yielded a patient-wise DSC of 0.853 and a LWS of 89.19%, which outperformed the standard U-Net (patient-wise DSC of 0.815) and modified U-Net only (patient-wise DSC of 0.846). Combining GAN architecture with modified U-Net also reduced the FPR, less than 1 on average. Post-processing with SAM further did not affect LWS and FPR, but effectively enhanced the patient-wise DSC by 2%–3% for the U-Net-based models.

Conclusion: The modifications to standard U-Net notably improves the detection and segmentation of BMs in BB images, and applying SAM as post-processing can further enhance the precision of segmentation results.

Key Words: Black-blood image, brain metastases, deep learning, auto-segmentation, Segment Anything Model

Received: July 15, 2024 **Revised:** February 13, 2025 **Accepted:** March 7, 2025 **Published online:** May 19, 2025

Co-corresponding authors: Hong In Yoon, MD, PhD, Department of Radiation Oncology, Yonsei Cancer Center, Heavy Ion Therapy Research Institute, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.

E-mail: yhi0225@yuhs.ac and

Hojin Kim, PhD, Department of Radiation Oncology, Yonsei Cancer Center, Heavy Ion Therapy Research Institute, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.

E-mail: hjhenrykim@yuhs.ac

*Sang Kyun Yoo and Tae Hyung Kim contributed equally to this work.

•The authors have no potential conflicts of interest to disclose.

© Copyright: Yonsei University College of Medicine 2025

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Brain metastases (BMs) are the most common intracranial tumors and comprise over half of all clinically diagnosed brain tumors in adults.^{1,2} Palliative whole-brain radiotherapy (WBRT) has been previously recommended. However, with advances in systemic therapy, the treatment perspectives for BMs have evolved. Recently, the American Society of Clinical Oncology Society for Neuro-Oncology-American Society for Radiation Oncology (ASCO-SNO-ASTRO) published guidelines for BMs in solid tumors.³ The evolution of radiotherapy (RT) modalities has significantly altered treatment strategies regarding RT. Firstly, the decreased role of WBRT is presented. Secondly, stereotactic radiosurgery (SRS) is increasingly used to treat BM. SRS offers the advantage of less cognitive decline compared to WBRT, which is particularly recommended for patients with a small number of BMs.⁴ Therefore, with the increasing use of SRS in radiation oncology, accurate detection and delineation of BMs are essential.⁵

Contrast-enhanced magnetic resonance imaging is the preferred imaging modality for detecting and delineating BMs,⁵⁻⁷ allowing the detection of tiny metastases with high spatial resolution and low partial volume effects. Several studies have attempted to automatically detect and segment BMs on T1 weighted gadolinium enhanced (T1Gd) MR images using deep learning (DL) algorithms.⁸⁻¹³ However, T1Gd MR simultaneously enhances BMs and intracranial vessels, causing difficulties in the auto-segmentation of BMs. Black-blood (BB) imaging, on the other hand, is known to selectively suppress blood signals, enhancing the visualization of small lesions compared to T1Gd.^{14,15} Thus, once DL-based auto-segmentation is involved in BB images, more accurate and reliable BMs detection and segmentation can be achieved.¹⁶⁻¹⁹

Although several DL-based algorithms have been developed for the detection and delineation of BMs,⁸⁻¹³ only a few studies have attempted to extend them to BB images.¹⁶⁻¹⁹ Several studies that applied DL-based algorithms to BB images have reported improved BMs detection performance, particularly for small metastases. For instance, Kottlors, et al.¹⁶ demonstrated that a neural network architecture trained on BB images achieved significantly higher accuracy, with an area under the curve of 0.87, compared to 0.53 on T1Gd images. Oh, et al.¹⁷ also evaluated BMs detection using BB images, achieving a sensitivity of 87.95% and an average false-positive rate (FPR) of 14.48. Additionally, Kikuchi, et al.¹⁸ showed that their BMs detection model achieved the sensitivity of 91.7%, with a FPR of 1.5, which was slightly higher than that of the observer test. However, these studies did not directly address the segmentation of BMs on BB images. Another study Park, et al.¹⁹ investigated BMs segmentation on BB images, but their network required combining BB images with other imaging to achieve improved segmentation performance.

In this study, we investigated methods to enhance the per-

formance of BMs auto-segmentation through DL algorithms on BB images. Specifically, we propose 1) modifying the neural network architecture to improve sensitivity and reduce false positives and 2) employing post-processing using the Segment Anything Model (SAM) to refine the segmentation performance.

MATERIALS AND METHODS

Patient cohort

The patient cohort comprised 50 patients diagnosed with BMs from lung cancer who underwent either WBRT or SRS between January 2019 and December 2020 at the Yonsei Cancer Center. The dataset consisted of BB images with manual delineation of the BMs by an expert radiation oncologist. The study was approved by the Institutional Review Board of Severance Hospital (4-2021-0306), and the requirement for informed consent was waived due to the retrospective nature of the study. Table 1 summarizes the baseline characteristics of the patients with BMs. In the training set, five patients had a single tumor, two had two tumors, two had three tumors, and 31 had four or more tumors. In the test set, two patients had a single tumor, one had two tumors, two had three tumors, and five had four or more tumors. Tumor volumes in the training set ranged from 0.003 cc to 101.672 cc, with a median of 0.049 cc. In the test set, volumes ranged from 0.007 cc to 16.642 cc, with a median of 0.069 cc.

