RESEARCH

Predicting local control of brain metastases after stereotactic radiotherapy with clinical, radiomics and deep learning features

Hemalatha Kanakarajan^{1*}, Wouter De Baene^{1*}, Patrick Hanssens^{2,3} and Margriet Sitskoorn¹

Abstract

Background and purpose Timely identification of local failure after stereotactic radiotherapy for brain metastases allows for treatment modifications, potentially improving outcomes. While previous studies showed that adding radiomics or Deep Learning (DL) features to clinical features increased Local Control (LC) prediction accuracy, their combined potential to predict LC remains unexplored. We examined whether a model using a combination of radiomics, DL and clinical features achieves better accuracy than models using only a subset of these features.

Materials and methods We collected pre-treatment brain MRIs (TR/TE: 25/1.86 ms, FOV: 210×210×150, flip angle: 30°, transverse slice orientation, voxel size: 0.82×0.82×1.5 mm) and clinical data for 129 patients at the Gamma Knife Center of the Elisabeth-TweeSteden Hospital. Radiomics features were extracted using the Python radiomics feature extractor and DL features were obtained using a 3D ResNet model. A Random Forest machine learning algorithm was employed to train four models using: (1) clinical features only; (2) clinical and radiomics features; (3) clinical and DL features; and (4) clinical, radiomics, and DL features. The average accuracy and other metrics were derived using K-fold cross validation.

Results The prediction model utilizing only clinical variables provided an Area Under the receiver operating characteristic Curve (AUC) of 0.85 and an accuracy of 75.0%. Adding radiomics features increased the AUC to 0.86 and accuracy to 79.33%, while adding DL features resulted in an AUC of 0.82 and accuracy of 78.0%. The best performance came from combining clinical, radiomics, and DL features, achieving an AUC of 0.88 and accuracy of 81.66%. This model's prediction improvement was statistically significant compared to models trained with clinical features alone or with the combination of clinical and DL features. However, the improvement was not statistically significant when compared to the model trained with clinical and radiomics features.

Conclusion Integrating radiomics and DL features with clinical characteristics improves prediction of local control after stereotactic radiotherapy for brain metastases. Models incorporating radiomics features consistently outperformed those utilizing clinical features alone or clinical and DL features. The increased prediction accuracy of our integrated model demonstrates the potential for early outcome prediction, enabling timely treatment modifications to improve patient management.

Keywords Brain metastases, Deep learning, Local control, Radiomics, Stereotactic radiotherapy

*Correspondence: Hemalatha Kanakarajan H.Kanakarajan@tilburguniversity.edu Wouter De Baene w.debaene@tilburguniversity.edu Full list of author information is available at the end of the article

© The Author(s) 2024. Open Access Th International License, which permits ar as you give appropriate credit to the or you modified the licensed material. You parts of it. The images or other third pa

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/bv-nc-nd/40/.



Open Access

Introduction

Metastatic brain tumors represent the most prevalent form of intracranial malignancies [1]. Brain metastases manifest in approximately 20–40% of individuals diagnosed with cancer [2, 3]. While any tumor has the potential to spread to the brain, the predominant types include lung cancer, breast cancer, melanoma, and gastrointestinal cancers [1]. The prevalence of brain metastases is increasing [4]. The adoption of sophisticated imaging methods for diagnosis, alongside the implementation of innovative chemotherapeutic approaches for systemic cancer treatment, may contribute to the increased likelihood of finding and developing brain metastases [1].

Currently, the prognosis of patients with brain metastases is poor with median overall survival of a few weeks to months in untreated patients [5]. The survival of patients with brain metastases depends upon prompt diagnosis and treatment efficacy. The standard treatment options are surgical resection and radiotherapy [5]. Surgery is recommended for patients with a single large tumor in a reachable location [6]. The three principal modalities of radiotherapy for brain metastases are Whole-Brain Radiation Therapy (WBRT), Single-fraction Stereotactic Radiosurgery (SRS), and hypo-fractionated Stereotactic Radiotherapy (SRT). WBRT was the main treatment in the past for patients with multiple brain metastases [7]. There has been a shift from WBRT to SRT and SRS due to the adverse effects of WBRT, such as fatigue and cognitive decline [8]. Through SRS, multiple non-coplanar beams are converged to deliver a single, high radiation dose to a targeted region whereas SRT delivers multiple, smaller doses of radiation over time. In SRS and SRT, the delivered radiation is confined to the lesion and there is a rapid dose fall-off at the edge of the treatment volume. Since the radiation dose is not delivered to the healthy brain tissue, there is a reduced likelihood of posttreatment cognitive decline compared to WBRT [9].

The assessment of Local Control (LC) of brain metastases is an important clinical endpoint. A stable disease after treatment is categorized as LC while a progressive disease indicates a Local Failure (LF) [10]. It may require several months before local changes of the treated lesions become evident on follow-up scans. Considering that the median survival of patients with brain metastases following radiotherapy can range between 5 months and 4 years [11, 12], timely identification of LF subsequent to radiotherapy is crucial as it offers the opportunity for timely tailored treatment modifications, ensuring that patients receive the most effective care and maximizing their chances of a favorable prognosis.

Cancer imaging analysis driven by Artificial Intelligence (AI) has the potential to revolutionize medical practice by revealing previously undisclosed characteristics from routinely obtained medical images [13]. These features can serve as valuable inputs for the development of machine learning models aimed at predicting the treatment response or LC of brain metastases [13]. This is particularly important given the advancement in Graphical Processing Unit (GPU) processing capabilities and the availability of large amounts of training data which have led to a rapid expansion in neural networks and deep learning techniques for regression and classification tasks [14]. Deep learning models have demonstrated significant potential in identifying crucial and unique features within medical image data across a range of applications, including cancer treatment [15–17]. Deep learning uses artificial neural networks to automatically learn features from raw data. In medical imaging, deep learning methods are applied directly to the images themselves, learning hierarchical representations of the data. Deep learning has been particularly successful in tasks like image classification, object detection, and segmentation [18]. The information extracted by the deep learning models from the tumor images can be used to predict treatment outcome [18-20]. Jalalifar et al. [21] introduced a novel deep learning architecture to predict the outcome of LC in brain metastasis treated with stereotactic radiation therapy using treatment-planning magnetic resonance imaging (MRI) alongside standard clinical attributes [21]. Their findings highlighted that the addition of deep learning features to the clinical features significantly enhanced the prediction accuracy.

