# Artificial-intelligence-driven measurements of brain metastases' response to SRS compare favorably with current manual standards of assessment

**Kayla Prezelski**, **Dylan G. Hsu, Luke del Balzo, Erica Heller, Jennifer Ma, Luke R.G. Pike, Åse Ballangrud, and Michalis Aristophanous**

All author affiliations are listed at the end of the article

Corresponding Author: Dylan G. Hsu, PhD, Department of Medical Physics, Memorial Sloan Kettering Cancer Center, 1275 York Ave., New York, NY 10065, USA (Hsud3@mskcc.org).

## Abstract

**Background**. Evaluation of treatment response for brain metastases (BMs) following stereotactic radiosurgery (SRS) becomes complex as the number of treated BMs increases. This study uses artificial intelligence (AI) to track BMs after SRS and validates its output compared with manual measurements.

**Methods**. Patients with BMs who received at least one course of SRS and followed up with MRI scans were retrospectively identified. A tool for automated detection, segmentation, and tracking of intracranial metastases on longitudinal imaging, MEtastasis Tracking with Repeated Observations (METRO), was applied to the dataset. The longest three-dimensional (3D) diameter identified with METRO was compared with manual measurements of maximum axial BM diameter, and their correlation was analyzed. Change in size of the measured BM identified with METRO after SRS treatment was used to classify BMs as responding, or not responding, to treatment, and its accuracy was determined relative to manual measurements.

**Results**. From 71 patients, 176 BMs were identified and measured with METRO and manual methods. Based on a one-to-one correlation analysis, the correlation coefficient was $R^2 = 0.76$ ($P = .0001$). Using modified BM response classifications of BM change in size, the longest 3D diameter data identified with METRO had a sensitivity of 0.72 and a specificity of 0.95 in identifying lesions that responded to SRS, when using manual axial diameter measurements as the ground truth.

**Conclusions**. Using AI to automatically measure and track BM volumes following SRS treatment, this study showed a strong correlation between AI-driven measurements and the current clinically used method: manual axial diameter measurements.

## Key Points

- AI-driven measurements of brain metastases correlate to manual measurements.
- AI can alleviate labor-intensive processes in BM radiation treatment planning.
- AI provides longitudinal data to supplement clinical decision-making.

Brain metastases (BMs) are the most common central nervous system tumor,[1] and they are diagnosed in 170 000 patients in the United States annually.[2] Historically, cancer patients with metastases to the brain were frequently treated with palliative whole-brain radiation therapy with a median overall survival (OS) of only 4–7 months.[2] Stereotactic radiosurgery (SRS) was labor-intensive and performed with an invasive frame,[3] and it was therefore limited to select patients with small numbers of BMs. In recent years, advancements in delivery technologies[4] and an increase in evidence that using SRS to spare normal brain tissues results in improved quality of life, cognitive outcomes,[5] and local tumor control[6] resulted in a shift

## Importance of the Study

In this study, we aimed to determine how BM measurements obtained by AI compared with the manually collected data on a one-to-one basis as well as to investigate BM SRS response classification performance. We found a strong correlation between the manual and AI one-to-one measurements. The advantages of METRO over manual measurement, as found in this study, include its longest three-dimensional (3D) diameter measurement more accurately capturing 3D BMs that deviate from classical spherical shape, and its consistent reproducibility in contrast to manual measurements that are subject to human error. AI-driven automated BM response classification was accurate and therefore can assist in streamlining clinical course monitoring and decision-making for patients with BMs.

in radiotherapy (RT) protocols for BM management away from whole-brain radiation therapy towards targeting individual lesions with SRS.[7] In addition, advances in systemic therapy have led to improved extracranial disease control and an increase in OS in these patients.[8–12] The combination of prolonged survival with a local management approach for BMs means that patients may receive multiple courses of SRS after locoregional failure. Given the multiplicity of lesions treated at multiple time points, consistent and accurate monitoring of individual BMs is now crucial.

Evaluation of treatment response for each BM and planning of new SRS treatments become more complex as the number of BMs increases. Patients receive multiple MRI and CT scans for each treatment course and a follow-up MRI every 2–3 months following treatment for surveillance. Current commercial software does not provide an automated solution for tracking BMs, and the recent PACS-integrated longitudinal tracking tool has highlighted the desire to simplify this laborious process.[13] Manual methods are time-intensive and fraught with potential errors if sequential images are not co-registered and radiation plan information is not overlayed on the images. It is neither practical nor feasible to detail each metastasis in the official radiological report. Even when recorded, small and neighboring metastases in the same brain lobe, or gyrus, may be mistaken for one another across follow-up surveillance imaging.

The current method for evaluating BMs and their response to treatment is manual evaluation. BM size has been reported based on a measurement of the longest axial diameter.[14] According to the results of a study by Benson et al., given the complexities of the manual process, radiologists often will only comment directly on a few dominant, or clinically significant, BMs and make general assessments of the total number of BMs in the brain.[15] Bi et al. contend that artificial intelligence (AI) can be used to automate this laborious process and open big data analytics.[16] Although clinical decision-making will still rely on multiple factors, including patients' clinical presentation, performance status, and information from multiple imaging techniques (ie, perfusion, κ-trans, PET),[17] AI can provide quantitative longitudinal data. This can help augment the current standard of care and drive escalation/de-escalation of treatment and surveillance imaging frequency.

