Predicting Local Failure after Stereotactic Radiation Therapy in Brain Metastasis using Quantitative CT and Machine Learning*

Majid Jaberipour, Arjun Sahgal, Hany Soliman and Ali Sadeghi-Naini, *Senior Member, IEEE*

***Abstract*— Despite recent advances in cancer treatment, the prognosis of patients diagnosed with brain metastasis is still poor. The median survival is limited to months even for patients undergoing treatment. Radiation therapy is a main component of treatment for brain metastasis. However, radiotherapy cannot control local progression in up to 20% of the metastatic brain tumours. An early prediction of radiotherapy outcome for individual patients could facilitate therapy adjustments to improve its efficacy. This study investigated the potential of quantitative CT biomarkers in conjunction with machine learning methods to predict local failure after radiotherapy in brain metastasis. Volumetric CT images were acquired for radiation treatment planning from 120 patients undergoing stereotactic radiotherapy. Quantitative features characterizing the morphology and texture were extracted from different regions of each lesion. A feature reduction/selection framework was adapted to define a quantitative CT biomarker of radiotherapy outcome. Different machine learning methods were applied and evaluated to predict the local failure outcome at pre-treatment. The optimum biomarker consisting of two features in conjunction with an AdaBoost with decision tree could predict the local failure outcome with 71% accuracy on an independent test set (20 patients, 31 lesions). This study is a step forward towards prediction of radiotherapy outcome in brain metastasis using quantitative imaging and machine learning.**

*This Research was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada, Terry Fox Foundation, and the Lotte and John Hecht Memorial Foundation.

M. Jaberipour is with Physical Sciences Platform, Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; also with Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, ON, Canada (e-mail: majid.jaberipour@sunnybrook.ca).

A. Sahgal and H. Soliman are with the Department of Radiation Oncology, Sunnybrook Health Sciences Centre; also with the Department of Radiation Oncology, University of Toronto, Toronto, ON, Canada (e-mail: arjun.sahgal@sunnybrook.ca; hany.soliman@sunnybrook.ca).

A. Sadeghi-Naini is with the Department of Electrical Engineering and Computer Science, Lassonde School of Engineering, York University, Toronto, ON, Canada; also with the Department of Radiation Oncology and Physical Sciences Platform, Odette Cancer Centre and Sunnybrook Research Institute, Sunnybrook Health Sciences Centre; also with the Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada (e-mail: asn@yorku.ca; phone: 416-736-2100 x20590).

I. INTRODUCTION

A metastatic brain tumour, the most common neoplasms of the central nervous system, is formed when cancer cells located elsewhere in the body travel to the brain. These tumours are usually rounded, firm, and well-demarcated. The prognosis of brain metastasis patients is poor [1]–[4]: their median survival is limited to months even for patients undergoing treatment. Radiation therapy is a main component of treatment for patients with brain metastasis. Stereotactic radiation therapy (SRT) is a focused radiation technique that is frequently administrated for the management of patients with up to 10 brain metastases [5]. However, up to 20% of metastatic brain tumours progress locally after stereotactic radiation treatment [6]. Predicting the outcome of standard radiation therapy for individual patients at pre-treatment could facilitate treatment adjustments or salvage therapies for non-responding patients that are expected to improve overall survival and quality of life. In this context, quantitative imaging with radiomics has been demonstrated to be a valuable methodology to increase precision in diagnosis or to predict treatment response in cancer patients [7]–[16]. Radiomics is defined as extracting high dimensional data from medical images and subsequent mining of these data for improved decision support. Although radiomics is a natural extension of computer-aided diagnosis (CAD) systems, it is significantly different from them. Unlike CAD systems, which are directed toward delivering a single answer (i.e., presence of a lesion or cancer), radiomics is explicitly a process designed to extract a large number of quantitative features from digital images, representing these data in organized data structures, and subsequently mine the data for hypothesis generation, testing, or both [17], [18].

Computed tomography (CT) and magnetic resonance (MR) images play significant roles in the diagnosis and management of metastatic brain tumours. Magnetic resonance imaging (MRI) is the standard imaging modality to detect brain metastasis. CT can detect some of the tumours (less than MRI) and is helpful in identifying skull metastases or hemorrhage within metastases [4]. Furthermore, CT is used for radiation therapy simulation in treatment planning to derive the radiation dose maps. The quantitative features extracted from treatment-planning CT can potentially be used to predict radiation therapy outcome in brain metastasis and are expected to complement quantitative MRI features in this application.