Data pre-processing

The MR images in our datasets had different pixel spacing resolutions ranging from 0.195 to 0.417. To ensure consistency,

Table 1. Characteristics of 50 Patients with BM

Characteristic	Total (n=50)	Train-set (n=40)	Test-set (n=10)
Age (median, yr)	62.5 (range 31–81)	64.5 (range 40–81)	60 (range 31–76)
Sex			
Male	25	19	6
Female	25	21	4
Number of patients per tumor			
1 tumor	7	5	2
2 tumors	3	2	1
3 tumors	4	2	2
≥4 tumors	36	31	5
Tumor volume in cc			
Maximum	101.672	101.672	16.642
Minimum	0.003	0.003	0.007
Q1	0.018	0.018	0.024
Q2 (median)	0.049	0.049	0.069
Q3	0.203	0.023	0.254

BM, brain metastases.

we adjusted the data for pixel spacing by setting it to 0.4. The slice thickness for all the scans was consistently 1 mm. The intensity distribution of the MR images showed slight variations across patient scans, although the same simulator and pulse sequence were employed, which necessitated the use of intensity normalization to constrain the variations.²⁰⁻²² Among various methods, we adopted piecewise linear histogram matching,²³ called Nyul normalization, which is a data-driven approach that establishes a standard intensity scale from a set of representative images, which is subsequently applied to each MR image. This process aligns the histogram of the MR images with the landmarks of the standard scale, thereby enhancing the consistency across different images.

DL models

The U-Net architecture,²⁴ which is the most widely used architecture for semantic segmentation, consists of an encoding phase to capture context and a symmetric decoding phase for precise localization. In this study, the encoder had four layers, each comprising a convolutional layer, followed by batch normalization, ReLU activation, and a pooling layer. Similarly, the decoder had four layers, each incorporating a transposed convolutional layer for upsampling, followed by concatenation with the corresponding feature map from the encoder through skip connections. In this study, the network input and output were aggregated into three consecutive 2D slices, capturing partial volume information during the training of the network.

The major concern of the standard U-Net architecture is the loss of detailed image gradient information, which is mostly attributed to pooling operations in the encoding phase. A skip connection was devised to mitigate this loss by introducing the imaging features in the encoder to the newly upsampled features in the decoder. The modified U-Net provides additional amendments to conventional skip connections²⁵ as illustrated in Fig. 1A. The image features before applying the max-pooling operation (downsampling) were transmitted to the skip connection. The downsampled images resulting from pooling were then processed through the transpose convolution (deconvolution) operator, and the result was subtracted from the image features in the skip connection. This subtraction process allowed for identifying the imaging information lost during the pooling operation. The image-loss information was plugged into the newly upsampled features in the decoder, which helped preserve the image gradient information more effectively than the conventional skip connection. The modified U-Net, which was initially proposed by Seo, et al.,²⁵ includes an object-dependent adaptive filter that is activated depending on the object size. However, we chose not to incorporate this modification in the present study since our dataset predominantly consisted of small BMs (median volume, 0.049 cc).

To enhance the BMs detection and segmentation performance, the modified U-Net architecture uses a generative adversarial network (GAN)²⁶ consisting of a generator and discrim-

inator (Fig. 1B). The generator consisted of a modified U-Net. The discriminator distinguishes between real and generated segmentation maps. The basic idea of a GAN is to compete with the generator against the discriminator, which makes it challenging to discern which of the true and generated segmentation maps is genuine. The architecture of the discriminator consisted of a series of four layers, each comprising a convolutional layer, followed by batch normalization and PReLU activation.

SAM as post-processing

SAM is an open-source software provided by Meta AI (New York City, NY, USA) comprising three main components: an image encoder, a prompt encoder, and a mask decoder. The image encoder utilizes a vision transformer (ViT)²⁷ pre-trained with a mask autoencoder²⁸ to capture detailed image features. The prompt encoder processes manual annotations using positional encoding. The mask decoder efficiently maps the image and prompts embeddings, along with an output token, to produce a mask through self-attention and cross-attention. We opted to use SAM for post-processing, focusing its application on refining segmentation within predefined bounding boxes generated by the three U-Net-based networks. The three U-Net-based networks generated different bounding boxes around the BM-segmented regions, with selected the regions of interest (ROIs) serving as the SAM prompt encoder input. Finally, SAM refines the segmentation within these boxes to enhance accuracy (Fig. 1C).

Evaluation

All training tasks in this study utilized a single graphic processing unit (NVIDIA TITAN RTX). The investigated networks were implemented using the PyTorch framework.²⁹ Preprocessing steps, including resampling and intensity normalization, were implemented using scikit-learn³⁰ and an intensity-normalization library.²² During the training process, a batch size of four was selected, and the models underwent early stopping after 100 epochs. The networks were optimized using the Adam optimizer and DICE loss function. The learning rate was initially set to $2e-5$, and the scheduler dynamically adjusted it based on the validation loss. Furthermore, it reduced the rate by a factor of 0.5 if there was no improvement throughout two epochs; otherwise, the rate was reduced by a factor of zero.

To evaluate the robustness of our method, we set aside a test set comprising data from 10 randomly selected patients out of the total of 50 patients. The patient scans were excluded from the training process. The performances of the investigated networks, both with and without post-processing, were assessed using the dice similarity coefficient (DSC), lesion-wise sensitivity (LWS), and average FPR. The DSC values represent the accuracy of BMs segmentation, while the LWS and average FPR are associated with the detection capability of BMs on BB images. Statistical analyses were performed using SPSS ver. 27

(IBM Corp., Armonk, NY, USA), and significance was assessed using the Wilcoxon signed-rank test for non-normally distributed data.