This validation study uses in-house developed AI software[18,19] to determine how BM detection, auto-segmentation, and longitudinal tracking compares with the current clinically used method of manual BM measurement. BM measurements obtained by AI are compared to the manually collected data on a one-to-one basis. BM response to SRS is derived from the Response Assessment in Neuro-Oncology working group for Brain Metastases (RANO-BM) criteria, which has been modified for this study.[14]

## Materials and Methods

### Data Collection

Under an institutional review board approved protocol, RT-naïve patients with renal cell carcinoma (RCC) BMs who received cranial single- or hypo-fractionated SRS were retrospectively identified. RCC has relatively good OS and response to SRS, allowing the analysis of longitudinal BM response in this study. Patient demographic data was recorded.

A tool developed in-house for automated detection and segmentation of intracranial metastases on longitudinal images, MEtastasis Tracking with Repeated Observations (METRO),[18,19] was applied to the dataset. METRO utilizes a three-dimensional (3D) convolutional neural network to segment gross tumor volumes on pre- and post-treatment follow-up T1-weighted (T1w) MR images and automated rigid registration to identify and track each lesion on longitudinal scans. The software then calculates the BM volume and longest 3D diameter on all available images following SRS based on the diameter of Gd contrast enhancement. The longest 3D diameter is defined as the diameter of the sphere with the minimum volume needed to encompass the lesion in three dimensions. The METRO software was used to longitudinally track 176 BMs treated within this patient group on a per-lesion basis.

For the manual data collection, images were visualized and measured using Centricity Universal Viewer v.7.0. Manual measurement was conducted on the same 176 BMs by measuring the maximum axial diameter of the BM using T1w brain volume imaging (BRAVO) post-Gd-contrast sequences. Axial T1w post-contrast images were used when BRAVO post-Gd-contrast sequences were lacking.[20] Manual measurements were conducted by authors L.D.B. and J.M. and were reviewed by senior author L.R.G.P.

### METRO to Manual Diameter Comparison

A one-to-one comparison of the METRO longest 3D diameter vs. the manual expert-drawn axial diameter was made for each BM at all follow-up imaging time points. The

correlation between the longest 3D diameter identified by METRO and the manual axial diameter was determined by basic linear regression. For BMs that had a 100% discrepancy, ie, they were missed by either METRO or manual measurement methods, the median BM diameter between the two methods was compared to gain insight regarding the size of the missed lesions using the Mann–Whitney test. All resection cavities were excluded from this analysis because METRO was not trained on resection cavity data.[19]

## BM SRS Response Classification

To investigate clinically relevant outcome measurements, BM change in size after SRS treatment was investigated. Similar to previous work,[18] a time window of 6 ± 3 months of follow-up time (90–270 days) was chosen. This time window, although not long enough to radiologically differentiate between treatment effect vs. BM recurrence, was chosen to investigate BM SRS response while balancing the limiting factor of patient OS. The pretreatment longest 3D diameter for each BM, determined by METRO, was compared to the longest 3D diameter measured on the 6-month post-SRS follow-up MRI. Then, the percent change in the longest 3D diameter of each BM between pre-SRS and 6-month post-SRS follow-up was calculated. Response categories derived from criteria proposed by RANO-BM[14] were modified in accordance with this study and applied to categorize each BM's response to SRS. BMs that decreased in diameter by at least 80% or were found to have resolved entirely on follow-up imaging were classified as unappreciable. A 30% or greater decrease in diameter was classified as "decreasing," a 20% or greater increase in size was "increasing," and the remainder were "stable." Because this study is meant to comment on the volume of change in each BM, rather than the clinical disease status, the four classifications used are moderately analogous to RANO-BM classification categories of complete progressive disease, stable disease, partial response, and complete response.[14]

The response category classification from METRO was compared with the classification based on manual axial diameter measurements for the same lesions on the corresponding pretreatment and nearest 6-month follow-up MR images in the picture archiving and communication system (PACS, GE Healthcare). Using manual axial diameter measurement as the ground truth (the current clinically used method), the precision, recall, specificity, and F1-score of response assessment categories determined by METRO were calculated. Therefore, a true positive indicated that METRO correctly categorized a BM in line with manual measurements. Precision is the ratio of the correctly categorized BMs to all BMs that METRO placed in that category. Recall, or sensitivity, is the ratio of correctly categorized BMs to all BMs that were placed in that category by METRO, as determined by manual axial diameter measurement as the ground truth. Specificity is the ratio of the negatively categorized BMs by METRO to all BMs that were genuinely negative, as determined by manual axial diameter measurement as the ground truth. The F1 score is the average of precision and recall.