Radiomics is another research domain for extracting quantitative features from medical images for different clinical applications [22]. Radiomics focuses on extracting quantitative features from medical images, such as texture, shape and intensity characteristics. These features are then used to characterize tumors or other abnormalities in the images. While both radiomics and deep learning are used in medical imaging, radiomics focuses on extracting handcrafted features (such as the manually delineated tumor segmentations) from images while deep learning learns features directly from raw data using neural networks [18, 20]. Numerous studies have underscored the efficiency of radiomic-based machine learning algorithms in predicting treatment outcomes across different medical conditions [23, 24]. The radiomic-based machine learning algorithms have also been efficiently applied for the prediction of LC of brain metastases after radiotherapy [25-27]. Karami et al. [25] proposed a radiomics framework to predict the LC in patients with brain metastasis treated with SRT. Based on the radiomics features, Kawahara et al. [28] proposed a neural network model for predicting the local response of metastatic brain tumor to Gamma Knife Radiosurgery (GKRS). Liao et al. [26] and Andrei et al. [27] demonstrated the value

of combining radiomic features and clinical features to enhance the prediction of brain metastases responses after GKRS. Their findings show that the addition of radiomic features to the clinical features improved the accuracy of the prediction models for LC of brain metastases.

The studies that used either radiomics or deep learning features together with the clinical features to predict LC of brain metastases after SRT showed that the addition of either radiomics or deep learning features increased the prediction accuracy of the models. Radiomics involves extracting so-called hand-crafted features which are computed using predefined mathematical functions applied to the region of interest in the image. Deep learning, by contrast, does not require pre-definition of features but extracts more abstract, high-dimensional image information in an exclusively data-driven manner. Although both radiomics and deep learning compute similar types of features, prior studies have indicated that these features may provide complementary information, potentially improving the prediction accuracy of the models (e.g., Gao et al. [45]). Chang et al. [40] and Hosny et al. [43] integrated deep learning and radiomics features in their prediction models for lung cancer patients. The study of Chang et al. [40] showed that the addition of deep learning features increased the performance of the models predicting the individual prognosis of patients with non-small cell lung cancer. On the other hand, the study of Hosny et al. [43] showed that adding deep learning features to radiomics features significantly increased the prediction accuracy of mortality risk for one treatment group but not for the other. Although adding deep learning to radiomics seems to increase the prediction performance of models in different oncology domains, it is not established yet whether the combination of radiomics and deep learning features can lead to a significant increase in the performance of the models to predict local control of brain metastases, which is the objective of the current study. A model trained with all these combined features might predict LC with a higher accuracy than the other models trained with a subset of these features, offering a more comprehensive understanding of treatment response, potentially leading to more tailored and effective interventions which may result in improved treatment outcomes, prolonged patient survival, and enhanced quality of life.

Methods

Data collection

We retrospectively collected the clinical data from 199 brain metastases patients from the Gamma Knife Center of the Elisabeth-TweeSteden Hospital (ETZ) at Tilburg, The Netherlands. This study was approved by the ETZ science office and by the Ethics Review Board at Tilburg University. The patients underwent GKRS at the Gamma Knife Center. Patients with incomplete clinical data (i.e. missing value for one or more clinical variables) were excluded from our data set, resulting in 129 patients included in the analyses. For these 129 patients, pre-treatment contrast-enhanced (with triple dose gadolinium) brain MRIs were collected using a 1.5T Philips Ingenia scanner (Philips Healthcare, Best, The Netherlands) with a T1-weighted sequence (TR/TE: 25/1.86 ms, FOV: 210×210×150, flip angle: 30°, transverse slice orientation, voxel size: 0.82×0.82×1.5 mm). These highresolution whole-brain planning scans were made as part of clinical care at the Gamma Knife Center of the ETZ between 2015 and 2021. For all patients, the segmentations of the baseline ground truth were manually delineated by expert oncologists and neuroradiologists at ETZ. At ETZ, the FU MRI scans were made at 3, 6, 9, 12, 15, and 21 months after treatment. A tumor was defined as progressive (i.e. local recurrence or local failure (LF)) if there was a relative increase in tumor volume on any of these follow-up MRIs compared to pretreatment MRI. We distinguish increase of the volume due to Adverse Radiation Effect from volume increase due to tumor progression. We consider progression as enlargement due to tumor progression. We consider a tumor stable (i.e. local control) as: no growth, smaller or disappearance of the metastasis. The approach for defining the stable and progressive disease is in line with that of RANO-BM [44]. The difference, however, is that we use tumor volume instead of the unidimensional longest diameter to measure the progression as volumetric criteria seem to outperform the unidimensional RANO-BM criteria (Ocaña-Tienda et al., [46]). The pre-processing, feature extraction, model training, and evaluation were performed in Python (version 3.11).

Preprocessing

As a first preprocessing step, all the MRI scans were registered to standard MNI space using Dartel in SPM12 (Wellcome Trust Center for Neuroimaging, London, UK), implemented in Python using the Nipype (Neuroimaging in Python-Pipelines and Interfaces) software package (version 1.8.6) [29]. The voxel size of the normalized image was set to 1*1*1 mm³. For all other normalization configurations, the default values offered by SPM12 were used. Normalizing to standard MNI space rescales and transforms scans to match the voxel size and spatial resolution of the MNI template. This ensures features extracted from the segmentations are comparable across patients. One other preprocessing step was to combine the ground truth labels for patients with more than one brain metastasis in one single ground truth mask. FSL library (Release 6.0) was used for this integration.