RESULTS

Performance of three U-Net-based models

The comparison metrics for evaluating the performances of the three U-Net-based models are listed in Table 2. The modified

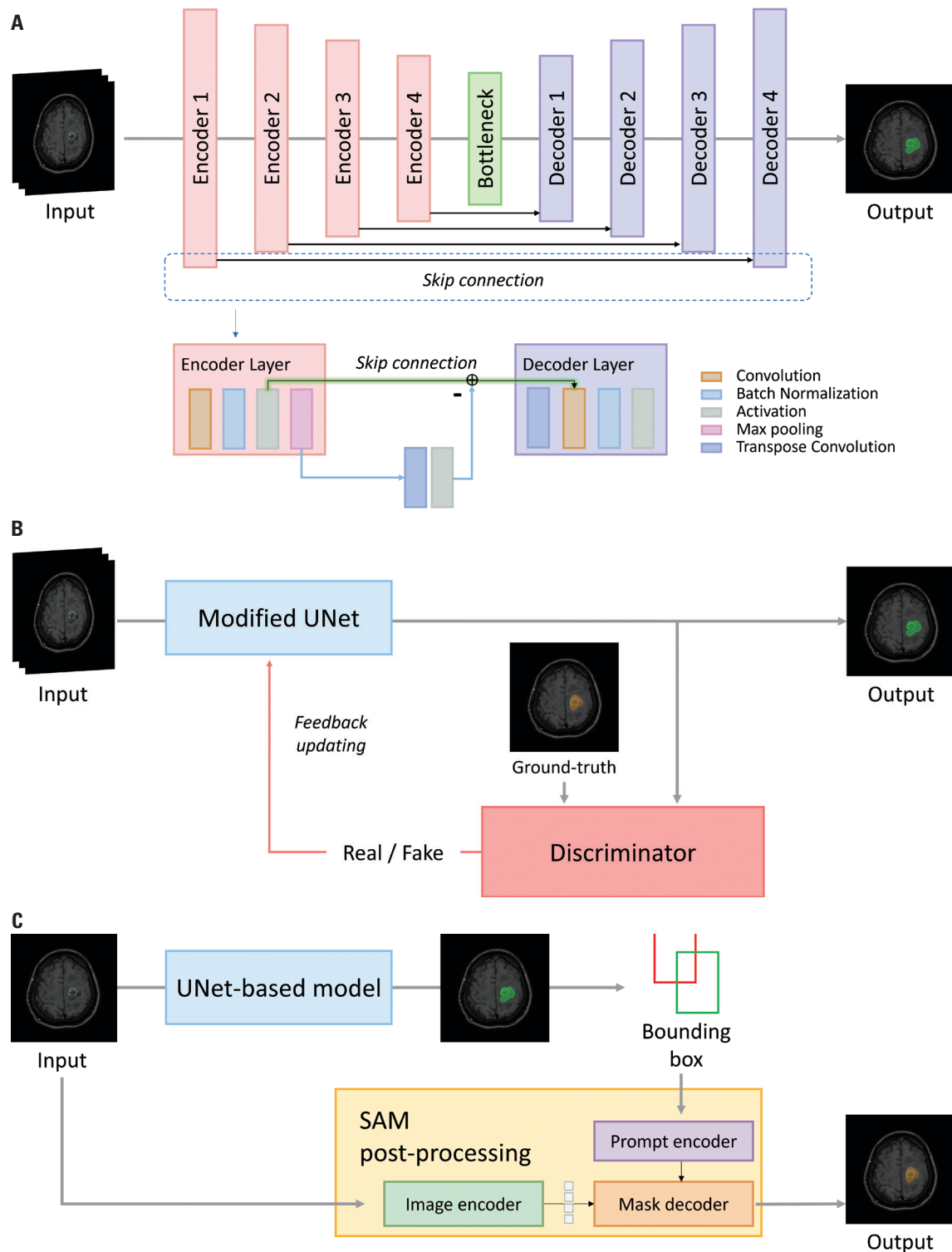


Fig. 1. A deep learning approach enhancing BMs detection and segmentation. (A) A modified U-Net. (B) A modified U-Net with GAN. (C) The post-processing step using SAM. BM, brain metastases; GAN, generative adversarial network; SAM, Segment Anything Model.

U-Net equipped with a specialized skip connection yielded a patient-wise DSC of 0.846, a LWS of 89.19%, and an average FPR of 1.4. These results outperformed those of the standard U-Net with a conventional skip connection. Additionally, as shown in Fig. 2, the modified U-Net with the GAN, including

Table 2. Comparison of the Results of Three U-Net-Based Models for BMs Detection and Segmentation

	Standard U-Net	Modified U-Net	Modified U-Net with GAN
LWS (%)	87.84	89.19	89.19
Average FPR	2.2	1.4	0.9* ($p=0.024$)
Patient-wise DSC	0.815 (± 0.09)	0.846 (± 0.06)	0.853 (± 0.06)
Tumor volume range (lesion-wise DSC)			
≥ 0.1 cc	0.724 (± 0.27)	0.712 (± 0.28)	0.715 (± 0.28)
0.06–<0.1 cc	0.754 (± 0.05)	0.761 (± 0.08)	0.763 (± 0.08)
0.04–<0.06 cc	0.540 (± 0.34)	0.635 (± 0.27)	0.674 (± 0.17)
0.02–<0.04 cc	0.567 (± 0.22)	0.553 (± 0.20)	0.575 (± 0.17)
<0.02 cc	0.515 (± 0.34)	0.558 (± 0.38)	0.556 (± 0.37)

BM, brain metastases; GAN, generative adversarial network; LWS, lesion-wise sensitivity; FPR, false-positive rate; DSC, dice similarity coefficient.

