See SunSCAN 3D in action:
Request your demo today

Introducing SunSCAN™ 3D

The Next-Generation Cylindrical Water Scanning System

SunSCAN 3D simplifies beam scanning with SRS-class accuracy and user-centered design.

It enables faster, easier workflows, and hyper-accurate dosimetry for today's busy clinics.

Learn more:
sunnuclear.com

SUN NUCLEAR
A MIRION MEDICAL COMPANY

SunSCAN™ 3D is not available for sale in all markets. CE Mark pending.

Received: 26 December 2021 | Revised: 28 May 2022 | Accepted: 30 May 2022

DOI: 10.1002/mp.15814

RESEARCH ARTICLE

MEDICAL PHYSICS

Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRI-based deep learning features

Seyed Ali Jalalifar[1] | **Hany Soliman**[2,3,4] | **Arjun Sahgal**[2,3,4] | **Ali Sadeghi-Naini**[1,2,4,5]

[1]Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, Ontario, Canada

[2]Department of Radiation Oncology, Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

[3]Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada

[4]Physical Sciences Platform, Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Ontario, Canada

[5]Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada

Correspondence
Ali Sadeghi-Naini, Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, LAS 3047, Lassonde Building, 4700 Keele Street, Toronto, ON M3J 1P3, Canada.
Email: asn@yorku.ca

Funding information
Natural Sciences and Engineering Research Council of Canada, Grant/Award Numbers: CRDPJ507521-16, RGPIN-2016-06472; Lotte and John Hecht Memorial Foundation; Terry Fox Foundation, Grant/Award Number: 1083

Abstract

Background: A considerable proportion of metastatic brain tumors progress locally despite stereotactic radiation treatment, and it can take months before such local progression is evident on follow-up imaging. Prediction of radiotherapy outcome in terms of tumor local failure is crucial for these patients and can facilitate treatment adjustments or allow for early salvage therapies.

Purpose: In this work, a novel deep learning architecture is introduced to predict the outcome of local control/failure in brain metastasis treated with stereotactic radiation therapy using treatment-planning magnetic resonance imaging (MRI) and standard clinical attributes.

Methods: At the core of the proposed architecture is an InceptionResentV2 network to extract distinct features from each MRI slice for local outcome prediction. A recurrent or transformer network is integrated into the architecture to incorporate spatial dependencies between MRI slices into the predictive modeling. A visualization method based on prediction difference analysis is coupled with the deep learning model to illustrate how different regions of each lesion on MRI contribute to the model's prediction. The model was trained and optimized using the data acquired from 99 patients (116 lesions) and evaluated on an independent test set of 25 patients (40 lesions).

Results: The results demonstrate the promising potential of the MRI deep learning features for outcome prediction, outperforming standard clinical variables. The prediction model with only clinical variables demonstrated an area under the receiver operating characteristic curve (AUC) of 0.68. The MRI deep learning models resulted in AUCs in the range of 0.72 to 0.83 depending on the mechanism to integrate information from MRI slices of each lesion. The best prediction performance (AUC = 0.86) was associated with the model that combined the MRI deep learning features with clinical variables and incorporated the inter-slice dependencies using a long short-term memory recurrent network. The visualization results highlighted the importance of tumor/lesion margins in local outcome prediction for brain metastasis.

Conclusions: The promising results of this study show the possibility of early prediction of radiotherapy outcome for brain metastasis via deep learning of MRI and clinical attributes at pre-treatment and encourage future studies on larger groups of patients treated with other radiotherapy modalities.

KEYWORDS
brain metastasis, deep learning, magnetic resonance imaging, stereotactic radiotherapy, therapy outcome prediction

1 | INTRODUCTION

Brain metastases are the most prevalent malignancy of the central nervous system, with an incidence rate of 10% to 30% among cancer patients.[1] As a fatal complication of systemic disease, in the United States alone, about 100 000 cancer patients develop brain metastasis annually.[1] Because of the overall improvement in healthcare and the longer survival of cancer patients, the incidence of brain metastasis is projected to rise.[2]

Timely diagnosis and precise treatment are critical in the survival of patients suffering from brain metastasis. The origin of cancer, size/number of metastases, and associated symptoms are important factors in planning a treatment strategy for brain metastasis. Surgery, radiation therapy, and chemotherapy are the main treatment options for the management of metastatic brain tumors. Surgical resection is recommended for patients with a single large tumor in an accessible location.[3] Whole-brain radiation therapy (WBRT), single-fraction stereotactic radiosurgery (SRS), and hypo-fractionated stereotactic radiotherapy (SRT) are the three main modalities of radiation therapy for brain metastasis. For decades, WBRT was the treatment of choice for patients with multiple brain metastases,[4] but it has been associated with adverse side effects including cognitive deterioration and fatigue.[5,6] Due to such side effects, there has been a shift away from WBRT to SRS and SRT over the past two decades. Systemic therapy, particularly targeted therapy that has good penetration through the blood-brain barrier, is increasingly being incorporated into the management of patients with brain metastases.[7]

Local response to radiotherapy is highly varied among patients despite administrating standardized dose/fraction regimens due to many tumor-related factors such as histology, tumor size, and location. Additionally, patient-related factors such as genetics, age, and performance status are also predictors of tumor response.[8] Local response of brain metastasis to stereotactic radiation therapy is assessed based on the changes in the tumor size on follow-up serial images,[9] and can be categorized into local control (LC; shrinking or stable tumor) versus local failure (LF; enlarging tumor). It may take months, however, before a local response is evident on follow-up images, let alone the fact that early changes in tumor size are not always correlated with long-term local response. In particular, post-treatment lesion enlargement on imaging may not always be a sign of tumor progression, but rather of a condition known as pseudo-progression due to adverse radiation effect (ARE).[10] Given the median survival of as short as 5 months and up to 4 years,[11,12] early prediction of LF after radiotherapy can facilitate effective treatment adjustments, potentially resulting in improved treatment outcomes, survival, and quality of life.

The therapeutic paradigm for brain metastasis has steadily shifted to focus more on tailored treatments based on the subgroups and predicted survival.[13] Recursive partitioning analysis (RPA) was one of the first methods to classify patient prognosis based on age, performance status, control of primary tumor, and extent of extracranial disease.[14–16] A success at the time but overly simplistic, RPA is now replaced by more sophisticated stratification methods such as diagnosis-specific graded prognostic assessment (DS-GPA).[13,17] In this method, a GPA of 4.00 and 0.00 is associated with the best and the worst prognosis, respectively. The DS-GPA is calculated using prognosis factors based on the primary site of cancer, age, and Karnofsky performance status (KPS).[18] The KPS is a clinical metric to quantify the ability of cancer patients to perform everyday activities, with a score in the range of 0-100.