Pre-processing was applied to improve the reliability of radiomics and deep learning feature extraction [30].

Clinical features

The list of clinical factors that we collected from the Gamma Knife Center of ETZ were gender, survival status, diagnosis of brain metastases within 30 days after diagnosis of primary tumor, prior brain treatment, prior SRS, prior WBRT, prior surgery, prior systemic treatment, presence of extracranial metastases, presence of lymph node metastases, presence of seizure, number of metastases at diagnosis, Karnofsky Performance Status score (KPS), occurrence of new metastases after GKRS, presence of extracranial tumor activity, primary tumor type, age at diagnosis of brain metastases, age at diagnosis of primary tumor, presence of local recurrence, tumor volume and treatment dose. For the treatment dose, we took the average value from the dose range. We extracted the tumor volume from the segmentations of the baseline ground truth and added it to the clinical data. We took the total tumor volume across the metastases for patients with more than one brain metastasis. The clinical data was converted to a python dataframe.

Radiomics features

The segment-based radiomics features were extracted from the T1 weighted pre-treatment MRI scans using the radiomics feature extractor of the python radiomics package. The seven groups of features extracted from the Region Of Interest (ROI) of the tumor segmentations were shape-based features (14 features), first-order features (18), Gray Level Cooccurrence Matrix (GLCM) (24) features, Gray Level Dependence Matrix (GLDM) (14) features, Gray Level Run Length Matrix (GLRLM) (16) features, Gray Level Size Zone Matrix (GLSZM) (16) features, and Neighbouring Gray Tone Difference Matrix (NGTDM) (5) features. The resulting 107 radiomics features were considered in this study. The list of radiomics features extracted are listed in the appendix. The radiomics features were then combined with the clinical features to form a combined python dataframe. The mathematical definitions of these radiomics features are given in the Pyradiomics feature documentation (https://pyrad iomics.readthedocs.io/en/latest/features.html). The IBSI (Image Biomarker Standardisation Initiative) provides standardized guidelines for extracting and reporting radiomics features from medical images. PyRadiomics has incorporated most aspects of IBSI recommendations. The minor deviations from this standard are in the Pyradiomics documentation (https://pyradiomics.readt hedocs.io/en/latest/faq.html).

Deep learning features

A 3D ResNet model [31, 32] pre-trained on the ImageNet challenge dataset [33] (without any custom training on our dataset) was used to extract the deep learning features from the manually segmented masks. Prior to input, the images were rescaled to $256 \times 256 \times 256$ using spline interpolation order 3, improving the accuracy of the model [34]. Additionally, the pixels were sample-wise scaled between -1 and 1 and the image was cropped around the segmentation mask without any padding. These preprocessing steps contribute to optimizing the performance of the 3D ResNet model in extracting meaningful features from the images [35].

A 3D convolution was applied on the training data. This convolutional layer was designed with 50 filters and a large kernel size of (7, 7, 7), while also employing a stride of (2, 2, 2) for down sampling. The purpose of this stride is to efficiently reduce the spatial dimensions of the input data, capturing broader information across the dataset while managing computational complexity [36].

To fine-tune this convolutional layer, we applied batch normalization and Rectified Linear Unit (ReLU) activation. Batch normalization helps the model to adapt to our dataset, improving its performance, stability and ability to generalize. ReLU is like a simple on/off switch for a neuron in a neural network. ReLU helps the neural networks learn by letting positive signals pass through unchanged while ignoring negative ones and thus enables neural networks to learn complex patterns effectively. ReLU transforms the features from the images to be compatible with the pretrained model, maintain consistency with the original training, and facilitate efficient gradient propagation during the fine-tuning process [37, 38].

Following this, we incorporated three fine-tuned ResNet blocks into the model. After adding the ResNet blocks, we applied global average pooling to reduce the spatial dimensions to $1 \times 1 \times 1$. Finally, a dense layer with softmax activation was added for classification. The 50 deep learning features extracted using this fine-tuned 3D ResNet model were then combined with the clinical and radiomics features to form a combined python data-frame. We extracted the deep features from the output of the GlobalAveragePooling3D layer which is the second-to-last layer in the model.

The complete process of pre-processing and feature extraction is summarized in Fig. 1.

Model training

The features with low variance (<0.01) were determined and excluded from the combined dataset to improve prediction accuracy. The list of the excluded features is included in the Appendix. The data was



Fig. 1 The process of pre-processing, feature extraction, and combining the data

then normalized, balanced using SMOTE [42] and was supplied to the Random Forest classifier. Experimental results from Chen et al. [39] demonstrated that the Random Forest machine learning algorithm achieves a better classification performance compared to other classification algorithms. Hence, we choose the Random Forest machine learning algorithm to predict LC from the combined data. The model was trained using a training data set that comprised 80%, 86% or 90% of the total data, depending on the k-value (as detailed below) and was subsequently cross-validated with corresponding validation data sets containing 20%, 14% or 10% of the data, respectively. The process of training and evaluation of the models is shown in Fig. 2. The binary outcome used in training and validation was the LC after treatment taken from the list of clinical features. The different models that we created were:

- 1. Random Forest classifier trained with clinical features only.
- 2. Random Forest classifier trained with the combination of clinical and deep learning features.
- 3. Random Forest classifier trained with the combination of clinical and radiomics features.
- 4. Random Forest classifier trained with the combination of clinical, radiomics and deep learning features.

Model evaluation

The performance of the model was evaluated by measuring the following metrics: classification accuracy, precision, F1 score, recall and AUC. The classification accuracy is the ratio of the number of correct predictions to the total number of input samples. The precision is the ability of the classifier not to label as positive a sample that is negative and recall is the ability of the classifier to find all the positive samples. In other words, precision is the ratio of true positive predictions to the total number of positive predictions made by the model, while recall is the ratio of true positive predictions to the total number of actual positives in the dataset. The F1 score in percentage gives the balance between how often the model is correct (precision) and how well it finds all the positive instances (recall). A Receiver Operating Characteristic (ROC) is a graphical plot which is created by plotting the true positive rate vs the false positive rate at various threshold settings. The AUC computes the area under the ROC curve. By doing so, the curve information is summarized in one number. Similar to the F1 score, the AUC reaches its best value at 1.