Asterisk (*) indicates the statistical significance of differences between the modified U-Net with GAN and the standard U-Net, as determined by p -values.

the discriminator, enhanced the segmentation performance, achieving a patient-wise DSC of 0.853, while the LWS remained unchanged. Notably, the modified U-Net with GAN significantly reduced the average FPR to less than 1 compared to the standard U-Net ($p=0.024$). While the improvement over the modified U-Net did not reach statistical significance ($p=0.059$), it still demonstrated a consistent trend toward enhanced performance.

The lesion-wise DSC across varying BMs volumes is also presented in Table 2. For BMs with volumes in the range of 0.04–0.06 cc, the modified U-Net and modified U-Net with GAN achieved notable improvements, yielding lesion-wise DSC of 0.635 and 0.674, respectively, as compared to the standard U-Net's 0.540. This indicates enhanced segmentation performance for smaller BMs. For BMs with volumes over 0.1 cc, the lesion-wise DSC across all three models was consistently around 0.715–0.724. However, for BMs with very small volumes, below 0.04 cc, the lesion-wise DSC was lower across all models.

Table 3 demonstrates that the LWS remained consistently high at 100% across all models when the number of BMs per patient was less than four, aligning with the ASCO-SNO-ASTRO guidelines 4 recommending SRS for such cases. Performance differences across models became apparent as the

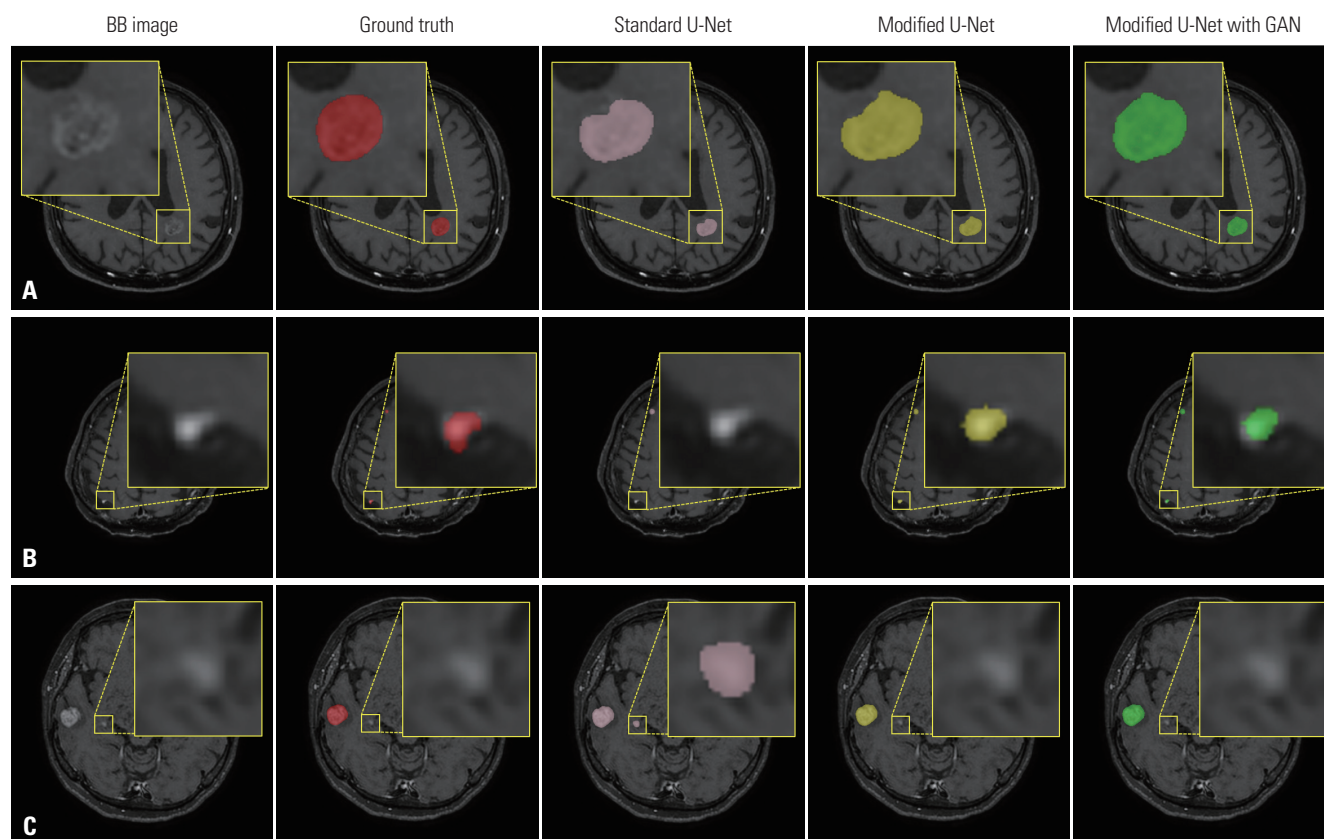


Fig. 2. The comparison of three U-Net-based models. The figure demonstrates the improvements in DSC, sensitivity, and false positives across different U-Net model variations. (A) True positives are shown, highlighting DSC improvements across the standard U-Net, modified U-Net, and modified U-Net with GAN. (B) False negatives are presented, illustrating the increase in sensitivity achieved by the modified models. (C) False positives are displayed, indicating the reduction in false positives with both modified models. Yellow bounding boxes indicate zoomed-in images to make differences clearer. DSC, dice similarity coefficient; GAN, generative adversarial network.

number of BMs per patient equaled or exceeded four. Regarding the volume of BMs, the largest variation across the models occurred in the range of 0.04–0.06 cc. Within this range, the standard U-Net yielded a relatively low LWS (88.89%), whereas the other two models achieved a LWS of 100%.