978-1-7281-1990-8/20/$31.00 ©2020 IEEE

Recently, Karimi et al. presented an MRI-based radiomic framework to predict the local control/local failure (LC/LF) outcome in patients with brain metastasis treated with hypo-fractionated SRT [19]. The quantitative MRI features were extracted from the images acquired at the baseline and the first follow up after SRT. The framework could predict the outcome with a cross-validated accuracy of 80%. To our knowledge, no previous study explored the potential of non-contrast CT through radiomic analysis for predicting the outcome in patients with brain metastasis treated with SRT. This study investigated the performance of non-contrast quantitative CT imaging in conjunction with machine learning methods to predict LC/LF outcome in brain metastasis patients undergoing hypo-fractionated SRT before the start of treatment.

II. Materials and Methods

A. Data Acquisition

This study was conducted in accordance with institutional research ethics board approval from Sunnybrook Health Sciences Centre (SHSC), Toronto Canada. Imaging and clinical data were collected from 120 patients (169 lesions) diagnoses with brain metastasis and treated with hypo-fractionated SRT. Imaging data included treatment-planning CT (non-contrast) and MRI (post-contrast T1w, and T2-FLAIR), as well as the follow-up MRI (on a 2-3 month schedule after the SRT) to determine the therapy outcomes. The RANO-BM criteria were used to determine the LC/LF outcome (ground truth) for each tumour [20]. The LC/LF outcome was defined as the local outcome identified in the last patient follow up within six years after the SRT. Out of the 120 patients, 83 and 37 patients had single and multiple lesions, respectively. Out of the 169 tumours, 106 tumours demonstrated an LC outcome whereas 63 tumours progressed locally after SRT (LF). Table I summarizes the characteristics of the patients.

TABLE I: PATIENT CHARACTERISTICS.

Age (years)	Range: 21-92;(years) Mean:63 years
Gender	Men: 48 (40%) Female: 72 (60%)
Histology	Lung cancer: 61 (51%) Breast cancer: 24 (20%) Melanoma cancer: 10 (8%) Colorectal cancer: 9 (8%) Other: 16 (13%)

B. Training and Test Sets

The patients were randomly divided into two independent training and test sets based on the stratified random sampling method [21]. The whole data set was first divided into five disjoint groups "patients with single tumour demonstrated an LC outcome", "patients with single tumour demonstrated an LF outcome", "patients with multiple tumours, all with an LC outcome", "patients with multiple tumours, all with an LF outcome", and finally "patients with multiple tumours, with both LC and LF outcomes". The test set was randomly selected from these disjoint groups with the same proportion presented in the whole data set. The training and test sets consisted of 100 (138) and 20 (31) patients (lesions), respectively. The training set was used for feature reduction/selection and creating the machine learning models (described below), whereas the test set was used for an independent evaluation of the trained models.

C. Quantitative CT Feature Extraction and Reduction

Treatment-planning T1w (post-contrast) and T2-FLAIR images were used to generate the tumour and lesion (tumour + edema) masks on MR images. To generate the tumour and lesion masks for CT images, the CT image was registered on the corresponding MR image using an affine registration with mutual Mattas information (MI) as the similarity metric. Then, the MRI masks were warped on the original CT image using the inverse of the registration transformation. In addition to the tumour and lesion masks, lesion-margin and tumour-margin masks were created with a 5 mm margin size for each lesion using morphological image analysis.

Prior to feature extraction, the image intensity was normalized in each image to have zero mean and unit variance. In addition to normalization, all images were resampled with an isotropic voxel size of $1 \times 1 \times 1\ mm^3$ to ensure a uniform scale in all directions when extracting the 3D features. Quantitative CT features were extracted from the tumour, lesion, tumour-margin, and the lesion-margin areas using the Pyradiomics package in python. A total of 430 features including morphological features, first-order statistics, and second-order texture features were derived. The morphological features were extracted from the region masks. The first-order statistics were extracted from the intensity histogram and the other texture features were derived based on the following statistical methods: gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM) and neighborhood gray-tone difference matrix (NGTDM). The extracted features were first assessed using a Pearson correlation analysis to exclude the highly correlated features with an $R^2 > 0.8$. The correlation analysis reduced the number of features from 430 to 82.