When considering the decision to re-treat a BM, the change in its size since the last SRS is of interest, and in addition to clinical presentation, can trigger the need for closer follow-up and acquisition of other image modalities (ie, PET, perfusion, delayed contrast) to verify active tumor tissue. Within a clinical context, automatic measurement and flagging of BMs increasing in volume is important because further images and evaluation are needed for such cases. Therefore, to best capture clinically relevant quantitative data, a modified BM SRS response classification system was applied to the dataset, which classified treated BMs as either responding or not responding to SRS. The "responding" category combines the BM SRS response classification criteria (unappreciable, decreasing, and stable categories); it was defined as a diameter percent change of <20% from pre-SRS measurement to approximately 6 months follow-up. The "not responding" category is comparable to the "increasing" or "progressive disease" category of RANO-BM in that the lesions have continued to increase since receiving SRS treatment and were defined as a diameter percent change of ≥20%. The AI-measured change in volume was also compared with the manual measurements described earlier.

## Results

There were 71 RT-naïve patients with RCC BMs who received cranial single- or hypo-fractionated SRS retrospectively identified. The cohort was 49:22 male to female, with a median age of 57 years (range 22–77) at primary tumor diagnosis, and median age of 62 years (range 24–81) at BM diagnosis. There were 176 BMs in the study, with a median of 1 BM per patient (IQR: 1–2, range: 1–19). Patients had a median of 1 (IQR: 1–2, range: 1–7) SRS courses.

Among the 71 patients, and a total of 176 unique BMs, there were 629 unique instances (ie, pre- and post-SRS MRI) in which BM measurements identified by both METRO and manual measurements were completed on the same patient MRI. Resection cavities ($n = 23$) were then excluded from further comparison. From the 606 remaining instances, 79 measurements were analyzed separately due to a 100% discrepancy between METRO and manual measurements. Of the 79 differences identified, 45 BMs were missed by METRO (ie, METRO indicated BM volume equal to zero whereas manual methods provided a non-zero volumetric measurement), and 34 BMs were missed by manual measurement methods (ie, METRO provided non-zero volumetric measurement whereas manual methods indicated BM volume equal to zero). Of the BMs missed by METRO ($n = 45$), the median BM manual diameter was 2.8 mm (IQR, 1.8–5). Of the BMs missed by manual measurement ($n = 34$), the median BM longest 3D diameter was 6.6 mm (IQR, 5.4–10.2) (Figure 1). BMs that were missed by manual measurement but contoured by METRO were confirmed to be true BMs by authors L.D.B. and L.R.G.P. as well as radiology reports where appropriate. The Mann-Whitney test comparison resulted in a statistically significant mean rank between these two groups ($P < .0001$), indicating that the BMs missed by METRO were smaller than those missed by manual methods.

Five hundred twenty-seven instances for comparison remained, representing 70 patients with 175 unique BMs. In

the one-to-one comparison of 527 instances between the longest 3D diameter determined by METRO, and the manually measured BM axial diameter, the correlation coefficient was $R^2 = 0.76$ ($P = .0001$) (Figure 2).

Eighty-four of the 176 BMs had at least one follow-up MRI within the specified time window. For patients with multiple follow-up scans within this time window, the scan closest to 6 months (180 days) was used, as determined by the minimal difference between the SRS start date and all follow-up MRIs within the time window. The median time between the SRS start date and the 6-month follow-up scan was 134 days (IQR, 107–162). Ninety-two BMs were excluded from the BM SRS response classification due to a lack of follow-up imaging in the time window, poor image quality, or other clinical factors. Among the 84 BMs that METRO tracked in the specified follow-up window of 6 ± 3 months, the calculated longest 3D diameter agreed with the manually measured axial

diameter classifications for 47 BMs (Table 1), resulting in an overall sensitivity and specificity of 0.72 and 0.85, respectively (Table 1). Figure 3A shows the classification differences between METRO and the manual classification; the manual vs. the METRO-defined percent change with the BM SRS response classification percentage cutoffs are shaded.

Using modified BM SRS response criteria defined in this study, METRO agreed with manual axial diameter-based classification for 13 out of 18 "not responding," ie, "increasing" BMs (Table 1), when referencing the change in manual axial diameter as the ground truth, and 63 out of 66 correctly classified as "responding." Five false negatives (FN1–FN5), ie, BMs that manual measurement-based classification determined to not respond to SRS, but METRO classified as responding to SRS, are labeled in Figure 3B.

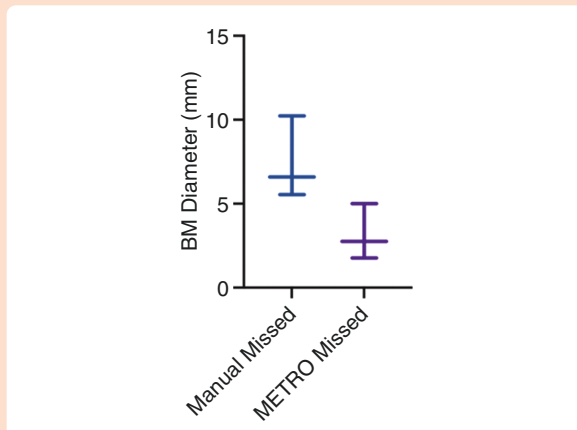Select cases from METRO are displayed in Figures 4 and 5.