A K-fold cross-validation was applied on the model. The different values that we used for K during cross-validation were 5, 7, and 10. The average accuracy and other metrics across the different folds was calculated. From the trained models, we also extracted the importance of the various factors for predicting the LC.



Fig. 2 Model training, evaluation and prediction

Comparison of models

To determine whether there was a significant difference in performance between the four models and to understand the practical significance of the observed differences, we statistically compared the accuracy of the four models. We employed the Friedman test in Python to analyze the significance of the difference in accuracies across all validation folds across all values of K. When the Friedman test showed significant difference in performance across the models, a pairwise comparison was performed using Wilcoxon signed-rank tests with a False Discovery Rate (FDR) correction using the Benjamini–Hochberg method. Statistical significance was set to p < 0.05.

Results

Patient characteristics

Table 1 shows the characteristics of patients included in our study. Among the 129 patients, 42% were male and 58% were female. The patients had an average age of 63 and an average tumor volume of 17,445 mm³. Sixty-nine % of the patients had a primary lung cancer and 94% of the patients had less than 10 brain metastases.

The average performance metrics for the four models across the cross-validation datasets are shown in Table 2. In bold is the value of the best score for the corresponding performance metric.

The accuracy of the Random Forest model trained with clinical features only was 75.0%. The model trained with the combination of clinical and deep learning features had an improved accuracy of 78.0%. The model trained with the combination of clinical and radiomics features was even higher at 79.33%. But the model with the highest prediction accuracy of 81.66% was the model trained with the combination of clinical, radiomics and deep learning features. This model also achieved the highest precision of 81.33%, F1 score of 86.44% and the best AUC of 0.88. The model trained with the combination of clinical and radiomics features showed the second highest accuracy of 79.3%, precision of 77.66%, F1 score of 78% and an AUC of 0.86. The model with the lowest prediction accuracy of 75.0% was the model trained with clinical features only. This model also achieved the lowest recall of 73.33%, F1 score of 73%, and a Precision of 74%. The ROC curve of the model trained with the combination of clinical, radiomics and deep learning features is shown in Fig. 3. The ROC curve of the other three models were added to the Appendix. We also calculated the average stability of the feature selection of the models using the Pearson's correlation across the cross-validation folds [47, 48]. The stability of the model with clinical features only is 88%. The stability of the clinical and radiomics

Table 1 Patient characteristics

Gender	
Male	54 (42%)
Female	75 (58%)
Extracranial tumor activity	
Yes	39
No	90
Diagnosis of brain metastases within 30 days after diagno- sis of primary tumor	
Yes	39
No	90
Prior brain treatment	
Yes	23
No	106
Prior SBS	100
Yes	15
No	114
Prior WBRT	
Yes	7
No	, 122
Prior surgery	122
Yes	8
No	121
Prior systemic treatment	121
Yes	76
No	53
Presence of extracranial metastases	55
Yes	54
No	75
Presence of lymph node metastases	/5
Yes	68
No	61
Presence of seizure	
Yes	18
No	111
KPS score	
60	3
70	14
80	31
90	38
100	43
Occurrence of new metastases after GKRS	
Yes	62
No	67
Presence of local recurrence	
Yes	40
No	89
Total tumor volume (mm ³)	
Average	17,445
	(88-88,029
Tracting out door (C.)	(T)(T) ⁼)
Average (minimum, meximum)	22 (10 25)
Average (minimum–maximum)	22 (18–25)

Table 1 (continued)

Age at diagnosis of brain metastases (years)	
Average (minimum–maximum)	63 (36–85)
Primary tumor type	
Lung	89
Melanoma	8
Breast	3
Others	29
Number of brain metastases	
1	30
2–3	50
4–10	41
>10	8

model and the model trained with combination of all features is 87%. The model with clinical and deep learning features had a feature selection stability of 70%.

Gini Importance, the standard measure of feature importance used by the Random Forest algorithm in scikit-learn, was used to define the variable importance. The variable importance of the top 10 factors for predicting the LC for each of the models were added to the appendix. For the model trained with the combination of clinical, radiomics and deep learning features, the top 10 features associated with the prediction of LC were tumor volume, original_shape_Elongation, shape_VoxelVolume, original_shape_Maxioriginal mum2DDiameterSlice, original_shape_Sphericity, original_firstorder_90Percentile, original_firstorder_Entropy, Average dose, original_shape_Flatness, and original_ shape_Maximum3DDiameter. Except for tumor volume and average dose which are clinical features, the rest of these top 10 features are radiomics features. There were no deep learning features in this list. Tumor volume emerged as the variable with the highest importance in the combined model.

Outcome of statistical analysis

We used the Friedman test to analyze the significance of the difference in accuracies across validation folds across all values of K. The results of the test indicate that there is a statistically significant difference in the accuracy of the four models with a Chi-squared value of 10.26 and a p-value of 0.016. This means that at least one of the models is performing significantly different from the others. To determine which models differ significantly, we performed pairwise Wilcoxon signed-rank tests with corrections for multiple comparisons using the FDR method. The outcomes of these tests are presented in Table 3. There is a statistically significant difference (after FDR correction) between the model trained with clinical features only and the models trained with the combination of all features (p-value=0.008) and the combination of clinical and radiomics features (p-value=0.042). Also, there is a significant difference between the models trained with clinical and deep learning features and the model trained with all features (p-value = 0.042).

Discussion

Brain metastases patients treated with SRT are at risk of developing local failure. Prompt diagnosis of local failure might increase treatment options and hence improve treatment outcomes. In this study, we trained and tested machine learning algorithms to predict the local failure using clinical features and T1-weighted MR imaging features. Four distinct models were developed: One model was trained with clinical features only, one with the combination of clinical and deep learning features, one with the combination of clinical and radiomics features and one with the combination of clinical, radiomics and deep learning features.