In Table 4, regardless of the number of BMs, a consistent trend is shown, in which the average FPR decreases in the order of standard U-Net, modified U-Net, and modified U-Net with GAN. With fewer than four BMs, the modified U-Net only and the modified U-Net with GAN achieved an average FPR of less than 1. When the number of BMs was equal to or exceeded four, the gaps among the models increased. Importantly, the modified U-Net with GAN constrained the FPR to approximately one under such conditions.

Table 3. LWS Depends on the Number of BMs Per Patient and the Volume Range of BMs

	Standard U-Net	Modified U-Net	Modified U-Net with GAN
Patients with tumor count <4	100	100	100
Patients with tumor count ≥4	84.75	86.44	86.44
Tumor volume range			
≥0.1 cc	93.55	93.55	93.55
0.06–<0.1 cc	100	100	100
0.04–<0.06 cc	88.89	100	100
0.02–<0.04 cc	92.31	92.31	92.31
<0.02 cc	66.67	66.67	66.67

LWS, lesion-wise sensitivity; BMs, brain metastases; GAN, generative adversarial network.

Table 4. Average FPR Depending on the Number of BMs Per Patient

	Standard U-Net	Modified U-Net	Modified U-Net with GAN
Patients with tumor count <4	1.17	0.83	0.67
Patients with tumor count ≥4	3.75	2.25	1.25

FPR, false-positive rate; BMs, brain metastases; GAN, generative adversarial network.

Performance of post-processed models

Table 5 lists the numerical performance of SAM-based post-processing models. The post-processing tasks enhanced segmentation performance, increasing the patient-wise DSC by 2% for the U-Net-based models, with statistical significance observed for the standard U-Net ($p=0.017$), the modified U-Net ($p=0.019$), and the modified U-Net with GAN ($p=0.037$). The modified U-Net with GAN achieved a patient-wise DSC of 0.873. For lesion-wise DSC, performance improved across all volume ranges. Notably, in the volume range of over 0.1 cc, all three models exhibited statistically significant improvements with $p<0.001$, while in the 0.02–0.04 cc range, the improvements were also significant with $p<0.01$. The enhancement in segmentation performance throughout post-processing was also evident in the delineations of BMs, as shown in Fig. 3. The predicted segmentation maps after post-processing closely resembled the ground-truth contours compared to the segmentation maps before processing. The post-processing, which utilized the bounding boxes from the U-Net-based approaches, did not affect LWS or the average number of FPR, but segmentation performance was improved.

DISCUSSION

The primary objective of this study was to enhance the automatic detection and segmentation of BMs in BB images. Detecting and segmenting BMs has been challenging due to their intrinsic features such as irregular shapes, small volumes, and unpredictable locations of occurrence. BB imaging suppresses the vessel elements that can facilitate BMs detection. To enhance the performance of DL-based models in terms of BMs detection and segmentation, we focused on two key aspects: modifications to the DL network architecture and post-processing using a foundation model (SAM)-based prompt.

For the network architecture, the standard U-Net was modified by incorporating a new concept of skip connection (modified U-Net) and adding a discriminator to the generator

Table 5. The Numerical Performance of Three Post-Processed Models for BMs Detection and Segmentation

After post-processing	Standard U-Net	Modified U-Net	Modified U-Net with GAN
LWS (%)	87.84	89.19	89.19
Average FPR	2.2	1.4	0.9
Patient-wise DSC	0.834 (± 0.08)* ($p=0.017$)	0.868 (± 0.05)* ($p=0.019$)	0.873 (± 0.05)* ($p=0.037$)
Tumor volume range (lesion-wise DSC)			
≥0.1 cc	0.754 (± 0.27)* ($p<0.001$)	0.746 (± 0.28)* ($p<0.001$)	0.752 (± 0.28)* ($p<0.001$)
0.06–<0.1 cc	0.775 (± 0.05)	0.788 (± 0.08)	0.787 (± 0.08)
0.04–<0.06 cc	0.576 (± 0.34)	0.667 (± 0.27)	0.718 (± 0.17)
0.02–<0.04 cc	0.638 (± 0.22)* ($p=0.002$)	0.607 (± 0.20)* ($p=0.002$)	0.630 (± 0.17)* ($p=0.003$)
<0.02 cc	0.534 (± 0.34)	0.575 (± 0.38)	0.586 (± 0.37)

BMs, brain metastases; GAN, generative adversarial network; LWS, lesion-wise sensitivity; FPR, false-positive rate; DSC, dice similarity coefficient.

Asterisks (*) indicate the statistical significance of differences between the post-processed model and the corresponding model before post-processing, as determined by p -values.