**Figure 1.** Brain metastases (BMs) missed by MEtastasis Tracking with Repeated Observations (METRO) ($n = 45$) had a median manually measured axial diameter of 2.8 mm (IQR, 1.8–5). BMs missed by manual methods ($n = 34$) had a median BM longest 3D diameter of 6.6 mm (IQR, 5.4–10.2).

## Discussion

The focus of this study was to determine how BM measurements obtained by AI compared with the manually collected data on a one-to-one basis as well as to investigate BM SRS response classification performance. A strong correlation was found between the manual and AI one-to-one measurements. As expected, the longest 3D diameter of a BM measured by METRO was found to be longer than the manual measurement of the longest 3D diameter on the axial plane, as shown in Figure 2. Because they are inherently different measurements, a perfect one-to-one correlation would not be expected. Although standardized across the field, the manual axial diameter measurement may not represent the full picture. For example, when lesions deviate from a spherical shape and expand in the coronal or sagittal planes, ie, a cylindrical BM (Figure 4), the longest 3D diameter can more accurately capture the maximum diameter of the BM. Accordingly, many of the points falling below
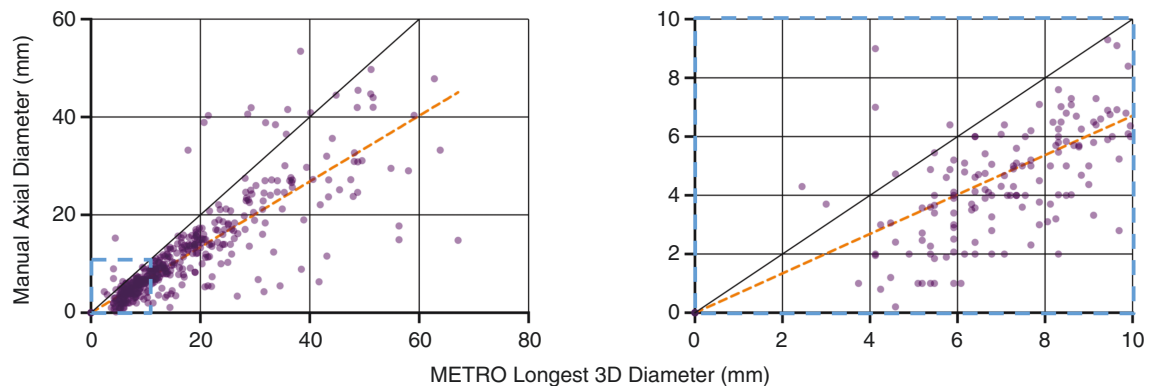


**Figure 2.** MEtastasis Tracking with Repeated Observations (METRO) longest 3D diameter calculated from the brain metastases (BMs) volume vs. manual axial diameter for $n = 527$ BMs. $R^2 = 0.76$.

**Table 1.** Confusion matrix to determine METRO's BM classification performance. Bold values indicate agreement between METRO and manual determination

| METRO predicted classification | Ground truth (manual axial diameter) | | | |
|---|---|---|---|---|
| | Increasing | Stable | Decreasing | Unappreciable |
| Increasing | **13** | 2 | 1 | 0 |
| Stable | 3 | **6** | 7 | 3 |
| Decreasing | 1 | 2 | **5** | 6 |
| Unappreciable | 1 | 1 | 10 | **23** |
| | Not responding | Responding | | |
| Not responding | **13** | 3 | | |
| Responding | 5 | **63** | | |
| **BM classification** | Precision | Recall | Specificity | F1-score |
| Increasing | 0.81 | 0.72 | 0.95 | 0.76 |
| Stable | 0.32 | 0.55 | 0.82 | 0.40 |
| Decreasing | 0.36 | 0.22 | 0.85 | 0.27 |
| Unappreciable | 0.66 | 0.72 | 0.77 | 0.69 |
| Not responding | 0.81 | 0.72 | 0.95 | 0.76 |

the line of equality are shown in Figure 2, indicating that the longest 3D diameter measured by METRO is, as expected, consistently larger than the corresponding manual axial diameter measurements.