Prognostic features could also be mined from textural information on medical images. A large body of research shows that imaging modalities, such as magnetic resonance imaging (MRI), potentially provide relevant prognostic information that, if appropriately retrieved, can be utilized to predict therapy outcome.[19–21] Radiomics is an emerging translational field of research concerned with the high-throughput mining of high-dimensional medical imaging data to discover quantitative diagnostic and prognostic features.[22] Studies show the effectiveness of radiomic features as prognostic factors. Karami et al. have proposed an MRI-based radiomic framework for early prediction of treatment outcome in brain metastasis patients treated with SRT.[23] Liao et al.[24] explored the use of radiomics and clinical features in conjunction with support vector machines to predict survival and local response of the tumor for patients diagnosed with brain metastasis and treated with Gamma Knife radiosurgery. Their study shows that combining clinical and radiomic features improve the capability of the model to predict both tumor's local response and overall survival. Mouraviev et al.[25] proposed extracting radiomic features from the tumor core and the peritumoral regions and trained a random forest classifier on these features to predict local control in brain metastasis treated with stereotactic radiosurgery. Their result shows that an optimized combination of

radiomic and clinical features resulted in a 19% increase in the area under the receiver operating characteristic (ROC) curve (AUC) compared to the clinical features alone. Some studies also demonstrate correlations between the radiomic signature of tumors and their phenotypes and genomic and proteomic profiles.[26]

While radiomic features are handcrafted, deep learning algorithms could be used to extract distinguishing relevant imaging features automatically. Deep learning models have proven to be quite effective at identifying important and distinct characteristics, particularly in image data.[27,28] Deep models have the capability to outline regions of interest automatically, capture textural changes within a lesion, discriminate between cancerous and non-cancerous cells, and potentially extract distinctive information from lesions to be later used for the task of outcome prediction.[29–38] Diamant et al.[39] hypothesized that convolutional neural networks could enhance the performance of traditional radiomics, by detecting image patterns that may not be covered by a traditional radiomic framework. They tested their hypothesis for the task of head and neck cancer therapy outcome prediction and their results show that deep models can explicitly recognize traditional radiomic features and perform accurate outcome prediction. A recent study by Cho et al.[40] suggests that using deep learning methods instead of classic machine learning results in brain metastasis detection with a lower false-positive rate.

Many scenarios in medical imaging analysis require dealing with a sequence of spatially connected images, i.e., a 3D volume, such as an MRI volume. Recurrent neural networks (RNN) are a suitable fit for dealing with spatial dependency. RNN allow quantifying the information persisting between image slices, whereas typical 2D convolutional neural networks (CNN) alone do not.[41] Long short-term memory (LSTM) networks are a special kind of RNNs capable of learning long-term dependencies.[42]

In this study, a novel deep learning framework is proposed and investigated, for the first time, to predict local failure in brain metastasis treated with SRT using the treatment planning MRI and clinical information available at pre-treatment. The framework consists of a CNN (InceptionResNetV2) to extract textural features from 2D slices in the input MRI volumes, followed by an LSTM network to account for the spatial dependency between the 2D slices. The framework is capable of integrating the conventional clinical factors such as histology, tumor location, size, and the number of brain metastases, with the deep learning features of MRI in a comprehensive data-driven model for therapy outcome prediction. The results of the study show that coupling the clinical factors and deep-learning-based MRI features associated with the entire tumor volume improves the performance of therapy outcome prediction model considerably. For further comparison, two other models with state-of-the-art architectures, namely, sequence to sequence (Seq2Seq) and transformer networks, were investigated to incorporate inter-slice dependencies. The results demonstrate that the proposed model with the LSTM network could outperform the other two models in therapy outcome prediction. The results obtained with advanced methods of visualizing the network's decision basis highlight the importance of tumor/lesion margin characteristics on MRI in therapy response prediction.

2 | MATERIALS AND METHODS

2.1 | Study protocol and data acquisition

This study was conducted in accordance with institutional research ethics approval from Sunnybrook Health Sciences Centre (SHSC), Toronto, Canada. The imaging and clinical data were collected from 124 patients diagnosed with brain metastasis and treated with SRT over five fractions. The imaging data applied in this study for therapy outcome prediction included gadolinium contrast-enhanced T1-weighted (T1w), and T2-weighted-fluid-attenuation-inversion-recovery (T2-FLAIR) images acquired before the treatment (baseline). The in-plane image resolution and the slice thickness were 0.5 and 1.5 mm for the T1w and 0.5 and 5 mm for T2-FLAIR images, respectively. The dataset also included the treatment-planning tumor contours for each patient delineated by expert oncologists and the edema contours outlined under their supervision. Among the 124 patients (156 lesions), 99 patients (116 lesions) were randomly selected for training and optimization of the predictive models (10 patients with 15 lesions as the validation set for optimizing the model hyperparameters) and 25 patients (40 lesions) were kept unseen as an independent test set. The distribution of samples in terms of clinical attributes including age, gender, tumor size, histology, and outcome were inspected in the training and test sets to ensure statistical similarity.

The patients were followed up with MRI after radiotherapy on a 2 to 3-month schedule. The median imaging follow-up for all patients was 8 months. The lesions were monitored on serial MRI and the local response was determined for each lesion by a radiation oncologist and neuroradiologist based on the RANO-BM criteria.[9] The local outcome was defined as LC (complete response, partial response, or stable disease) or LF (progressive disease) identified in the last patient follow-up. ARE was diagnosed and differentiated from local progression using serial imaging (including perfusion MRI) and/or histological confirmation[43] based on the report by Sneed et al.[44] In keeping with these, 93 and 63 lesions were categorized with an LC and LF outcome, respectively.

2.2 | Preprocessing

The baseline T1w and T2-FLAIR images were resampled with a voxel size of $0.5 \times 0.5 \times 1$ mm^3. The resampled MRI volumes had a size of $512 \times 512 \times 174$ voxels. The T2-FLAIR images were co-registered on T1w images using an affine registration. To ensure a local outcome prediction on separated lesions, the size of the smallest sub-volume that encompasses the entire region of interest (ROI), including the tumor and edema (lesion) and their 5-mm outer margin,[45] was determined for the individual lesions. Observations of the study presented in[45] demonstrate that MRI radiomic features derived from the lesion margin (3-5 mm) can contribute to radiotherapy outcome prediction models in brain metastasis. A sub-volume of $128 \times 128 \times 45$ voxels was determined as a fit standard for all lesions. Subsequently, the sub-volumes associated with individual lesions were cropped from the T1w and T2-FLAIR images. The tumor and edema contours were used to generate ROI masks (tumor + 5-mm margin for T1w; tumor + edema + 5-mm margin for T2-FLAIR) for each lesion that were used to mask out the areas outside the ROI within the cropped sub-volumes. The voxel intensities in each image were normalized to the range of 0 and 1.