Our findings align with and expand upon prior studies in this field. Jalalifar et al. [21] introduced a novel deep learning architecture to predict the outcome of LC in brain metastasis treated with stereotactic radiation therapy using treatment-planning magnetic resonance imaging and standard clinical attributes. The

Table 2 Average performance of the models along with their 95% confidence intervals on the validation datasets

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	AUC
Model trained with clinical features only	75.0	74.0	73.33	73.0	0.85
	(73.86, 76.13)	(72.04, 75.96)	(69.87, 76.79)	(71.86, 74.13)	(0.83, 0.86)
Model trained with the combination of clinical and deep learning features	78.0	75.33	83.0	78.33	0.82
	(76.04, 79.96)	(72.48, 78.18)	(79.08, 86.92)	(77.67, 78.98)	(0.80, 0.84)
Model trained with the combination of clinical and radiomics features	79.33	77.66	80.0	78.0	0.86
	(78.02, 80.64)	(75.93, 79.39)	(77.73, 82.26)	(78.0, 78.0)	(0.84, 0.88)
Model trained with the combination of clinical, radiomics and deep learning features	81.66	81.33	80.33	86.44	0.88
	(77.69, 85.64)	(79.60, 83.06)	(75.75, 84.90)	(83.27, 89.60)	(0.85, 0.91)



Fig. 3 ROC curve of model trained with clinical, deep learning and radiomics features

Table 3	Outcome	of statistical	anal	vsis
---------	---------	----------------	------	------

Models	p-value	Corrected p-value
Clinical features only vs clinical and deep learning features	0.615	0.615
Clinical features only vs clinical and radiomics features	0.0149	0.042
Clinical features only vs combination of all 3 features	0.0013	0.008
Clinical and deep learning features vs clinical and radiomics features	0.290	0.349
Clinical and deep learning features vs combination of all 3 features	0.021	0.042
Clinical and radiomics features vs combination of all 3 features	0.241	0.349

In bold is the statistically significant values

accuracy of the model developed with only the clinical features was 67.5%, but the addition of deep learning features to the clinical features substantially increased the prediction accuracy to 82.5%. Similarly, our model trained with clinical and deep learning features provided a prediction accuracy of 78.0% representing an improvement over clinical features only (75.0%). However, in contrast to Jalalifar et al., the addition of deep learning features in our study did not yield a similar magnitude of improvement.

Kawahara et al. [28] demonstrated the potential of radiomics features in this domain using a neural network model including only the radiomics features which achieved an accuracy of 78% in predicting the local response of metastatic brain tumors to GKRS. Our model trained with a combination of the clinical and radiomics features provided a prediction accuracy of 79.33%. Importantly, this model outperformed the model including the clinical and deep learning features. In this study, we combined both hand-crafted radiomics and

deep learning features with clinical features to predict LC of brain metastases. This model achieved an accuracy of 81.66%, representing the best performance among all models tested.

The statistical analysis showed that models incorporating radiomics significantly outperformed those trained solely on clinical features or on the combination of clinical and deep learning features, reinforcing the robustness of hand-crafted radiomics features in capturing imaging-based characteristics that are critical for predicting LC. Adding deep learning features to the clinical features only or to the combination of clinical and radiomics features did not significantly improve the prediction accuracy. These results suggest that while deep learning features have shown utility in prior studies (e.g. Jalalifar et al. [21]), their contribution in this specific context may be limited, possibly by their overlap with radiomics features. The outcome of our study is in similar lines with the study of Hosny et al. [43] who found that adding deep learning features to radiomics features did not significantly increase the prediction accuracy for the radiotherapy treatment group of lung cancer patients, although it did significantly increase prediction accuracy for the surgery treatment group.

Although the difference in performance between the model trained with combined features and the model trained with clinical and radiomics features did not reach statistical significance, the performance of the combined model is higher when compared to all the other models trained with a subset of the features. The ability to predict LC with high accuracy before initiating SRT treatment offers an invaluable opportunity for tailoring treatment strategies for the best outcomes. Providing clinicians with information on the risk of local recurrence for individual patients empowers them to discuss these risks with patients prior to SRT. The capability to predict LC prior to treatment not only aids in informed decisionmaking regarding SRT but also opens avenues to consider alternative treatment modalities such as systemic therapy or WBRT. Additionally, it can enable clinicians to explore alternative radiotherapy approaches such as fractionated SRT or SRS with higher dose, depending on the predicted risk of local recurrence. Conversely, in cases where the risk of local recurrence is deemed low, SRT may be favored over other treatment options. Ultimately, pre-treatment prediction of LC serves as a valuable tool for both clinicians and patients, facilitating shared decision-making and optimizing treatment plans tailored to the needs and risk profiles of individual patients. Furthermore, insights into the variable importances provided by the model could offer valuable insights into the features associated with the LC of brain metastases, potentially guiding future research and clinical decision-making.

Tumor volume emerging as the feature with highest variable importance shows that the local control after SRS is highly correlated to the total tumor volume. This is in accordance with the clinical study of Baschnagel et al. [41] who found that total brain metastasis volume was a strong and independent predictor for local control. Although this study focused on creating a model for predicting the LC after SRT, the same approach can be extended to other treatment options and to the prediction of other clinical endpoints like overall survival.

One limitation of this study lies in the brain metastases segmentation procedure. Expert oncologists and neuroradiologists at ETZ manually delineated the segmentations of the baseline ground truth on all the planning MRI scans used in this study. A fully automated model using an AI-based segmentation system would have been better. Another limitation of this study is that the local control is included as a binary outcome variable. This restricts the evaluation of the temporal aspect of the local control which could provide valuable insights on the time at which the local failure occurs. Additionally, combining tumor volumes and using a single progression label simplifies the analysis and represents overall tumor volume, but it may restrict insights into the heterogeneity of metastases, especially in cases with varied lesion behavior.