D. Feature Selection Method for Imbalanced Data Set

In this subsection, we describe the methodology that was applied for feature selection on our imbalanced data set. The feature selection framework had two phases. In the first phase, sequential forward floating selection (SFFS) was used with a stratified *k*-fold ($k = 10$) cross validation (on patient level) to find the best feature set (up to 5 features) associated with each random balanced subset. Each balanced subset was generated using a random downsampling on the majority class (LC). This phase was run for *m* times (Figure 1). In the second phase, one of the feature sets obtained in phase I was selected as the final feature set (optimum biomarker; Figure 2). For this purpose, several random balanced subsets (*n* subsets) were generated and for each balanced subset, the cross-validated score of all unique feature sets obtained in phase I was calculated. The cross-validated scores were

subsequently averaged over all random balanced subsets and the feature set with the highest averaged score was selected as the optimum biomarker. The receiver operating characteristic (ROC) area under the curve (AUC) was used as the feature selection score in both phases of the framework.

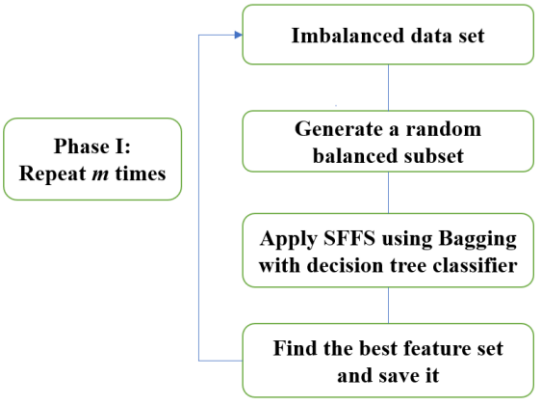

Figure 1: Feature selection-phase I.

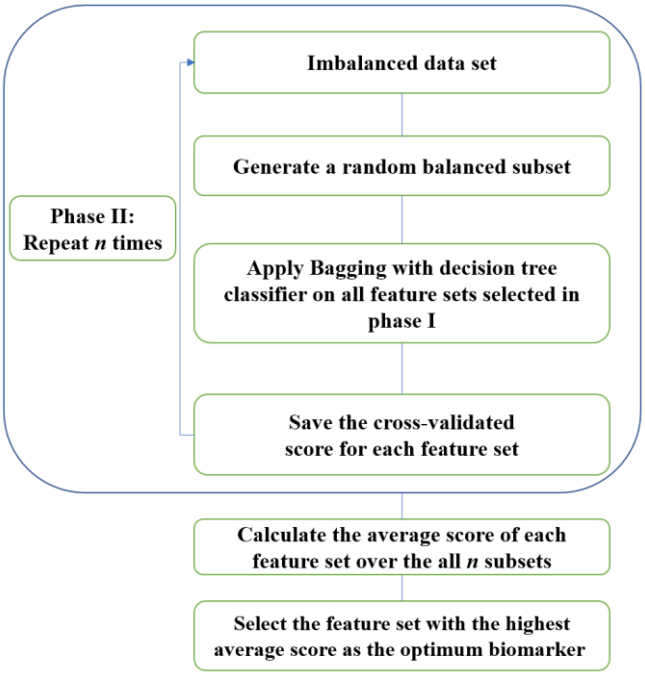

Figure 2: Feature selection-phase II.

E. Outcome Prediction Framework

Three machine learning algorithms (bootstrap aggregating (Bagging) with decision tree; *k*-nearest neighbor algorithm (*k*-NN), adaptive boosting (AdaBoost) with decision tree were utilized for local control and failure predictive-model construction with two mentioned features. Max voting ensemble [22] is a simple method that can be adapted for classification of an imbalanced data set very well. We benefited from this method to predict the LC/LF outcome of the lesions in the test set. Following this strategy, *p* random balanced sets were generated and applied to train different prediction models using a similar classifier. For each lesion in the test set, a max-voting was applied over the *p* outcomes predicted by these models to obtain the final predicted outcome for that lesion. The outcome prediction framework has been presented in Figure 3.