2.3 | Clinical features

Standard clinical features at the baseline including the histology (primary cancer), tumor location (infratentorial/ supratentorial), tumor size (longest diameter in mm), number of brain metastases, total dose (Gy), previous WBRT (yes/no), prior SRT/SRS (yes/no), GPA (from 0 to 4) along with age (year) and gender (male/female) were collected for each patient and their performance was investigated in therapy outcome prediction with and without the deep learning features of MRI. The categorical features were converted to vectors using one-hot encoding, while the continuous features (e.g., tumor size) were first discretized to categories and then converted to one-hot encoding format. A 3-layer fully-connected multi-layer perceptron (MLP) was trained and optimized using the training and validation data to predict LC/LF for each lesion solely with the clinical features. The MLP model included an input layer with 2 to 48 neurons (depending on the input features), one hidden layer with 10 neurons, and an output layer with two neurons (LC/LF). Feature selection was performed through an exhaustive search among all possible combinations of the features to obtain the best feature set based on the accuracy of the model on the validation set. The selected features were also coupled with the deep learning features in the comprehensive model developed for therapy outcome prediction (described further below).

2.4 | Deep learning and visualization framework

2.4.1 | System overview

Figure 1 demonstrates an overview of the deep learning framework developed and investigated for LC/LF outcome prediction using the baseline MRI data and clinical features. At the heart of the proposed system, an InceptionResNetV2 is trained to classify the local response of lesions using associated single slices of the T1w and T2-FLAIR images as two parallel input channels of the network. Using the trained network, for each MRI slice of the lesion, 256 features are extracted from the last fully-connected layer of the network (Figure 1a). The clinical features are then fused with the extracted features from the 2D MRI slices through concatenation and fed to either an LSTM (Figure 1b), Seq2Seq (Figure 1c), or transformer network (Figure 1d) to incorporate 3D spatial dependencies that exist within volumetric MRI in predicting the therapy outcome of each lesion. The clinical features were fused with the deep-learning features of each MRI slice before feeding them to the LSTM, Seq2Seq, or transformer network to let this information propagate through the network during the training phase. The LSTM network consists of two layers with $N = 45$ LSTM cells in each layer. Each cell is connected to its adjacent cell in the same layer, propagating the current cell state to the next one. The first layer of LSTM provides the input to the second layer and is connected to it in the same fashion that the original input is fed to the LSTM network. At the last cell of the second layer of the network, a dense layer with two output units and a *softmax* activation classifies the outcome of each lesion into either LC or LF. The Seq2Seq network has an encoder–decoder style LSTM architecture. Each of the encoder and decoder components consists of three LSTM layers (each layer with $N = 45$ cells). Each LSTM encoder/decoder outputs a cell state and a hidden state based on the previous cell state, previous hidden state, and the current input. A weighted sum of the hidden states in the encoder generates the context vector that is fed to the decoder to set the initial cell state and hidden state of the decoders. Each LSTM decoder produces a prediction, and the prediction of the last LSTM layer is applied as the overall outcome prediction of the Seq2Seq network for each lesion. The transformer input consists of 45 fixed-length vectors along with a trainable class (*cls)* token with the number of transformer blocks being 1, the number of attention heads being 2, and the hidden layer size of feed-forward network for input classification being 32. Through matrix multiplication, the transformer architecture allows the information to propagate from one slice to another in a way that all slices are considered when making a classification, with emphasis on more important slices. Further details on different components in

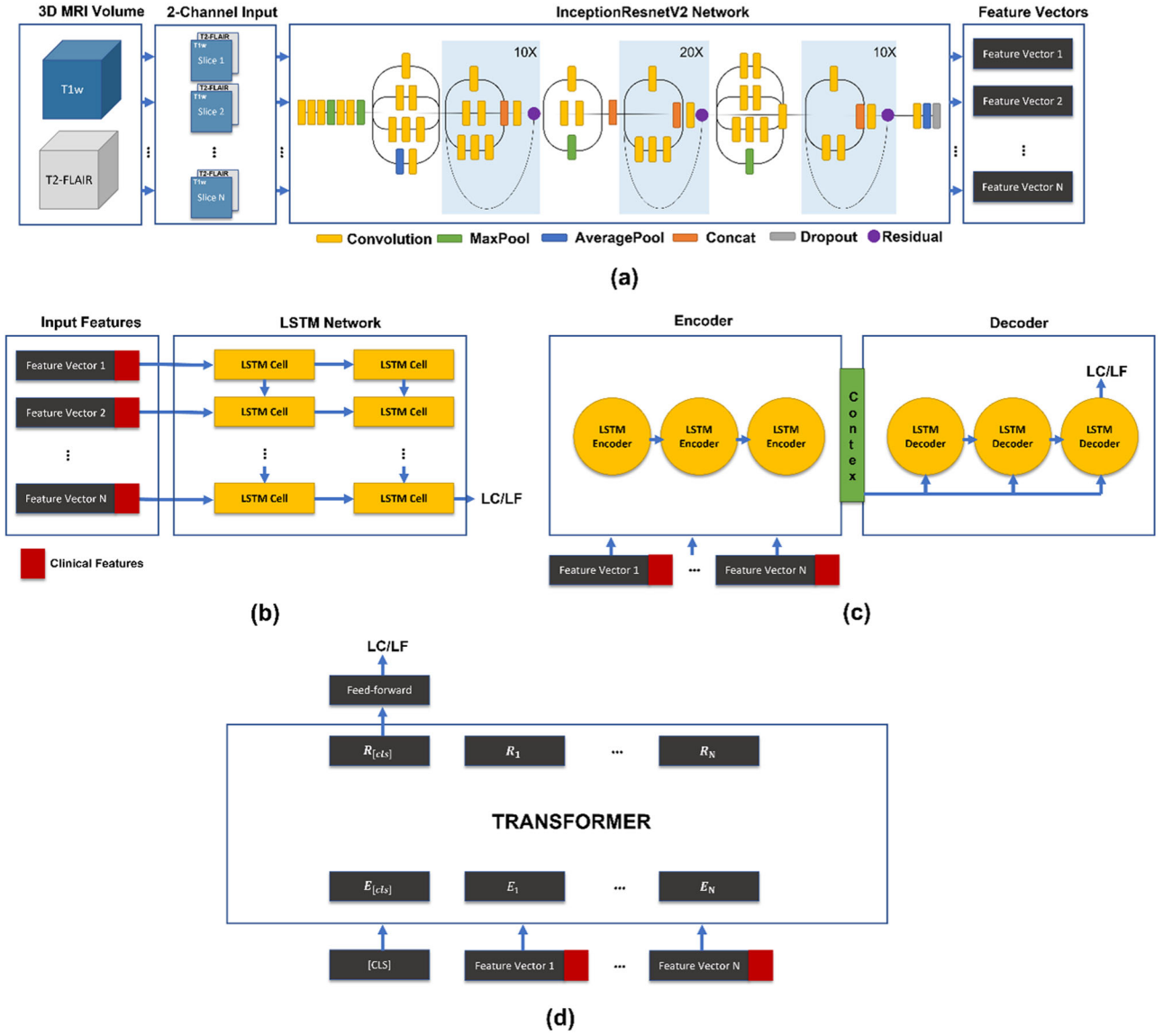

FIGURE 1 System overview. (a) Slice by slice feature extraction from MRI volumes using the InceptionResnetV2, (b) a two-layer LSTM, (c) a Seq2Seq network, and (d) a transformer network. The input feature vectors for all three networks are a concatenation of the features coming from the InceptionResNet and the clinical features.

the proposed framework, and the InceptionResNetV2, LSTM, Seq2Seq, and transformer architectures have been provided in the Supporting Information.