It is important to note that in this study, we exclusively used the T1 weighted MRI scans. Exploring additional sequences and extracting radiomics and deep learning features from them could potentially improve the accuracy of the prediction models even more. In addition, for a more rigorous evaluation of the efficacy and robustness of the models, further investigations involving larger patient cohorts, preferably with multi-institutional data are warranted. Furthermore, the inclusion of an external validation dataset could significantly improve the generalizability of the prediction model, strengthening confidence in its clinical applicability across diverse patient populations and healthcare settings.

Conclusion

The findings of this study show that the machine learning model trained with the combination of clinical, radiomics and deep learning features predict LC of brain metastases with a higher accuracy than models trained with a subset of these features. While the integration of multiple feature types generally improved predictive performance, adding deep learning features did not statistically enhance the model's performance. The increased prediction accuracy can lead to more tailored and effective interventions, resulting in improved treatment outcomes, prolonged patient survival, and enhanced quality of life.

Appendix Full list of Radiomics features

original shape Elongation original_shape_Flatness original_shape_LeastAxisLength original_shape_MajorAxisLength original_shape_Maximum2DDiameterColumn original_shape_Maximum2DDiameterRow original_shape_Maximum2DDiameterSlice original_shape_Maximum3DDiameter original_shape_MeshVolume original_shape_MinorAxisLength original_shape_Sphericity original_shape_SurfaceArea original_shape_SurfaceVolumeRatio original_shape_VoxelVolume original_firstorder_10Percentile original_firstorder_90Percentile original_firstorder_Energy original_firstorder_Entropy original firstorder InterguartileRange original_firstorder_Kurtosis original_firstorder_Maximum original_firstorder_MeanAbsoluteDeviation original_firstorder_Mean original_firstorder_Median original_firstorder_Minimum original_firstorder_Range $original_first order_RobustMeanAbsoluteDeviation$ original_firstorder_RootMeanSquared original_firstorder_Skewness original_firstorder_TotalEnergy original_firstorder_Uniformity original_firstorder_Variance original_glcm_Autocorrelation original_glcm_ClusterProminence original_glcm_ClusterShade original_glcm_ClusterTendency original_glcm_Contrast original_glcm_Correlation original_glcm_DifferenceAverage original_glcm_DifferenceEntropy original_glcm_DifferenceVariance original_glcm_Id original_glcm_Idm original_glcm_Idmn original_glcm_Idn original_glcm_Imc1 original_glcm_Imc2 original_glcm_InverseVariance original glcm JointAverage original_glcm_JointEnergy

original glcm JointEntropy original_glcm_MCC original_glcm_MaximumProbability original_glcm_SumAverage original_glcm_SumEntropy original_glcm_SumSquares original_gldm_DependenceEntropy original gldm DependenceNonUniformity original_gldm_DependenceNonUniformityNormalized original_gldm_DependenceVariance original_gldm_GrayLevelNonUniformity original gldm GrayLevelVariance original_gldm_HighGrayLevelEmphasis original_gldm_LargeDependenceEmphasis original_gldm_LargeDependenceHighGrayLevelEmphasis original gldm LargeDependenceLowGrayLevelEmphasis original_gldm_LowGrayLevelEmphasis original gldm SmallDependenceEmphasis original_gldm_SmallDependenceHighGrayLevelEmphasis original gldm SmallDependenceLowGrayLevelEmphasis original_glrlm_GrayLevelNonUniformity original_glrlm_GrayLevelNonUniformityNormalized original glrlm GrayLevelVariance original_glrlm_HighGrayLevelRunEmphasis original glrlm LongRunEmphasis original_glrlm_LongRunHighGrayLevelEmphasis original_glrlm_LongRunLowGrayLevelEmphasis original glrlm LowGrayLevelRunEmphasis original glrlm RunEntropy original_glrlm_RunLengthNonUniformity original_glrlm_RunLengthNonUniformityNormalized original_glrlm_RunPercentage original glrlm RunVariance original_glrlm_ShortRunEmphasis original glrlm ShortRunHighGrayLevelEmphasis original_glrlm_ShortRunLowGrayLevelEmphasis original glszm GrayLevelNonUniformity original_glszm_GrayLevelNonUniformityNormalized original_glszm_GrayLevelVariance original_glszm_HighGrayLevelZoneEmphasis original glszm LargeAreaEmphasis original_glszm_LargeAreaHighGrayLevelEmphasis original glszm LargeAreaLowGrayLevelEmphasis original_glszm_LowGrayLevelZoneEmphasis original_glszm_SizeZoneNonUniformity original glszm SizeZoneNonUniformityNormalized

original_glszm_SmallAreaEmphasis	original_glrlm_GrayLevelNonUniformityNormalized
original_glszm_SmallAreaHighGrayLevelEmphasis	original_glrlm_LowGrayLevelRunEmphasis
original_glszm_SmallAreaLowGrayLevelEmphasis	original_glrlm_RunLengthNonUniformityNormal-
original_glszm_ZoneEntropy	ized
original_glszm_ZonePercentage	original_glrlm_RunPercentage
original_glszm_ZoneVariance	original_glrlm_ShortRunEmphasis
original_ngtdm_Busyness	original_glrlm_ShortRunLowGrayLevelEmphasis
original_ngtdm_Coarseness	original_glszm_GrayLevelNonUniformityNormal-
original_ngtdm_Complexity	ized
original_ngtdm_Contrast	original_glszm_LowGrayLevelZoneEmphasis
original_ngtdm_Strength	original_glszm_SmallAreaEmphasis
	original_glszm_SmallAreaLowGrayLevelEmphasis
List of excluded features	original_ngtdm_Coarseness
	DL_5
original_firstorder_Uniformity	DL_13
original_glcm_Id	DL_31
original_glcm_Idm	DL_34
original_glcm_Idmn	DL_43
original_glcm_Idn	DL_48
original_glcm_Imc2	
original_glcm_InverseVariance	See Figs.
original_glcm_JointEnergy	4,
original_glcm_MCC	5,
original_glcm_MaximumProbability	6 and
original_gldm_LowGrayLevelEmphasis	7.
original_gldm_SmallDependenceLowGrayLevelEm-	
phasis	