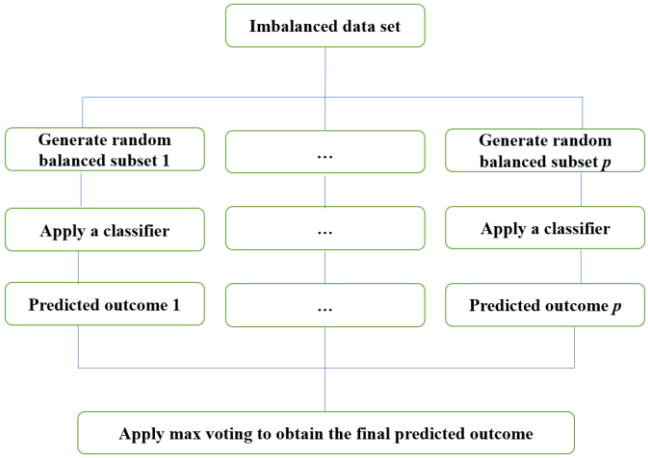

Figure 3: Outcome prediction framework.

III. Results

The feature selection framework was applied to the 82 quantitative CT features remained after feature reduction. The optimum biomarker consisted of two features including the "Shaped-3D-Elongation" and "GLDM-3D-Dependence-Nonuniformity-Normalized" from the tumour region. The *m*, *n*, and *p* parameters of the feature selection and outcome prediction frameworks were set to 501. Three different classifiers were investigated separately to predict the LC/LF outcome of each lesion. For each classifier, a grid search in conjunction with a stratified 10-fold cross-validation was used to tune parameters of the models. The results of outcome prediction on the independent test set have been presented in Table II for different classifiers. Obtained results demonstrated that the AdaBoost with decision tree outperformed the other classifiers in predicting the LC/LF outcome of the lesions using the optimum quantitative CT biomarker (consisting of two features). This classifier could predict the LC/LF outcome of lesions treated with SRT with a sensitivity, specificity, and accuracy of 76.9%, 66.7%, and 71.0%, respectively. Sensitivity (specificity) in this study refers to the proportion of the lesions with an LF (LC) outcome that was predicted to have an LF (LC) outcome by the model.

TABLE II: RESULTS OF OUTCOME PREDICTION WITH DIFFERENT CLASSIFIERS

Classifier	Accuracy	Specificity	Sensitivity	AUC
AdaBoost with decision tree	71.0%	66.7%	76.9%	0.72
Bagging with decision tree	64.5%	66.7%	61.5%	0.64
k-NN (*k* =5)	64.5%	61.1%	69.2%	0.65

IV. Discussion

This study demonstrated a relatively good potential of non-contrast quantitative CT in conjunction with machine learning to differentiate local control and failure outcome in metastatic brain tumours treated with SRT at pre-treatment. The optimum quantitative CT biomarker obtained through a multi-phase feature reduction/selection framework consisted of one morphological and one texture feature extracted from the tumour region. The best machine learning model adapted could predict the LC/LF outcome with an accuracy of 71% and AUC of 0.72 on an independent test set. The results imply that non-contrast CT can potentially contribute to a multi-modal quantitative imaging framework (CT/MRI) for *a priori* outcome prediction in brain metastasis treated with SRT. As CT and MRI are acquired as part of the standard of care in brain metastasis for radiation treatment planning, the complementary information they may provide can potentially be utilized with minimal overhead to optimize robust machine learning models to predict LC/LF outcome with high sensitivity and specificity. Investigating such multi-modal quantitative imaging strategy has been planned as a future direction of this study.

References

[1] I. T. Gavrilovic and J. B. Posner, "Brain metastases: Epidemiology and pathophysiology," *J. Neurooncol.*, vol. 75, no. 1, pp. 5–14, 2005.

[2] K. J. Stelzer, "Epidemiology and prognosis of brain metastases," *Surg. Neurol. Int.*, vol. 4, no. Suppl 4, pp. S192-202, 2013.

[3] R. Soffietti *et al.*, "EFNS Guidelines on diagnosis and treatment of brain metastases: Report of an EFNS Task Force," *Eur. J. Neurol.*, vol. 13, no. 7, pp. 674–681, 2006.