2.4.2 | Model training

The IncpetionResNetV2 network was trained for the task of therapy outcome prediction using all single slice two-channel images of T1w and T2-FLAIR (128 × 128 pixels ×2 channels of T1w and T2-FLAIR) associated with each lesion in the training set. The two-channel images included a standardized ROI encompassing the entire lesion (tumor + edema) and its 5 mm margin on the 2D imaging plane. As mentioned before, this network was developed to be used solely as a feature extractor in the framework. To serve this purpose, the prediction accuracy of the network on the validation set must improve as the ability of the network to predict therapy outcome is closely related to how distinctive the derived features in the last layer of the network are. In order to optimize the prediction accuracy of the IncpetionResNet, the network weights were initialized through pretraining, and the training was performed using a decreased batch size and learning rate through a curriculum learning strategy as described below. The network pretraining was performed using ImageNet, and then the BraTS dataset[46] on the task of brain tumor type classification.

Small batches can offer a regularization effect.[47] To obtain maximum generalization and prevent overfitting, a batch size of one was used during the network training, while to maintain stability due to the high variance

of the gradient, the learning rate was set to a very small value of 10^{-6}. While this strategy was successful in improving the network performance on the validation set, it resulted in a longer training process because of the small learning rate and also very small batch size.[48] The idea of curriculum learning is to start the network training on easier subtasks and gradually increase the task difficulty.[49] Since outcome prediction using single 2D images can be extremely difficult in some cases, for a smoother network training, the network was initially trained only on the MRI slices of each lesion with a larger tumor cross-section. This is usually the middle slices of a lesion in the axial plane of an MRI volume. Gradually, more challenging images were added to the training set. Practically, this method of feeding data to the network resulted in smoother training and better validation accuracy and loss.

The LSTM and Seq2Seq networks were trained with a learning rate of 0.0001 and batch size of 2. An *RMSprop* optimizer was used for optimizing the categorical cross-entropy loss function in both networks. A tanh was used as the activation function in all layers of the LSTM and Seq2Seq networks. Both networks were trained for a total of 500 epochs. The transformer was trained for 200 epochs with a batch size of eight and a learning rate of 0.0001. An Adam optimizer was used for optimizing the sparse categorical cross-entropy function. The *RELU* activation function was utilized for hidden layers of the transformer while the activation function for the last feed-forward layer (classifier) was *softmax*.

All the experiments were run in Python and models were trained and tested using Keras[50] with TensorFlow[51] backend. Additionally, we used the scikit-learn package[52] to calculate performance metrics and the matplotlib package[53] for visualization. All model trainings were performed on a single GeForce RTX TI 2080 graphic card. The training process took about 18 h (~54 m parameters), 2 h (60k parameters), 3 h (6 m parameters), and 5 h (7 m parameters), for the InceptionResNet, LSTM, Seq2Seq, and the transformer networks respectively. The total inference time for a single input was 27, 29, and 31 ms for the LSTM, Seq2Seq, and transformer networks.

2.4.3 | Visualization of network decision basis

The outcome prediction framework was supplemented by a visualization algorithm to illustrate the contribution of different areas of ROI on MRI to the network's prediction for each lesion. Specifically, the visualization algorithm generates a heatmap that color-codes the importance of different regions on the input images for the network's conclusion and can be used to interpret the rationale behind its decision for each case. A modified version of the prediction difference analysis (PDA) technique was adopted in the applied visualization method in conjunction with a sliding window analysis.[54] A 2×2 pixel sliding black square was used to occlude a small region of the input image iteratively before feeding it to the trained network for outcome prediction. The absolute difference in the output probability of the network (i.e., $|p_{\text{input}} - p_{\text{occluded\_input}}|$) was recorded in each iteration as a metric to measure the contribution of the occluded region and applied to generate the heatmap. A higher difference between the obtained probabilities typically demonstrates more important regions of the image with more-telling information for and a higher impact on the network's prediction.

3 | RESULTS

The clinical characteristics of the patients in this study have been summarized in Table 1. Among the 124 patients, 40% were male and 60% were female. The patients had an average age of 62 ± 15 years and an average tumor size of 2 ± 1.03 cm. The average GPA for the patients was 2.2.

The exhaustive search process to select the best clinical features for outcome prediction resulted in a set of four features including the histology, tumor location, tumor size, and number of brain metastases. Using these features, the optimized MLP network could predict the therapy outcome with an accuracy of 68%, a sensitivity of 65%, and a specificity of 70% on the independent test set. These results set the ground to investigate in the next step whether integrating clinical variables with the deep learning features extracted from MRI can improve the accuracy of radiotherapy outcome prediction.

Table 2 summarizes the performance of the deep learning networks on the validation and test set before and after integrating the clinical features. The table presents the results in terms of accuracy, sensitivity, specificity, and AUC. Since the InceptionResNet processes the MRI slices individually, to predict the therapy outcome for each lesion using this network solely, the output probability of the network was averaged over all slices associated with the entire lesion volume before thresholding it to obtain the overall prediction. The results presented in Table 2 suggest that using the LSTM, Seq2Seq, or transformer network with the InceptionResNet to incorporate inter-slice dependencies outperformed a simple averaging over slices. Specifically, coupling the Seq2Seq, transformer, and LSTM with the InceptionResNet improved the sensitivity of the predictive model from 65% to 77% and its specificity from 74% to 78%, 78%, and 83%, respectively, on the independent test set. Further, integrating the clinical variables and deep learning features of MRI in the framework improved the performance of the predictive model. Whereas the predictive models with only the