Fig. 4 Variable importance for LC in decreasing order of significance for top 10 factors. A—model trained with clinical factors only. B—model trained with clinical and deep learning factors. C- model trained with clinical and radiomics features. D—model trained with the combination of clinical, radiomics and deep learning features



Fig. 5 ROC curve of model trained with clinical features only



Fig. 6 ROC curve of model trained with clinical and deep learning features



Fig. 7 ROC curve of model trained with clinical and radiomics features

Abbreviations

- SRS Stereotactic Radiosurgery
- LC Local Control
- ETZ Elisabeth-TweeSteden Hospital
- AUC Area Under the receiver operating characteristic Curve
- WBRT Whole-Brain Radiation Therapy
- SRT Hypo-fractionated Stereotactic Radiotherapy
- LF Local Failure
- Al Artificial Intelligence
- GKRS Gamma Knife Radiosurgery
- MRI Magnetic Resonance Imaging
- KPS Karnofsky Performance Status score
- ROC Receiver Operating Characteristic
- GPU Graphical Processing Unit
- ReLU Rectified Linear Unit

Acknowledgements

We would like to acknowledge the support provided by Eline Verhaak for this research and thank her for helping us during the retrospective collection of clinical data from the Gamma Knife Center of the Elisabeth-TweeSteden Hospital at Tilburg, The Netherlands.

Author contributions

Conceptualization, H.K., W.d.B., P.H., and M.S.; Methodology, H.K.; Formal Analysis, H.K.; Writing—Review & Editing, H.K., W.d.B., P.H., and M.S.; Supervision, W.d.B., and M.S.

Funding

This research is supported by KWF Kankerbestrijding and NWO Domain AES, as part of their joint strategic research programme: Technology for Oncology IL. The collaboration project is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public–private partnerships.

Availability of data and materials

The data used for this study is available at ETZ and is accessible after approval from the ETZ Science office. No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

This study is part of the AI in Medical Imaging for novel Cancer User Support (AMICUS) project at Tilburg University. This project is approved by the Ethics Review Board at the Tilburg University. Consent to participate: The data did not contain any identifiable personal information, therefore the need for consent to participate was waived by the Institutional Review Board Elisabeth-TweeSteden Hospital (ETZ), Tilburg, The Netherlands (Study number: L1267.2021-AMICUS).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands. ²Gamma Knife Center, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands. ³Department of Neurosurgery, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands.

Received: 2 July 2024 Accepted: 17 December 2024 Published online: 30 December 2024

References

- Nayak L, Lee EQ, Wen PY. Epidemiology of brain metastases. Curr Oncol Rep. 2011;14(1):48–54. https://doi.org/10.1007/s11912-011-0203-y.
- 2. Patchell RA. The management of brain metastases. Cancer Treat Rev. 2003;29(6):533–40. https://doi.org/10.1016/S0305-7372(03)00105-1.
- Bradley KA, Mehta MP. Management of brain metastases. Semin Oncol. 2004;31(5):693–701. https://doi.org/10.1053/j.seminoncol.2004.07.012.
- 4. Rogers LR. Neurologic complications of cancer. Neuro Oncol. 2009;11(1):96–7. https://doi.org/10.1215/15228517-2008-118.
- Preusser M, et al. Brain metastases: pathobiology and emerging targeted therapies. Acta Neuropathol. 2012;123(2):205–22. https://doi.org/10. 1007/s00401-011-0933-9.
- Carapella CM, Gorgoglione N, Oppido PA. The role of surgical resection in patients with brain metastases. Curr Opin Oncol. 2018;30(6):390–5. https://doi.org/10.1097/cco.00000000000484.
- Brown PD, Ahluwalia MS, Khan OH, Asher AL, Wefel JS, Gondi V. Wholebrain radiotherapy for brain metastases: evolution or revolution? J Clin Oncol. 2018;36(5):483–91. https://doi.org/10.1200/JCO.2017.75.9589.
- Brown PD, 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):401–9. https:// doi.org/10.1001/jama.2016.9839.
- Peña-Pino I, Chen CC. Stereotactic radiosurgery as treatment for brain metastases: an update. Asian J Neurosurg. 2023;18(02):246–57. https:// doi.org/10.1055/s-0043-1769754.
- Ko PH, Kim HJ, Lee JS, Kim WC. Tumor volume and sphericity as predictors of local control after stereotactic radiosurgery for limited number (1–4) brain metastases from nonsmall cell lung cancer. Asia Pac J Clin Oncol. 2020;16(3):165–71. https://doi.org/10.1111/ajco.13309.
- Sperduto PW, et al. Estimating survival in patients with lung cancer and brain metastases. JAMA Oncol. 2017;3(6):827–827. https://doi.org/10. 1001/jamaoncol.2016.3834.
- Sperduto PW, et al. Estimating survival in melanoma patients with brain metastases: an update of the graded prognostic assessment for melanoma using molecular markers (Melanoma-molGPA). Int J Radiat Oncol Biol Phys. 2017;99(4):812–6. https://doi.org/10.1016/j.ijrobp.2017.06.2454.
- Jalalifar SA, Soliman H, Sahgal A, Sadeghi-Naini A. A self-attention-guided 3D deep residual network with big transfer to predict local failure in brain metastasis after radiotherapy using multi-channel MRI. IEEE J Transl Eng Health Med. 2023;11:13–22. https://doi.org/10.1109/jtehm.2022.3219625.
- 14. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436– 44. https://doi.org/10.1038/nature14539.
- Jia S, Jiang S, Lin Z, Li N, Xu M, Yu S. A survey: Deep learning for hyperspectral image classification with few labeled samples. Neurocomputing. 2021;448:179–204. https://doi.org/10.1016/j.neucom.2021.03.035.
- Sarvamangala DR, Kulkarni RV. Convolutional neural networks in medical image understanding: a survey. Evolut Intell. 2021. https://doi.org/10. 1007/s12065-020-00540-3.
- 17. Wang M, Zhang Q, Lam S, Cai J, Yang R. A review on application of deep learning algorithms in external beam radiotherapy automated treatment planning. Front Oncol. 2020. https://doi.org/10.3389/fonc.2020.580919.
- Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. Annu Rev Biomed Eng. 2017;19(1):221–48. https://doi.org/10.1146/annur ev-bioeng-071516-044442.
- Wetzer E, Harlin H, Lindblad J, Sladoje N. When texture matters: texturefocused cnns outperform general data augmentation and pretraining in oral cancer detection. IEEE. 2020. https://doi.org/10.1109/isbi45749.2020. 9098424.
- Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Process Mag. 2019;36(4):132–60. https://doi. org/10.1109/msp.2019.2900993.
- Jalalifar SA, Soliman H, Sahgal A, Sadeghi-Naini A. Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRIbased deep learning features. Med Phys. 2022;49(11):7167–78. https:// doi.org/10.1002/mp.15814.
- O'Connor JPB, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017;14(3):169–86. https://doi.org/10.1038/nrclinonc. 2016.162.
- 23. Rizzo S, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp. 2018. https://doi.org/10.1186/s41747-018-0068-z.