In addition, for lesions with irregular shapes, it can be difficult for an observer to manually measure the maximal diameter and identify the true maximal extent of a lesion. On the other hand, an automated method can easily identify the true full extent of the disease through a geometrical calculation (Figure 5).
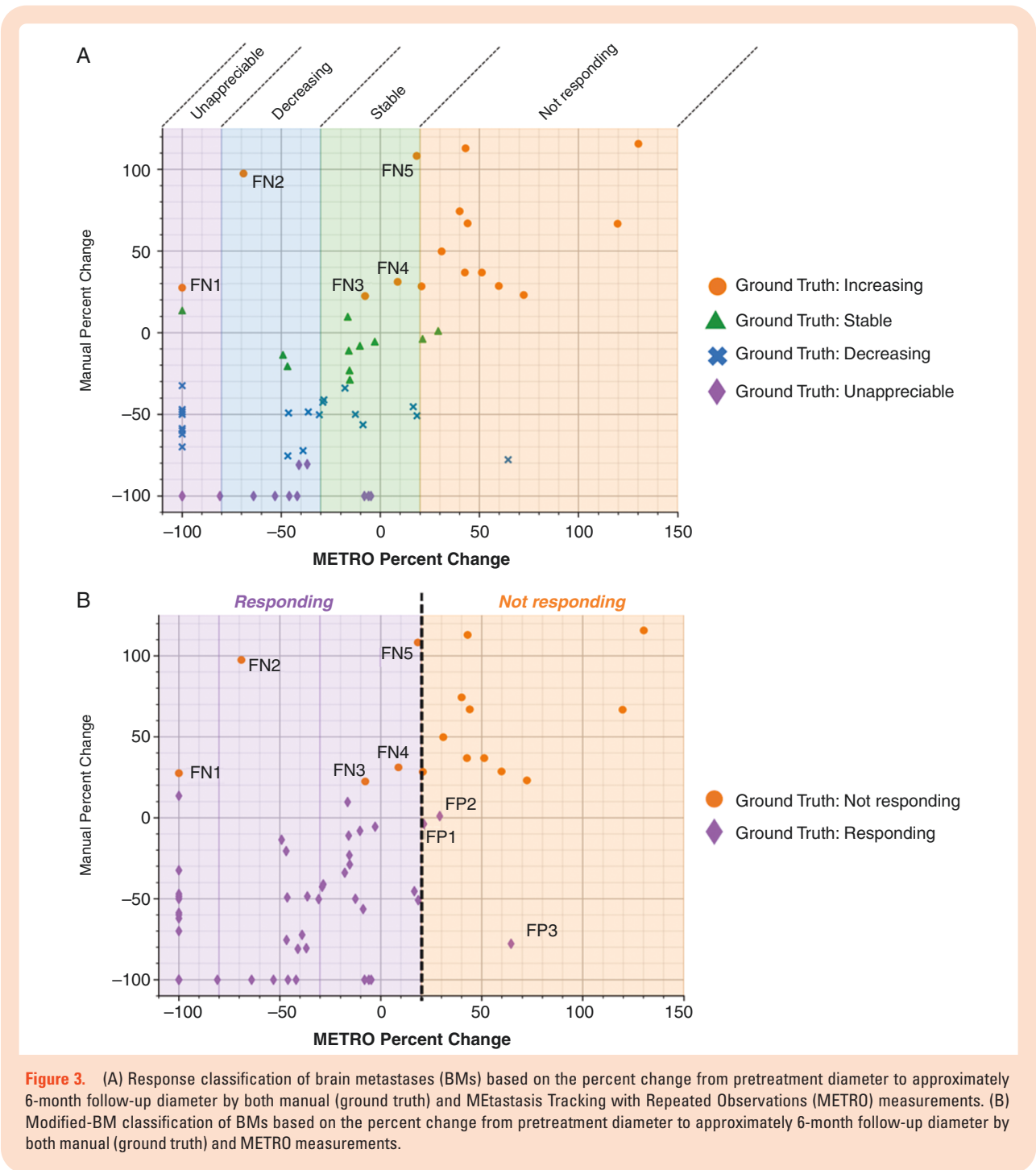
Lastly, AI software, such as METRO, offers 3D volumetric information about a BM. This can provide a more complete picture of how the BM is responding rather than a simple diameter estimate. In addition to volume and diameter, many other metrics indicative of size and shape can be easily obtained by METRO, such as elongation and sphericity.

METRO AI suffers from the limitations of the detection and segmentation algorithm that it utilizes for lesion identification. In a previous study, our group described and characterized the 3D CNN algorithm utilized in METRO for BM identification and obtaining the volume measurement.[19] The results of the study demonstrated that the algorithm has near 100% detection sensitivity for larger lesions, which drops under 90% for lesions smaller than 5 mm, and METRO follows that trend, as shown in Figure 1. BMs that fell below the detection sensitivity of METRO were missed ($n = 45$). However, the manual detection also missed BM measurements ($n = 34$), and the manual misses were in a less predictable manner because they included lesions of all sizes, as shown in Figure 1. The known and well-characterized performance of an AI algorithm can be seen as advantageous compared with the more randomized human error seen in the detection by manual human measurements. The 34 BMs missed by manual measurements occurred on 20 unique BMs over consecutive follow-up MRIs. This is likely partially attributed to the scenario in which a BM shows no contrast enhancement

in one time point and then shows enhancement on a later follow-up MRI. The recurrence may not be found due to human error because they stopped tracking once the enhancement was gone.

METRO displayed good precision, recall, and specificity in identifying increasing and unappreciable BMs in the BM SRS response classification (Table 1). In terms of the intermediate categories, ie, stable and decreasing, which have a narrow percent change range, performance was reduced, as shown in Figure 3A. Specificity remained high across all 4 categories (Table 1). For most BMs studied, which were <20 mm in diameter, an approximately 2 mm difference in the diameter measurement can make a 10–20% difference in percent change that would cause a BM to become reclassified, especially among categories that only span a narrow percentage range. In addition, many of the disagreements between METRO and the manual measurement classification in the decreasing category resulted from small lesions (<5 mm) not being detected by METRO and therefore incorrectly classified as unappreciable. When extrapolating such misclassifications to the clinical management of patients, a >50% reduction in the size of a small BM vs. a 100% complete disappearance would be considered a success in SRS and would not require further surveillance or treatment.