TABLE 1 Patient characteristics

Clinical features/outcome	Training set (99 patients and 116 lesions)	Test set (25 patients and 40 lesions)
Tumor size (Longest diameter)	Range: 0.4-7 cm Mean: 1.99 cm	Range: 0.6-6.6 cm Mean: 2.06 cm
Age	62 ± 15 years	63 ± 17 years
Gender		
Male	39 patients (39%)	11 patients (44%)
Female	60 patients (61%)	14 patients (56%)
Tumor location		
Supratentorium	87 lesions (75%)	29 lesions (72.5%)
Infratentorium	29 lesions (25%)	11 lesions (27.5%)
Histology		
Lung cancer	58 lesions (50%)	23 lesions (57.5%)
Breast cancer	26 lesions (22%)	9 lesions (22.5%)
Melanoma cancer	9 lesions (8%)	3 lesions (7.5%)
Colorectal cancer	7 lesions (6%)	0 lesions (0%)
RCC cancer	8 lesions (7%)	1 lesion (2.5%)
Other	8 lesions (7%)	4 lesions (10%)
Total dose (over 5 fractions)		
22.5 Gy	1 lesion (1%)	0 lesions (0%)
25 Gy	20 lesions (17%)	8 lesions (20%)
27.5 Gy	6 lesions (5%)	2 lesions (5%)
30 Gy	73 lesions (63%)	20 lesions (50%)
32.5 Gy	7 lesions (6%)	6 lesions (15%)
35 Gy	9 lesions (8%)	4 lesions (10%)
Previous WBRT		
Yes	45 lesions (39%)	9 lesions (22.5%)
No	71 lesions (61%)	31 lesions (77.5%)
Prior SRT/SRS		
Yes	1 lesion (1%)	0 lesions (0%)
No	115 lesions (99%)	40 lesions (100%)
Number of brain metastases		
One lesion	34 patients (34%)	9 patients (36%)
Two lesions	35 patients (35%)	7 patients (28%)
Three lesions	11 patients (11%)	4 patients (16%)
More than three lesions	19 patients (19%)	5 patients (20%)
Graded prognostic assessment (GPA)		
0.00-1.00	15 patients (15%)	3 patients (12%)
1.01-2.00	39 patients (39%)	14 patients (56%)
2.01-3.00	36 patients (36%)	3 patients (12%)
3.01-4.00	9 patients (9%)	5 patients (20%)
SRT outcome		
LC	70 lesions (60%)	23 lesions (57.5%)
LF	46 lesions (40%)	17 lesions (42.5%)

TABLE 2 Results of therapy outcome prediction in terms of accuracy, sensitivity, specificity, and AUC for different models

Network	Validation set				Independent test set			
	Acc. (%)	Sens. (%)	Spec. (%)	AUC	Acc. (%)	Sens. (%)	Spec. (%)	AUC
MLP/Clinical features	60	50	66.7	0.65	67.5	65	70	0.68
InceptionResNet	66.7	66.7	66.7	0.69	70	65	74	0.72
InceptionResNet + Seq2Seq	73.3	66.7	77.8	0.76	77.5	76.5	78.2	0.81
InceptionResNet + Transformer	73.3	66.7	77.8	0.75	77.5	76.5	78.2	0.81
InceptionResNet + LSTM	80	83.3	77.8	0.83	80	76.5	82.6	0.83
InceptionResNet + Seq2Seq with Clinical Feature Fusion	80	83.3	77.8	0.81	77.5	70.6	82.6	0.8
InceptionResNet + Transformer with Clinical Feature fusion	80	83.3	77.8	0.82	77.5	70.6	82.6	0.82
InceptionResNet + LSTM with Clinical Feature Fusion	86.7	83.3	88.9	0.88	82.5	76.5	87	0.86

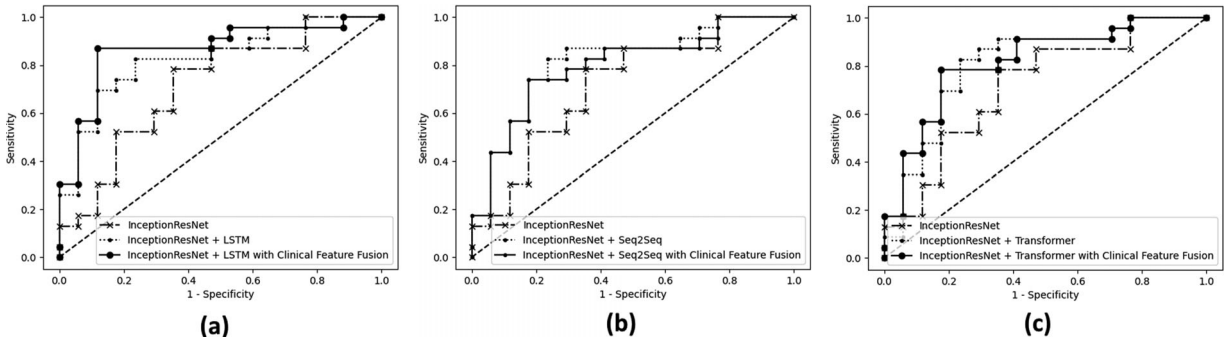

FIGURE 2 The ROC curves on the independent test set for different predictive models

clinical variables and with the MRI deep learning features solely demonstrated an outcome prediction accuracy of 68% and 78%-80%, respectively, the models with the combined deep learning and clinical features demonstrated an accuracy of 78%-83% on the test set. The Seq2Seq and transformer networks could not benefit from incorporating the clinical features that can be due to their overcomplex architecture for this application. The results of the ROC analysis (Figure 2) also support these observations and demonstrate the benefits of incorporating the 3D spatial dependencies in MRI deep learning features and fusing the clinical information to improve the performance of the outcome prediction system. The best results were obtained by the InceptionResNet + LSTM model with clinical feature fusion where the predictive model demonstrated an AUC of 0.86 on the independent test set, compared to an AUC of 0.72 and 0.8 for the InceptionResNet, and InceptionResNet + LSTM without the clinical features, respectively.

Figure 3 demonstrates the results of survival analysis on the two cohorts of patients in the test set stratified based on the outcome prediction at the baseline using different models. The Kaplan–Meier progression-free survival curves are presented for the patients in cohort 1 (with their all lesions having a predicted outcome of LC) and cohort 2 (with at least one lesion with a predicted outcome of LF). A log-rank test applied on the survival curves of the two cohorts for each model demonstrated no statistically significant difference for the InceptionResNet or the clinical model, an approaching significance ($p = 0.05$) for the InceptionResNet + LSTM model, and a significant difference for the InceptionResNet + LSTM model with clinical feature fusion.

Figure 4 depicts the visualization heatmaps associated with the T1w and T2-FLAIR images of four representative lesions obtained through the prediction difference analysis. The heatmaps demonstrate the impact level of different regions on MRI on the decision of the InceptionResNet in predicting the LC/LF outcome for each lesion. The visualization results imply that the tumor/lesion margin areas are particularly among the high-impact regions on both T1w and T2-FLAIR images with higher attention gained from the network for outcome prediction.