- 24. Bae S, et al. Radiomic MRI phenotyping of glioblastoma: improving survival prediction. Radiology. 2018;289(3):797–806. https://doi.org/10.1148/radiol.2018180200.
- 25. Karami E, Ruschin M, Soliman H, Sahgal A, Stanisz GJ, Sadeghi-Naini A. An MR radiomics framework for predicting the outcome of stereotactic radiation therapy in brain metastasis. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, vol. 2019, pp. 1022–1025, 2019. https://doi.org/10.1109/EMBC.2019.88565 58
- Liao C-Y, et al. Enhancement of radiosurgical treatment outcome prediction using MRI radiomics in patients with non-small cell lung cancer brain metastases. Cancers. 2021;13(16):4030. https://doi.org/10.3390/cance rs13164030.
- Mouraviev A, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. Neuro Oncol. 2020. https://doi.org/10.1093/neuonc/noaa007.
- Kawahara D, Tang X, Lee CK, Nagata Y, Watanabe Y. Predicting the local response of metastatic brain tumor to gamma knife radiosurgery by radiomics with a machine learning method. Front Oncol. 2020;10: 569461. https://doi.org/10.3389/fonc.2020.569461.
- Gorgolewski K, et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. Front Neuroinf. 2011. https://doi.org/10.3389/fninf.2011.00013.
- Fan C, Chen M, Wang X, Wang J, Huang B. A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. Front Energy Res. 2021. https://doi.org/10. 3389/fenrg.2021.652801.
- Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9(4):611–29. https://doi.org/10.1007/s13244-018-0639-9.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp. 770–778, Jun 2016. https://doi.org/10.1109/cvpr.2016.90
- Russakovsky O, et al. ImageNet large scale visual recognition challenge. Int J Comput Vis. 2015;115(3):211–52. https://doi.org/10.1007/ s11263-015-0816-y.
- Keek SA, et al. Predicting adverse radiation effects in brain tumors after stereotactic radiotherapy with deep learning and handcrafted radiomics. Front Oncol. 2022. https://doi.org/10.3389/fonc.2022.920393.
- He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer, pp. 630–645, 2016, https://doi.org/10.1007/978-3-319-46493-0_ 38
- Riad R, Teboul O, Grangier D, Zeghidour N. Learning strides in convolutional neural networks. arXiv (Cornell University), Feb 2022. https://doi. org/10.48550/arxiv.2202.01653
- Hara K, Saito D, Shouno H, Analysis of function of rectified linear unit used in deep learning. In: 2015 International Joint Conference on Neural Networks (JJCNN). IEEE, pp. 1–8, 2015. https://doi.org/10.1109/ijcnn.2015. 7280578
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2012;60(6):84–90. https:// doi.org/10.1145/3065386.
- Chen R-C, Dewi C, Huang S-W, Caraka RE. Selecting critical features for data classification based on machine learning methods. J Big Data. 2020. https://doi.org/10.1186/s40537-020-00327-4.
- Chang R, Qi S, Wu Y, Yue Y, Zhang X, Qian W. Nomograms integrating CT radiomic and deep learning signatures to predict overall survival and progression-free survival in NSCLC patients treated with chemotherapy. Cancer Imaging. 2023. https://doi.org/10.1186/s40644-023-00620-4.
- Baschnagel AM, et al. Tumor volume as a predictor of survival and local control in patients with brain metastases treated with Gamma Knife surgery. J Neurosurg. 2013;119(5):1139–44. https://doi.org/10.3171/2013.7. jns13431.
- Elreedy D, Atiya AF, Kamalov F. A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Mach Learn. 2023. https://doi.org/10.1007/s10994-022-06296-4.

- Hosny A, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med. 2018. https://doi.org/ 10.1371/journal.pmed.1002711.
- 44. Douri K, et al. Response assessment in brain metastases managed by stereotactic radiosurgery: a reappraisal of the RANO-BM criteria. Curr Oncol. 2023;30(11):9382–91. https://doi.org/10.3390/curroncol30110679.
- 45. Gao Y, et al. An integrated model incorporating deep learning, handcrafted radiomics and clinical and US features to diagnose central lymph node metastasis in patients with papillary thyroid cancer. BMC Cancer. 2024. https://doi.org/10.1186/s12885-024-11838-1.
- Ocña-Tienda B, et al. Volumetric analysis: rethinking brain metastases response assessment. Neuro-Oncol Adv. 2023. https://doi.org/10.1093/ noajnl/vdad161.
- 47. Nogueira SO, Gavin B. Measuring the stability of feature selection, p. 442–457; 2016. https://doi.org/10.1007/978-3-319-46227-1_28
- Demircioğlu A. Benchmarking feature selection methods in radiomics. Invest Radiol. 2022. https://doi.org/10.1097/rli.00000000000855.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.