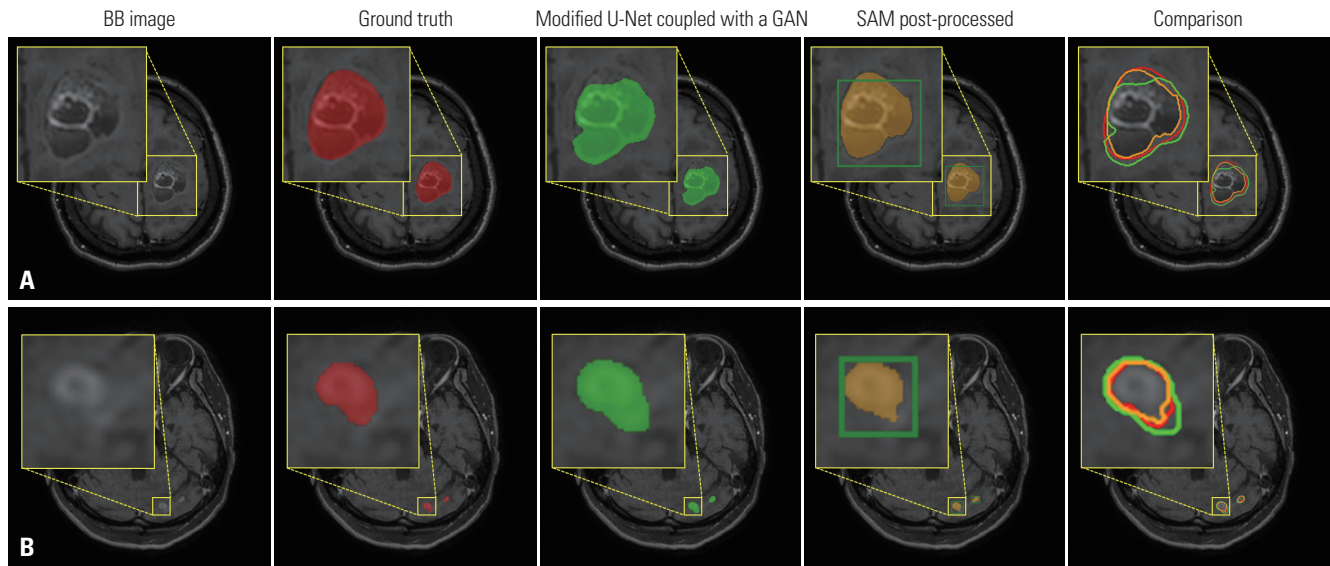


Fig. 3. The cases achieved enhancement through post-processing. The top row (A) displays under-segmentation (center) and its refinement (SAM post-processed), while the bottom row (B) shows over-segmentation (center) and its refinement (SAM post-processed). In the SAM post-processed images, the green box represents the bounding box used as input for the SAM's prompt encoder, generated by the U-Net-based model. In the comparison column, red contours denote the ground truth, green contours indicate the model's initial segmentation, and yellow contours show the post-processed segmentation, illustrating refinement in both under- and over-segmentation cases. Yellow bounding boxes indicate zoomed-in images to make differences clearer. SAM, Segment Anything Model.

(modified U-Net with GAN). The enhanced skip connections were designed to mitigate information loss during the encoding phase, allowing more effective retention of gradient information in BMs detection and segmentation. Additionally, the inclusion of a discriminator in the modified U-Net improved segmentation performance by encouraging the model to generate more realistic and accurate segmentation maps. The modified U-Net and modified U-Net with GAN models demonstrated substantial performance improvements, both achieving an LWS of 89.19%. Notably, the modified U-Net with GAN effectively minimized false positives, reducing the average FPR from 2.2 to 0.9 ($p=0.024$), demonstrating its robustness in enhancing detection reliability, as shown in Table 2.

In the post-processing stage, we used SAM to further refine the segmentation maps. While SAM alone can provide segmentation results for given medical images, it may not produce qualified segmented maps when the model is trained with different types of images. Furthermore, since SAM was designed for a broad range of segmentation tasks, applying it to our dataset posed challenges, as it segmented not only the GTV but also other structures, potentially leading to difficulty in interpreting multiple GTVs as distinct labels. Alternatively, adjusting the model to improve BMs detection and segmentation predictions could be beneficial, but fine-tuning may be challenging due to the limited dataset size used for tuning compared to the original SAM model's training data. Given these considerations, we employed SAM as a post-processing step, focusing on narrow ROIs. As shown in Tables 2 and 5, this approach contributed to improved segmentation performance, with the modified U-Net with GAN achieving a significant in-

crease in patient-wise DSC from 0.853 to 0.873. Furthermore, lesion-wise DSC significantly improved after post-processing ($p<0.001$ for over 0.1 cc, $p<0.01$ for 0.02–0.04 cc), demonstrating enhanced segmentation performance across different lesion sizes and reinforcing the robustness of our approach.

A significant commonality between our study and previous studies^{16–19} is the recognition of the potential benefits of employing BB images. Previous studies demonstrated the efficacy of utilizing BB images by achieving notable improvements in BMs detection performance. A previous study¹⁹ reported that the average FPR reached 0.59. However, the network was designed for training using both T1Gd and BB images. Additionally, the datasets used in the previous study may have differed from our datasets in terms of the proportion of patients with a small volume of BMs. Notably, our proposed network detected BMs at 100% with an average FPR of 0.67 when the number of BMs was less than four. This implies that the proposed network has the potential to be applied to patients with SRS according to the ASCO-SNO-ASTRO guidelines.⁴ In general, studies on BMs from lung cancer cases reported a median volume of 0.1 cc to 4 cc,³¹ whereas our dataset focused on tumors with a significantly smaller median volume. Specifically, our dataset included a total of 787 BMs, with approximately 20% (165 BMs) containing both necrosis and viable tumor tissue. Notably, around 80% of the BMs in this study were very small and did not contain necrosis. This aspect underscores the challenge of detecting small BMs, highlighting the effectiveness of our approach in identifying and segmenting these difficult-to-detect cases.