4 | DISCUSSION

In this study, the possibility of early prediction of local outcome for brain metastasis patients treated with SRT was investigated using deep learning of the treatment-planning MRI and clinical variables. A comprehensive outcome prediction framework was developed to derive

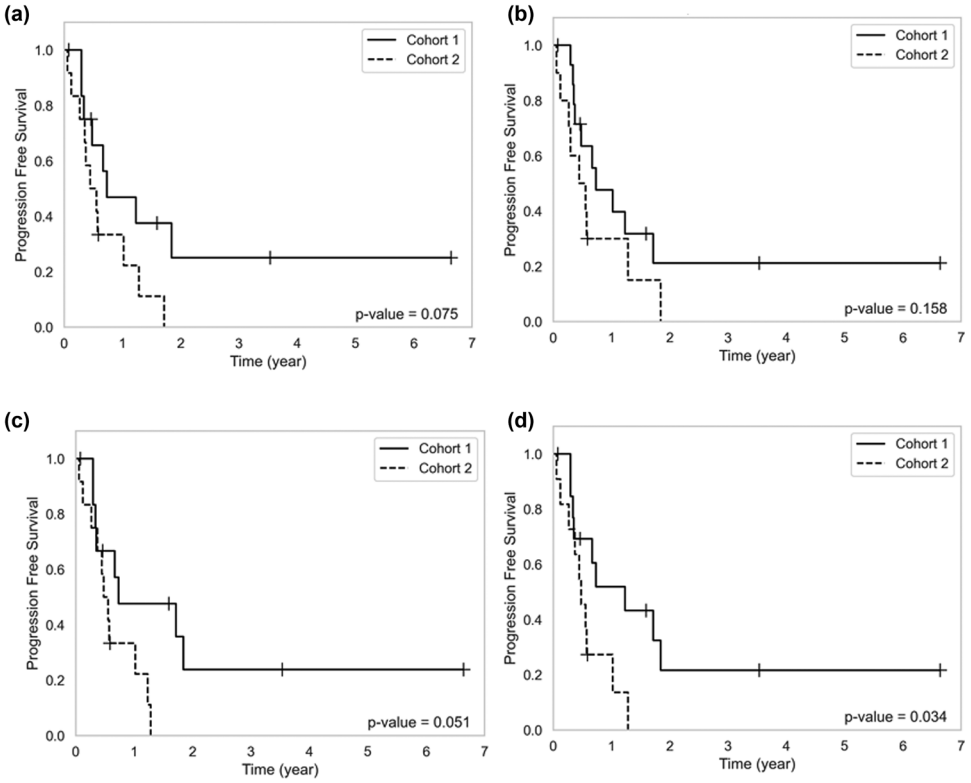

FIGURE 3 Kaplan–Meier progression-free survival curves for two cohorts of patients stratified at the baseline based on the outcome prediction by different models: (a) MLP with clinical features only, (b) InceptionResNet, (c) InceptionResNet + LSTM, and (d) InceptionResNet + LSTM with clinical feature fusion. Cohort 2 includes the patients in the independent test set who had at least one lesion with a predicted outcome of local failure, and cohort 1 includes all other patients in the independent test set.

optimal deep learning-based MRI features, integrate them with standard clinical variables, and incorporate 3D spatial dependencies within volumetric MRI of each lesion to predict the therapy outcome. To extract features from MRI, an InceptionResNet architecture was adapted, mainly because this architecture proved to excel at learning distinctive features from images in many tasks by applying residual connections and inception blocks. To incorporate spatial dependency between slices in each MRI volume, three different neural network architectures, namely, LSTM, Seq2Seq, and transformer networks were utilized and investigated. The results demonstrated a notable improvement in prediction performance of the model after integrating the recurrent or attention-based neural networks and clinical features. While using only the clinical variables or MRI features from single slices resulted in an AUC of 0.68 and 0.72, respectively, coupling the LSTM, Seq2Seq, and transformer network with the InceptionResNet improved the AUC to 0.83, 0.81, and 0.81, respectively. The clinical feature fusion with the MRI deep learning features improved the results further, with the best result of 83%, 77%, 87%, and 0.86 for the accuracy, sensitivity, specificity, and AUC obtained for the InceptionResNet + LSTM model with clinical feature fusion.

Results of risk stratification through survival analysis further highlighted the benefits of incorporating the 3D spatial dependencies of MRI features as well as the clinical feature fusion. Among different models explored, only the InceptionResNet + LSTM model with clinical feature fusion could stratify the patients into two cohorts with a statistically significant difference in progression-free survival, with the cohorts identified by the InceptionResNet + LSTM model without any clinical features approached significance.

A visualization module was integrated with the outcome prediction framework to provide insights on the contribution level of different areas of lesion on MRI to the network's decisions. The results highlighted the importance of the tumor/lesion margin areas in radiotherapy outcome prediction. This observation is in agreement with the findings of previous studies on MRI radiomics for treatment outcome prediction in brain metastasis.[25,45] Nests of tumor cells may exist for several millimeters outside the confines of the distinct metastatic brain lesion.[55] The information provided through the visualization modules, such as the one presented in this work, regarding the contribution of margin areas of a lesion to its predicted outcome, can potentially be beneficial during radiation treatment planning to reduce the chance of local failure in brain metastasis.[55]

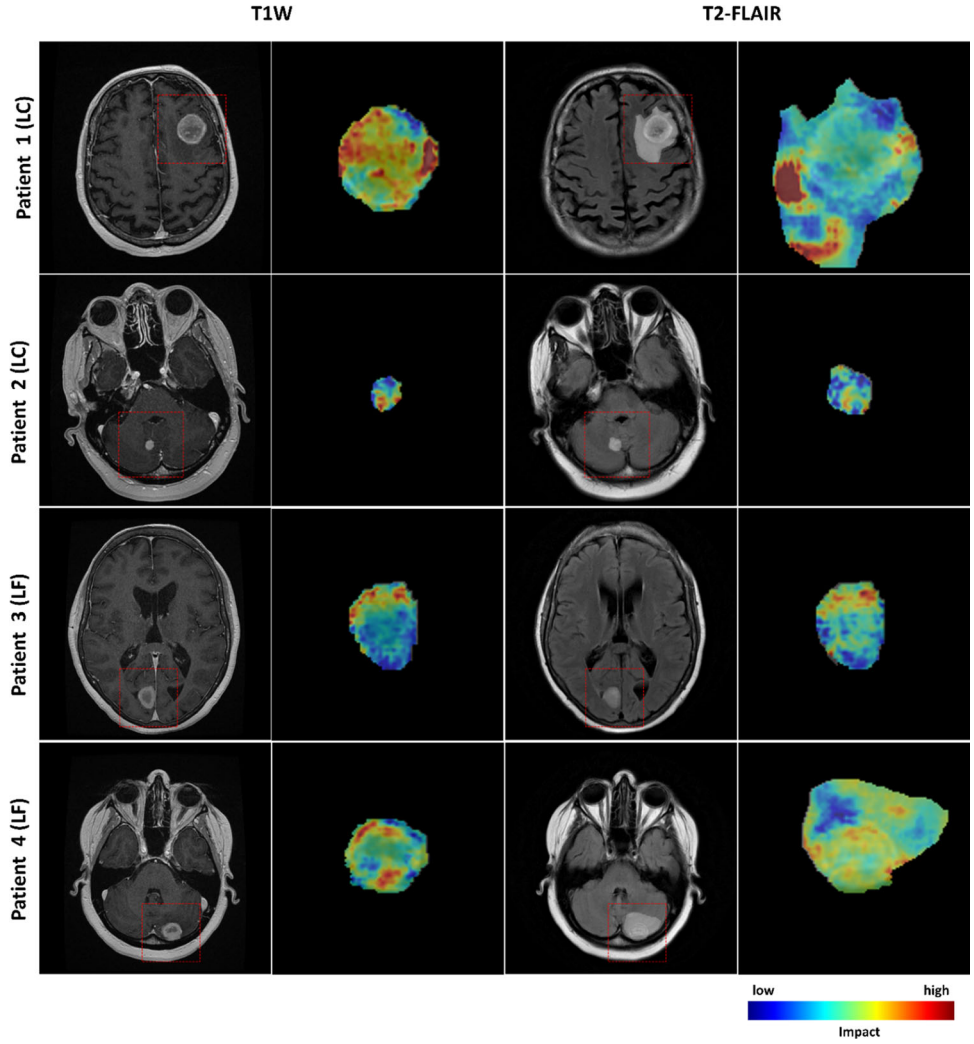

FIGURE 4 Treatment-planning MRI and deep learning visualization heatmaps acquired for four representative patients with metastatic lesions treated with SRT, two with an LC and the other two with an LF outcome. The heatmaps are overlaid on the ROIs input to the InceptionResNet and demonstrate the impact level of different areas on each image for the network's prediction, as calculated through a modified PDA. The network predicted the outcome of all four lesions correctly.