Identifying and tracking the increase in BM volume after treatment will have a significant impact on the patient's continuous care. BM size increase will prompt closer follow-up, potentially with different image modalities, or additional MRI sequences, to establish whether the increase in volume is due to active tumor growth, treatment effect, or radiation necrosis. To investigate the performance of METRO for BMs that continue to increase after RT, a clinically based binary decision was applied to these metrics. When combining the three BM SRS response classification categories of stable, decreasing, and unappreciable to a binary clinical decision of not responding or responding (Table 1), METRO had a similar

**Figure 3.**   (A) Response classification of brain metastases (BMs) based on the percent change from pretreatment diameter to approximately 6-month follow-up diameter by both manual (ground truth) and MEtastasis Tracking with Repeated Observations (METRO) measurements. (B) Modified-BM classification of BMs based on the percent change from pretreatment diameter to approximately 6-month follow-up diameter by both manual (ground truth) and METRO measurements.

performance to the manual observer. A specificity of 0.95, ie, a reliable classification of a BM as responding to SRS or not increasing in size, can be helpful in clinical decision-making, especially in complex patient cases where there are multiple new and old lesions to consider. In turn, this would allow the clinician to focus on BMs of interest in an expedited manner, ultimately allowing for improved treatment management. The sensitivity of 0.72 is not sufficiently high to confirm that a BM is not increasing in size. The metric of classification

performance is impacted by the difference between the longest 3D diameter determined by METRO and the corresponding manual axial diameter measurements, which further highlights the need to correlate BMs with additional clinical imaging studies.

BMs that increased in size, but were incorrectly classified as responding by METRO, were further investigated (*n* = 5). False negative 1 (Figure 3B) was likely misclassified due to a lack of contrast uptake by the BM, which is relied upon by METRO to properly segment it from the
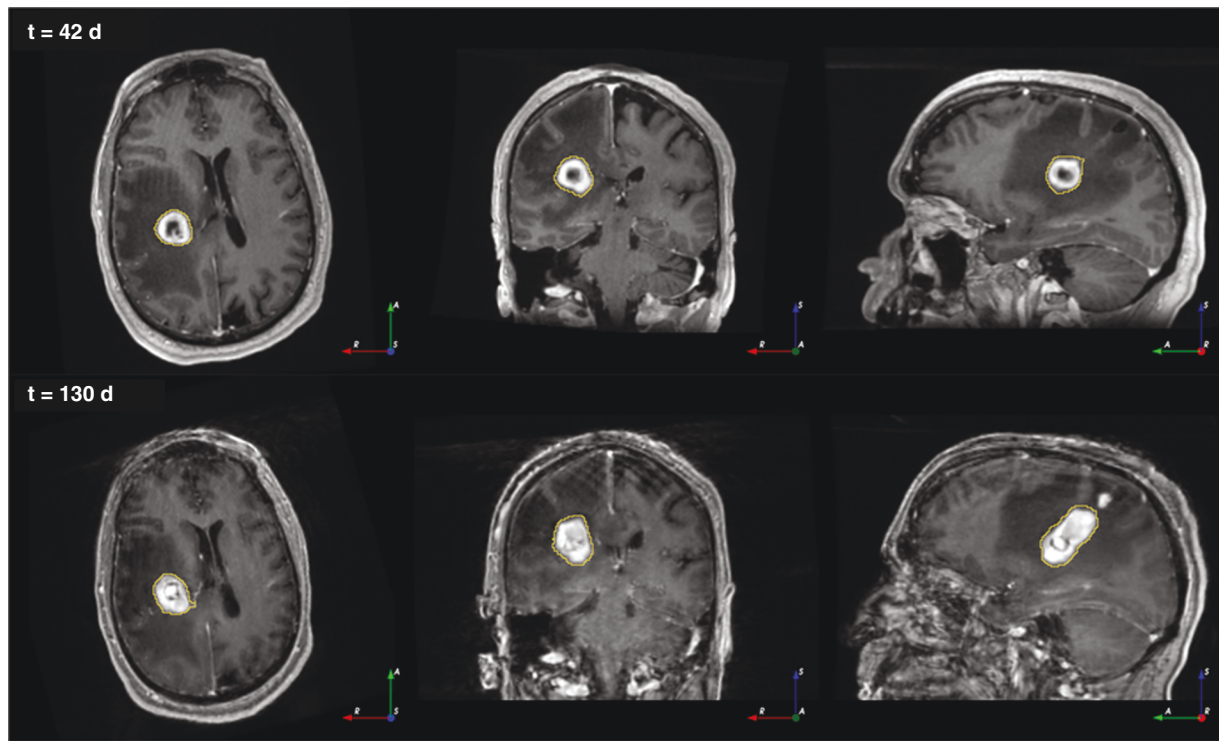
**Figure 4.** Brain metastases measurements by manual methods and MEtastasis Tracking with Repeated Observations (METRO) at two time points, t = 42 days and t = 130 days, showing an increased discrepancy between measurements as the lesion becomes cylindrical rather than spherical at a later time point.
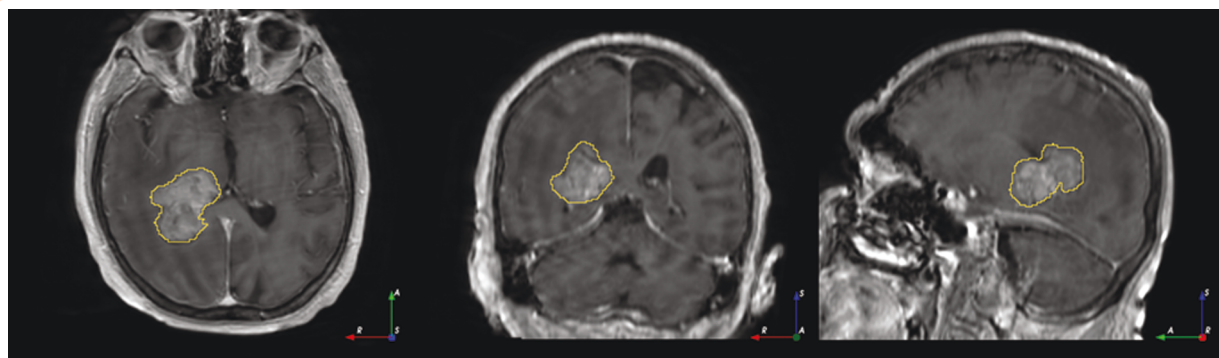