[4] E. C. A. Kaal, M. J. B. Taphoorn, and C. J. Vecht, "Symptomatic management and imaging of brain metastases," *J. Neurooncol.*, vol. 75, no. 1. pp. 15–20, 2005.

[5] M. Yamamoto *et al.*, "Stereotactic radiosurgery for patients with multiple brain metastases (JLGK0901): a multi-institutional prospective observational study," *Lancet Oncol.*, vol. 15, no. 4, pp. 387–95, 2014.

[6] A. Nagai, Y. Shibamoto, M. Yoshida, K. Wakamatsu, and Y. Kikuchi, "Treatment of Single or Multiple Brain Metastases by Hypofractionated Stereotactic Radiotherapy Using Helical Tomotherapy," *Int. J. Mol. Sci.*, vol. 15, no. 4, pp. 6910–6924, 2014.

[7] P. Grossmann *et al.*, "Defining the biological basis of radiomic phenotypes in lung cancer," *Elife*, vol. 6, pp. 1–22, 2017.

[8] W. T. Tran *et al.*, "Predicting breast cancer response to neoadjuvant chemotherapy using pretreatment diffuse optical spectroscopic texture analysis," *Br. J. Cancer*, vol. 116, no. 10, pp. 1329–1339, 2017.

[9] A. Sadeghi-Naini *et al.*, "Breast-Lesion Characterization using Textural Features of Quantitative Ultrasound Parametric Maps," *Sci. Rep.*, vol. 7, p. 13638, 2017.

[10] H. Tadayyon, A. Sadeghi-Naini, and G. J. Czarnota, "Noninvasive characterization of locally advanced breast cancer using textural analysis of quantitative ultrasound parametric images," *Transl. Oncol.*, vol. 7, no. 6, pp. 759–767, 2014.

[11] H. Tadayyon *et al.*, "A priori prediction of neoadjuvant chemotherapy response and survival in breast cancer patients using quantitative ultrasound," *Sci. Rep.*, vol. 7, p. 45733, 2017.

[12] A. Sadeghi-Naini *et al.*, "Imaging innovations for cancer therapy response monitoring," *Imaging Med.*, vol. 4, no. 3, pp. 311–327, Jun. 2012.

[13] M. J. Gangeh, A. Sadeghi-Naini, M. Diu, H. Tadayyon, M. S. Kamel, and G. J. Czarnota, "Categorizing Extent of Tumor Cell Death Response to Cancer Therapy Using Quantitative Ultrasound Spectroscopy and Maximum Mean Discrepancy," *IEEE Trans. Med. Imaging*, vol. 33, no. 6, pp. 1390–1400, 2014.

[14] H. Tadayyon *et al.*, "Quantitative ultrasound assessment of breast tumor response to chemotherapy using a multi-parameter approach," *Oncotarget*, vol. 7, no. 29, pp. 45094–45111, 2016.

[15] M. J. Gangeh, A. Sadeghi-Naini, M. S. Kamel, and G. J. Czarnota, "Assessment of cancer therapy effects using texton-based characterization of quantitative ultrasound parametric images," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*, 2013, pp. 1372–1375.

[16] H. Tadayyon *et al.*, "Quantification of Ultrasonic Scattering Properties of In Vivo Tumor Cell Death in Mouse Models of Breast Cancer," *Transl. Oncol.*, vol. 8, no. 6, pp. 463–73, 2015.

[17] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.

[18] R. T. H. M. Larue, G. Defraene, D. De Ruysscher, P. Lambin, and W. Van Elmpt, "Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures," *Br. J. Radiol.*, vol. 90, no. 1070. pp. 1–10, 2017.

[19] E. Karami *et al.*, "Quantitative MRI Biomarkers of Stereotactic Radiotherapy Outcome in Brain Metastasis.," *Sci. Rep.*, vol. 9, p. 19830, 2019.

[20] N. U. Lin *et al.*, "Response assessment criteria for brain metastases: proposal from the RANO group," *Lancet Oncol.*, vol. 16, no. 6, pp. e270–e278, 2015.

[21] C.-E. Särndal, *Stratified Sampling*. Springer, 2003.

[22] V. Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, P. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, A. and and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, and E. Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, "Scikit-learn: Machine Learning in python," *J. Mach. Learn. Res.*, vol. 2825--2830, no. 12, pp. 2825–2830, 2011.