5 | CONCLUSION

The findings of this study show the possibility of early prediction of therapy outcome using a combination of quantitative MRI features and clinical attributes. This is in agreement with observations of previous studies that suggest the diagnostic and prognostic power of hand-crafted textural features in various imaging modalities such as CT,[56] MRI,[57] and ultrasound[58] in different cancer sites. The study here highlights the advantage of using deep learning architectures in combination with RNN for automated extraction of optimal quantitative features from volumetric MRI that can be effectively coupled with standard clinical variables for accurate radiotherapy outcome prediction. In conclusion, the promising results obtained in this study encourage future investigations on larger cohorts of patients. The results reported in this paper were obtained on an independent test set. However, to evaluate the efficacy and robustness of the framework in clinic more rigorously, further investigations are required preferably on multi-institutional data.

ACKNOWLEDGMENTS

This research was supported by the Natural Sciences and Engineering Research Council of Canada (Grant #: CRDPJ507521-16 and RGPIN-2016-06472), Lotte and John Hecht Memorial Foundation, and the Terry Fox Foundation (Grant #: 1083). Ali Sadeghi-Naini holds the York Research Chair in Quantitative Imaging and Smart Biomarkers, and an Early Researcher Award from the Ontario Ministry of Colleges and Universities.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Sacks P, Rahman M. Epidemiology of brain metastases. *Neurosurg Clin N Am*. 2020;31(4):481-488. https://doi.org/10.1016/j.nec.2020.06.001
2. Fox BD, Cheung VJ, Patel AJ, Suki D, Rao G. Epidemiology of metastatic brain tumors. *Neurosurg Clin N Am*. 2011;22(1):1-6. https://doi.org/10.1016/j.nec.2010.08.007
3. Carapella CM, Gorgoglione N, Oppido PA. The role of surgical resection in patients with brain metastases. *Curr Opin Oncol*. 2018;30(6):390-395. https://doi.org/10.1097/CCO.0000000000000484
4. Brown PD, Ahluwalia MS, Khan OH, Asher AL, Wefel JS, Gondi V. Whole-brain radiotherapy for brain metastases: evolution or revolution? *J Clin Oncol*. 2018;36(5):483-491. https://doi.org/10.1200/JCO.2017.75.9589
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. *J Am Med Assoc*. 2016;316(4):401-409. https://doi.org/10.1001/jama.2016.9839
6. Brown PD, Ballman KV, Cerhan JH, et al. Postoperative stereotactic radiosurgery compared with whole brain radiotherapy for resected metastatic brain disease (NCCTG N107C/CEC·3): a multicentre, randomised, controlled, phase 3 trial. *Lancet Oncol*. 2017;18(8):1049-1060. https://doi.org/10.1016/S1470-2045(17)30441-2
7. Venur VA, Karivedu V, Ahluwalia MS. Systemic therapy for brain metastases. *Handbook of Clinical Neurology*. 2018;149:137-153.
8. Vogenberg FR, Barash CI, Pursel M. Personalized medicine: part 1: evolution and development into theranostics. *Pharmacol Ther*. 2010;35(10):560-576.
9. Lin NU, Lee EQ, Aoyama H, et al. Response assessment criteria for brain metastases: proposal from the RANO group. *Lancet Oncol*. 2015;16(6):e270-e278. https://doi.org/10.1016/S1470-2045(15)70057-4
10. Wiggenraad R, Bos P, Verbeek-De Kanter A, et al. Pseudo-progression after stereotactic radiotherapy of brain metastases: lesion analysis using MRI cine-loops. *J Neurooncol*. 2014;119(2):437-443. https://doi.org/10.1007/s11060-014-1519-x
11. Sperduto PW, Yang TJ, Beal K, et al. Estimating survival in patients with lung cancer and brain metastases. *JAMA Oncol*. 2017;3(6):827. https://doi.org/10.1001/jamaoncol.2016.3834
12. Sperduto PW, Jiang W, Brown PD, 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*. 2017;99(4):812-816. https://doi.org/10.1016/j.ijrobp.2017.06.2454
13. Sperduto PW, Kased N, Roberge D, et al. Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases. *J Clin Oncol*. 2012;30(4):419-425. https://doi.org/10.1200/JCO.2011.38.0527
14. Saito EY, Viani GA, Ferrigno R, et al. Whole brain radiation therapy in management of brain metastasis: results and prognostic factors. *Radiat Oncol*. 2006;1:20. https://doi.org/10.1186/1748-717X-1-20
15. Nieder C, Mehta MP. Prognostic indices for brain metastases – usefulness and challenges. *Radiat Oncol*. 2009;4(1):10. https://doi.org/10.1186/1748-717X-4-10
16. Nieder C, Spanne O, Mehta MP, Grosu AL, Geinitz H. Presentation, patterns of care, and survival in patients with brain metastases. *Cancer*. 2011;117(11):2505-2512. https://doi.org/10.1002/cncr.25707
17. Soliman H, Das S, Larson DA, Sahgal A. Stereotactic radiosurgery (SRS) in the modern management of patients with brain metastases. *Oncotarget*. 2016;7(11):12318-12330. https://doi.org/10.18632/oncotarget.7131
18. Schag CC, Heinrich RL, Ganz PA. Karnofsky performance status revisited: reliability, validity, and guidelines. *J Clin Oncol*. 1984;2(3):187-193. https://doi.org/10.1200/JCO.1984.2.3.187
19. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. 2018;2(1):36. https://doi.org/10.1186/s41747-018-0068-z
20. Kapadia A, Mehrabian H, Conklin J, et al. Temporal evolution of perfusion parameters in brain metastases treated with stereotactic radiosurgery: comparison of intravoxel incoherent motion and dynamic contrast enhanced MRI. *J Neurooncol*. 2017;135(1):119-127. https://doi.org/10.1007/s11060-017-2556-z
21. Desmond KL, Mehrabian H, Chavez S, et al. Chemical exchange saturation transfer for predicting response to stereotactic radiosurgery in human brain metastasis. *Magn Reson Med*. 2016;78(3):1110-1120. https://doi.org/10.1002/mrm.26470
22. O'connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2017;14(3):169-186. https://doi.org/10.1038/nrclinonc.2016.162
23. 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. *Annu Int Conf IEEE Eng Med Biol Soc*. 2019;2019:1022-1025. https://doi.org/10.1109/EMBC.2019.8856558
24. Liao C-Y, Lee C-C, Yang H-C, 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/cancers13164030
25. Mouraviev A, Detsky J, Sahgal A, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. *Neuro Oncol*. 2020;22(6):797-805. https://doi.org/10.1093/neuonc/noaa007
26. Aerts H, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. 2014;5:4006. https://doi.org/10.1038/ncomms5006
27. Lin Y-Z, Nie Z-H, Ma H-W. Structural damage detection with automatic feature-extraction through deep learning. *Comput Civ Infrastruct Eng*. 2017;2(12):1025-1046. https://doi.org/10.1111/mice.12313
28. Zhao W, Du S. Spectral–spatial feature extraction for hyperspectral image classification: a dimension reduction and deep learning approach. *IEEE Trans Geosci Remote Sens*. 2016;54(8):4544-4554. https://doi.org/10.1109/TGRS.2016.2543748
29. You C, Zhao R, Liu F, et al. Class-aware generative adversarial transformers for medical image segmentation. *arXiv*:2201.10737v2, 2022. https://doi.org/10.48550/arXiv.2201.10737
30. Yang L, Ghosh RP, Franklin JM, et al. NuSeT: a deep learning tool for reliably separating and analyzing crowded cells. *PLoS Comput Biol*. 2020;16(9):e1008193. https://doi.org/10.1371/journal.pcbi.1008193
31. You C, Zhao R, Staib L, Duncan JS. Momentum contrastive voxel-wise representation learning for semi-supervised volumetric medical image segmentation. *arXiv*:2105.07059v4. 2021. https://doi.org/10.48550/arXiv.2105.07059
32. You C, Zhou Y, Zhao R, Staib L, Duncan JS. SimCVD: simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation. *arXiv*:2108.06227v4. 2021. https://doi.org/10.48550/arXiv.2108.06227
33. Lagree A, Mohebpour M, Meti N, et al. A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks. *Sci Rep*. 2021;11(1):8025. https://doi.org/10.1038/s41598-021-87496-1
34. Jalalifar A, Soliman H, Sahgal A, Sadeghi-Naini A. A cascaded deep-learning framework for segmentation of metastatic brain tumors before and after stereotactic radiation therapy. *Annu Int Conf IEEE Eng Med Biol Soc*. 2020;2020:1063-1066. https://doi.org/10.1109/EMBC44109.2020.9175489