Despite the various advantages stated in this study, there

were certain limitations. First, the patient cohort included datasets from a single institution and consisted of only 50 patients, a relatively small sample size compared to previous studies on BMs segmentation. Although this size is modest, our study focused on enhancing the segmentation and detection accuracy of small-volume BMs in BB images by amending network architecture and employing foundation model with relatively limited sample size. The modifications of the skip connection of the U-Net and foundation model-based post-processing methods were demonstrated to effectively function in this work. Second, while convolutional neural network-based approaches have been employed for current automated BMs segmentation, recent advancements such as ViTs³²⁻³⁴ may offer potential improvements. ViTs are believed to overcome the drawbacks of convolution-based networks, which mostly rely on local image information. BMs segmentation requires a very small volume of BMs, and the degree of improvement from ViTs may be insignificant. Third, this study utilized SAM as a post-processing method, which successfully enhanced the segmentation performance, as evidenced by the increase in patient-wise DSCs. However, the post-processing method with bounding boxes from the U-Net inferences was unable to identify false-negative BMs (related to sensitivity) or suppress false-positive BMs (related to average FPR). From our observations, applying the given SAM model to our datasets did not perform well because the given model was trained with numerous types of natural images and not solely BB images. Another approach for fine-tuning the given network failed to achieve better performance than the U-Net-based approach. As a further study, expanding the dataset to include multiple institutions and using federated transfer learning could improve the generalizability of DL-based models for BMs segmentation. Additionally, developing a domain-specific SAM model trained on BB images specifically for BMs segmentation may further enhance the accuracy in identifying and segmenting BMs, addressing the limitations noted with the current SAM model.

In conclusion, the proposed work developed DL-based auto-segmentation networks to enlarge the BMs detection and segmentation performance in BB images. Modifying the skip connection in the standard U-Net notably enhanced the sensitivity of BMs in BB images, and incorporating the discriminator helped reduce false positives in BMs detection. Additionally, it was demonstrated that applying SAM as a post-processing step further improved the segmentation performance of BMs in BB images.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (NRF-2021R1G1A1092804) and Eulji University in 2021.

AUTHOR CONTRIBUTIONS

Conceptualization: Tae Hyung Kim, Jin Sung Kim, Sung Soo Ahn, Eui Hyun Kim, Wonmo Sung, and Hong In Yoon. **Data curation:** Tae Hyung Kim, Sang Kyun Yoo, and Hong In Yoon. **Formal analysis:** Sang Kyun Yoo and Tae Hyung Kim. **Funding acquisition:** Tae Hyung Kim and Hong In Yoon. **Investigation:** Tae Hyung Kim, Sang Kyun Yoo, and Hojin Kim. **Methodology:** Tae Hyung Kim, Sang Kyun Yoo, and Hojin Kim. **Project administration:** Hojin Kim and Hong In Yoon. **Resources:** Tae Hyung Kim and Hong In Yoon. **Software:** Sang Kyun Yoo and Tae Hyung Kim. **Supervision:** Hojin Kim and Hong In Yoon. **Validation:** Sang Kyun Yoo, Tae Hyung Kim, Hojin Kim, and Hong In Yoon. **Visualization:** Sang Kyun Yoo and Tae Hyung Kim. **Writing—original draft:** Sang Kyun Yoo and Tae Hyung Kim. **Writing—review & editing:** Sang Kyun Yoo, Tae Hyung Kim, Hojin Kim, and Hong In Yoon. **Approval of final manuscript:** all authors.

ORCID iDs

Sang Kyun Yoo	https://orcid.org/0000-0001-6967-0932
Tae Hyung Kim	https://orcid.org/0000-0002-5205-3775
Jin Sung Kim	https://orcid.org/0000-0003-1415-6471
Sung Soo Ahn	https://orcid.org/0000-0002-0503-5558
Eui Hyun Kim	https://orcid.org/0000-0002-2523-7122
Wonmo Sung	https://orcid.org/0000-0002-0172-0646
Hojin Kim	https://orcid.org/0000-0002-4652-8682
Hong In Yoon	https://orcid.org/0000-0002-2106-6856

REFERENCES

- Gavrilovic IT, Posner JB. Brain metastases: epidemiology and pathophysiology. *J Neurooncol* 2005;75:5-14.
- Barnholtz-Sloan JS, Sloan AE, Davis FG, Vignea FD, Lai P, Sawaya RE. Incidence proportions of brain metastases in patients diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. *J Clin Oncol* 2004;22:2865-72.
- Sung KS. Clinical practice guidelines for brain metastasis from solid tumors. *Brain Tumor Res Treat* 2024;12:14-22.
- Vogelbaum MA, Brown PD, Messersmith H, Brastianos PK, Burri S, Cahill D, et al. Treatment for brain metastases: ASCO-SNO-ASTRO guideline. *J Clin Oncol* 2022;40:492-516.
- Kaufmann TJ, Smits M, Boxerman J, Huang R, Barboriak DP, Weller M, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. *Neuro Oncol* 2020;22:757-72.
- Pope WB. Brain metastases: neuroimaging. *Handb Clin Neurol* 2018;149:89-112.
- Takeda T, Takeda A, Nagaoka T, Kunieda E, Takemasa K, Watanabe M, et al. Gadolinium-enhanced three-dimensional magnetization-prepared rapid gradient-echo (3D MP-RAGE) imaging is superior to spin-echo imaging in delineating brain metastases. *Acta Radiol* 2008;49:1167-73.
- Yoo SK, Kim TH, Chun J, Choi BS, Kim H, Yang S, et al. Deep-learning-based automatic detection and segmentation of brain metastases with small volume for stereotactic ablative radiotherapy. *Cancers (Basel)* 2022;14:2555.
- Liu Y, Stojadinovic S, Hryckushko B, Wardak Z, Lau S, Lu W, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS One* 2017;12:e0185844.
- Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G. Deep learning