**Figure 5.** An irregularly shaped brain metastasis segmented by the MEtastasis Tracking with Repeated Observations (METRO) process is shown.

surrounding brain tissue. False negative 2 (Figure 3B) was likely misclassified resulting from MRI limitations due to patient motion artifact. False negatives 3, 4, and 5 (FN3, FN4, and FN5, Figure 3B) were likely misclassified due to the difference between calculating the 3D longest diameter (METRO) vs. the manual axial diameter. For such BMs, an increase in diameter was measured manually on the axial plane whereas they remained stable on the plane of the longest 3D diameter. For these cases, BM volume from METRO, where the lesion is evaluated as a 3D structure,

would likely be the most accurate way to capture the BM change in size.

Limitations of MRI technology, such as patient motion artifact and resolution, will continue to affect the accuracy of AI-driven detection and segmentation of BMs in terms of maximizing sensitivity and decreasing false positive detections.[19,21–26] The high specificity of METRO in classifying BMs that were responding to treatment will allow physicians to quickly focus on BMs of interest for potential re-treatment and spend less time manually analyzing

BMs that have a confirmed reduction in volume. The AI-driven segmentation of BMs from METRO provided lesion volumes. In current practice, such segmentation would require significant manual effort and a lot of time to accomplish by clinicians; however, AI-driven segmentation makes this data feasible to obtain in a relatively quick timeframe with minimal user interaction. Although no current classification system uses BM volume, this method will provide more accurate information for non-spherical BMs as well as information on a patient's overall intracranial tumor burden[27,28] and therefore can assist clinicians when following their patients post-treatment.

## Conclusions

In this study, manual axial diameter measurements of BMs were compared with measurements created by METRO, an AI-based tool for automated detection and segmentation of intracranial metastases. Although METRO measures BMs based on their longest 3D diameter, the measurement strongly correlates to that of the currently used clinical method. The known and well-characterized limitations of detection sensitivity in METRO may have an advantage over that of random human error when considering missed lesions, especially when considering the time savings for clinicians who would not have to obtain BM size estimates. Furthermore, METRO performs well with detecting and measuring large BMs, which have the highest impact on patient quality of life and OS. With further refinement of training datasets for METRO, its ability to aid clinical decision-making will be enhanced. Because AI tools, such as METRO, can be used to acquire large volumes of data over long treatment courses for patients, applications of big data analytics will be of value to study BM growth curves and how they correlate to clinical characteristics that can refine longitudinal BM management and patient care.

## Keywords

brain metastases | deep learning | imaging | longitudinal tumor tracking | radiation therapy

## Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Authorship statement

Conceptualization: K.P., D.G.H, A.B., M.A.; manual data collection: L.D.B., J.M., L.R.G.P.; METRO data collection: K.P., D.G.H.; formal analysis: K.P., E.H.; METRO Software: D.G.H, A.B., M.A.; writing: original draft: K.P.; writing: review and editing: K.P., D.G.H., L.D.B., L.R.G.P, A.B., M.A.
Data Availability
Data will be made available upon reasonable request.

## Affiliations

Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, New York, USA (K.P., D.G.S., E.H., Å.B, M.A.); Saint Louis University School of Medicine, St. Louis, Missouri, USA (K.P.); Medical College of Georgia, Athens, Georgia, USA (L.D.B.); Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, USA (L.D.B., J.M., L.R.G.P.); Biomarker Development Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA (L.R.G.P.)