35. 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):132-160. https://doi.org/10.1109/MSP.2019.2900993
36. Wetzer E, Gay J, Harlin H, Lindblad J, Sladoje N. When texture matters: texture-focused CNNs outperform general data augmentation and pretraining in oral cancer detection. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*; April 2020: 517-521. https://doi.org/10.1109/ISBI45749.2020.9098424
37. Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. 2017;19:221-248. https://doi.org/10.1146/annurev-bioeng-071516-044442
38. Lyu Q, You C, Shan H, Wang G. Super-resolution MRI through deep learning. *arXiv*:1810.06776v1. 2018. https://doi.org/10.48550/arXiv.1810.06776
39. Diamant A, Chatterjee A, Vallières M, Shenouda G, Seuntjens J. Deep learning in head & neck cancer outcome prediction. *Sci Rep*. 2019;9(1):2764. https://doi.org/10.1038/s41598-019-39206-1
40. Cho SJ, Sunwoo L, Baik SH, Bae YJ, Choi BS, Kim JH. Brain metastasis detection using machine learning: a systematic review and meta-analysis. *Neuro Oncol*. 2021;23(2):214-225. https://doi.org/10.1093/neuonc/noaa232
41. Rumelhart DE, >Hinton GE, Williams RJ. Learning internal representations by error propagation. In: Rumelhart DE, McClelland JL, eds. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*. MIT Press; 1987:318-362.
42. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
43. Truong MT, St Clair EG, Donahue BR, et al. Results of surgical resection for progression of brain metastases previously treated by gamma knife radiosurgery. *Neurosurgery*. 2006;59(1):86-97. https://doi.org/10.1227/01.NEU.0000219858.80351.38
44. Sneed PK, Mendez J, Vemer-Van Den Hoek JGM, et al. Adverse radiation effect after stereotactic radiosurgery for brain metastases: incidence, time course, and risk factors. *J Neurosurg*. 2015;123(2):373-386. https://doi.org/10.3171/2014.10.JNS141610
45. Karami E, Soliman H, Ruschin M, et al. Quantitative MRI biomarkers of stereotactic radiotherapy outcome in brain metastasis. *Sci Rep*. 2019;9:19830. https://doi.org/10.1038/s41598-019-56185-5
46. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging*. 2015;34(10):1993-2024. https://doi.org/10.1109/TMI.2014.2377694
47. Wilson DR, Martinez TR. The general inefficiency of batch training for gradient descent learning. *Neural Netw*. 2003;16(10):1429-1451. https://doi.org/10.1016/S0893-6080(03)00138-2
48. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016.
49. Bengio Y, Louradour J, Collobert R, Weston J. Curriculum learning. In: *Proceedings of the 26th Annual International Conference on Machine Learning - ICML '09*. 2009:1-8, https://doi.org/10.1145/1553374.1553380
50. Chollet F, et al. Keras, *GitHub*, 2015. https://github.com/fchollet/keras
51. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv*:1603.04467v2. 2016. http://arxiv.org/abs/1603.04467
52. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.
53. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. https://doi.org/10.1109/MCSE.2007.55
54. Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: *European Conference on Computer Vision (ECCV)*. 2014:818-833. https://doi.org/10.1007/978-3-319-10590-1_53
55. Hardesty DA, Nakaji P. The current and future treatment of brain metastases. *Front Surg*. 2016;3:30. https://doi.org/10.3389/fsurg.2016.00030
56. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol*. 2015;114(3):345-350. https://doi.org/10.1016/j.radonc.2015.02.015
57. Kniep HC, Madesta F, Schneider T, et al. Radiomics of brain MRI: utility in prediction of metastatic tumor type. *Radiology*. 2019;290(2):479-487. https://doi.org/10.1148/radiol.2018180946
58. Sadeghi-Naini A, Papanicolau N, Falou O, et al. Low-frequency quantitative ultrasound imaging of cell death in vivo. *Med Phys*. 2013;40(8):082901. https://doi.org/10.1118/1.4812683

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Jalalifar SA, Soliman H, Sahgal A, Sadeghi-Naini A. Predicting the outcome of radiotherapy in brain metastasis by integrating the clinical and MRI-based deep learning features. *Med. Phys*. 2022;49:7167–7178. https://doi.org/10.1002/mp.15814