- enables automatic detection and segmentation of brain metastases on multisequence MRI. *J Magn Reson Imaging* 2020;51:175-82.
11. Bousabarah K, Ruge M, Brand JS, Hoevels M, Rueß D, Borggreffe J, et al. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. *Radiat Oncol* 2020;15:87.
12. Buchner JA, Kofler F, Etzel L, Mayinger M, Christ SM, Brunner TB, et al. Development and external validation of an MRI-based neural network for brain metastasis segmentation in the AURORA multicenter study. *Radiother Oncol* 2023;178:109425.
13. Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, et al. Deep learning-based detection and segmentation-assisted management of brain metastases. *Neuro Oncol* 2020;22:505-14.
14. Park J, Kim J, Yoo E, Lee H, Chang JH, Kim EY. Detection of small metastatic brain tumors: comparison of 3D contrast-enhanced whole-brain black-blood imaging and MP-RAGE imaging. *Invest Radiol* 2012;47:136-41.
15. Lee S, Park DW, Lee JY, Lee YJ, Kim T. Improved motion-sensitized driven-equilibrium preparation for 3D turbo spin echo T1 weighted imaging after gadolinium administration for the detection of brain metastases on 3T MRI. *Br J Radiol* 2016;89:20150176.
16. Kottlors J, Geissen S, Jendreizik H, Große Hokamp N, Fervers P, Pennig L, et al. Contrast-enhanced black blood MRI sequence is superior to conventional T1 sequence in automated detection of brain metastases by convolutional neural networks. *Diagnostics (Basel)* 2021;11:1016.
17. Oh JH, Lee KM, Kim HG, Yoon JT, Kim EJ. Deep learning-based detection algorithm for brain metastases on black blood imaging. *Sci Rep* 2022;12:19503.
18. Kikuchi Y, Togao O, Kikuchi K, Momosaka D, Obara M, Van Cauteren M, et al. A deep convolutional neural network-based automatic detection of brain metastases with and without blood vessel suppression. *Eur Radiol* 2022;32:2998-3005.
19. Park YW, Jun Y, Lee Y, Han K, An C, Ahn SS, 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;31:6686-95.
20. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical normalization techniques for magnetic resonance imaging. *Neuroimage Clin* 2014;6:9-19.
21. Fortin JP, Sweeney EM, Muschelli J, Crainiceanu CM, Shinohara RT; Alzheimer's Disease Neuroimaging Initiative. Removing inter-subject technical variability in magnetic resonance imaging studies. *Neuroimage* 2016;132:198-212.
22. Reinhold JC, Dewey BE, Carass A, Prince JL. Evaluating the impact of intensity normalization on MR image synthesis. *Proc SPIE Int Soc Opt Eng* 2019;10949:109493H.
23. Nyúl LG, Udupa JK, Zhang X. New variants of a method of MRI scale standardization. *IEEE Trans Med Imaging* 2000;19:143-50.
24. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A, editors. *Medical image computing and computer-assisted intervention – MICCAI 2015*. Cham: Springer; 2015. p.234-41.
25. Seo H, Huang C, Bassenne M, Xiao R, Xing L. Modified U-Net (mU-Net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images. *IEEE Trans Med Imaging* 2020;39:1316-25.
26. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, editors. *Advances in neural information processing systems 27 (NIPS 2014)*. Red Hook (NY): Curran Associates; 2015.
27. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. *arXiv [Preprint]*. 2020 [accessed on 2024 April 1]. Available at: <https://doi.org/10.48550/arXiv.2010.11929>.
28. He K, Chen X, Xie S, Li Y, Dollár P, Girshick R. Masked autoencoders are scalable vision learners. *arXiv [Preprint]*. 2021 [accessed on 2024 April 1]. Available at: <https://doi.org/10.48550/arXiv.2111.06377>.
29. Paszke A, Gross S, Chintala S, Chanan G, Yang E, DeVito Z, et al. Automatic differentiation in PyTorch. *OpenReview*. 2017 [accessed on 2024 April 1]. Available at: <https://openreview.net/forum?id=BJJsrmlfCZ>.
30. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;12:2825-30.
31. Roh YH, Park JE, Kang S, Yoon S, Kim SW, Kim HS. Prognostic value of MRI volumetric parameters in non-small cell lung cancer patients after immune checkpoint inhibitor therapy: comparison with response assessment criteria. *Cancer Imaging* 2023;23:102.
32. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al, editors. *Advances in neural information processing systems 30 (NIPS 2017)*. Red Hook (NY): Curran Associates; 2018.
33. Shaw P, Uszkoreit J, Vaswani A. Self-attention with relative position representations. *arXiv [Preprint]*. 2018 [accessed on 2024 April 1]. Available at: <https://doi.org/10.48550/arXiv.1803.02155>.
34. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: a survey. *Med Image Anal* 2023;88:102802.