## References

1. Tabouret E, Chinot O, Metellus P, et al. Recent trends in epidemiology of brain metastases: An overview. *Anticancer Res.* 2012;32(11):4655–4662.
2. Ellis TL, Neal MT, Chan MD. The role of surgery, radiosurgery and whole brain radiation therapy in the management of patients with metastatic brain tumors. *Int J Surg Oncol.* 2012;2012:1–10.
3. Lightstone AW, Benedict SH, Bova FJ, Solberg TD, Stern RL. American Association of Physicists in Medicine Radiation Therapy Committee. Intracranial stereotactic positioning systems: Report of the American Association of Physicists in Medicine Radiation Therapy Committee Task Group No. 68. *Med Phys.* 2005;32(7Part1):2380–2398.
4. Lee SK, Huang S, Zhang L, et al. Accuracy of surface-guided patient setup for conventional radiotherapy of brain and nasopharynx cancer. *J Appl Clin Med Phys.* 2021;22(5):48–57.
5. Brown PD, Jaeckle K, Ballman KV, et al. Effect of radiosurgery alone vs radiosurgery with whole brain radiation therapy on cognitive function in patients with 1 to 3 brain metastases: A randomized clinical trial. *JAMA.* 2016;316(4):401–409.
6. El Gantery MM, Abd El Baky HM, El Hossieny HA, Mahmoud M, Youssef O. Management of brain metastases with stereotactic radiosurgery alone versus whole brain irradiation alone versus both. *Radiat Oncol.* 2014;9(1):116.

7. Gondi V, Bauman G, Bradfield L, et al. Radiation therapy for brain metastases: An ASTRO clinical practice guideline. *Pract Radiat Oncol.* 2022;12(4):265–282.

8. Bailleux C, Eberst L, Bachelot T. Treatment strategies for breast cancer brain metastases. *Br J Cancer.* 2021;124(1):142–155.

9. Venur VA, Leone JP. Targeted therapies for brain metastases from breast cancer. *Int J Mol Sci.* 2016;17(9):1543.

10. Lin NU. Intracranial efficacy and survival with tucatinib plus trastuzumab and capecitabine for previously treated HER2-positive breast cancer with brain metastases in the HER2CLIMB Trial. *J Clin Oncol.* 2020;38(23):2610–2619.

11. Curigliano G, Mueller V, Borges V, et al. Tucatinib versus placebo added to trastuzumab and capecitabine for patients with pretreated HER2+ metastatic breast cancer with and without brain metastases (HER2CLIMB): Final overall survival analysis. *Ann Oncol.* 2022;33(3):321–329.

12. Murthy RK, Loi S, Okines A, et al. Tucatinib, trastuzumab, and capecitabine for HER2-positive metastatic breast cancer. *N Engl J Med.* 2020;382(7):597–609.

13. Cassinelli Petersen G, Bousabarah K, Verma T, et al. Real-time PACS-integrated longitudinal brain metastasis tracking tool provides comprehensive assessment of treatment response to radiosurgery. *Neurooncol Adv.* 2022;4(1):vdac116.

14. Lin NU, Lee EQ, Aoyama H, et al. Response Assessment in Neuro-Oncology (RANO) group. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol.* 2015;16(6):e270–e278.

15. Benson J, Burgstahler M, Zhang L, Rischall M. The value of structured radiology reports to categorize intracranial metastases following radiation therapy. *Neuroradiol J.* 2019;32(4):267–272.

16. Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: Clinical challenges and applications. *CA Cancer J Clin.* 2019;69(2):127–157.

17. Mehrabian H, Detsky J, Soliman H, Sahgal A, Stanisz GJ. Advanced magnetic resonance imaging techniques in management of brain metastases. *Front Oncol.* 2019;9(1):440.

18. Hsu DG, Ballangrud A, Prezelski K, et al. Automatically tracking brain metastases after stereotactic radiosurgery. *Phys Imag Radiat Oncol.* 2023;27:100452.

19. Hsu DG, Ballangrud A, Shamseddine A, et al. Automatic segmentation of brain metastases using T1 magnetic resonance and computed tomography images. *Phys Med Biol.* 2021;66(17). doi:10.1088/1361-6560/ac1835

20. Kaufmann TJ, Smits M, Boxerman J, et al. Consensus recommendations for a standardized brain tumor imaging protocol for clinical trials in brain metastases. *Neuro Oncol.* 2020;22(6):757–772.

21. 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. doi:10.1016/j.compbiomed.2018.02.004

22. 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.

23. Rudie JD, Weiss DA, Colby JB, et al. Three-dimensional U-Net convolutional neural network for detection and segmentation of intracranial metastases. *Radiol Artif Intell.* 2021;3(3):e200204.

24. Liu Y, Stojadinovic S, Hrycushko B, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. *PLoS One.* 2017;12(10):e0185844.

25. 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.

26. 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.

27. Kim IK, Starke RM, McRae DA, et al. Cumulative volumetric analysis as a key criterion for the treatment of brain metastases. *J Clin Neurosci.* 2017;39:142–146.

28. Hirshman BR, Wilson B, Ali MA, et al. Superior prognostic value of cumulative intracranial tumor volume relative to largest intracranial tumor volume for stereotactic radiosurgery-treated brain metastasis patients. *Neurosurgery.* 2018;82(4):473–480.