IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 8, 2019, accepted May 22, 2019, date of publication June 3, 2019, date of current version June 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920396

Predicting the Gene Status and Survival Outcome of Lower Grade Glioma Patients With Multimodal MRI Features

AHMAD CHADDAD[1], CHRISTIAN DESROSIERS[2], BASSAM ABDULKARIM[1], AND TAMIM NIAZI[1]

[1]Division of Radiation Oncology, Department of Oncology, McGill University, Montreal, QC H3A 0G4, Canada
[2]Ecole de Technologie Supérieure, Montreal, QC H3C 1K3, Canada

Corresponding author: Ahmad Chaddad (ahmad.chaddad@mail.mcgill.ca)

This work was supported by the McGill University.

ABSTRACT We propose a novel class of multimodal image features based on the joint intensity matrix (JIM) to model fine-grained texture signatures in the radiomic analysis of lower grade glioma (LGG) tumors. Experiments use expanded JIM features to predict the genetic status and the survival outcome of LGG patients with preoperative T1-weighted, T1-weighted post-contrast, fluid attenuation inversion recovery (FLAIR), and T2-weighted MR images from The Cancer Imaging Archive (n=107). Texture features were extracted from regions of interest labeled by a radiation oncologist and summarized by 19 parameters. These parameters are then used to contrast mutant and wild gene type groups (i.e., IDH1, ATRX, TP53, and 1p/19q codeletion) via the Wilcoxon test, and to compare short and long survival patient groups with the Kaplan-Meier estimator. Random forest (RF) classification is employed to predict gene status (i.e., mutation or wild) and survival outcome (i.e., short or long survival), as well as to identify highly group-informative features. A subset of JIM features show statistically significant relationships with LGG gene status (i.e., in IDH1 and ATRX, with corrected $p < 0.05$) and survival outcome (p=0.0001, HR=0.09, CI=0.03-0.3). A maximum classification AUC of 78.59% was obtained for predicting IDH1 status from combined JIM and GLCM features. Classification combining all features (i.e., volume, JIM, and GLCM) results in an AUC value of 86.79% (corrected p=0.04) in predicting short and long LGG patient survival outcomes, where JIM features are generally the most informative predictors.

INDEX TERMS IDH1, radiomics, LGG, survival.

I. INTRODUCTION

Gliomas form the most common type of primary brain tumor in adults, accounting for 30% of all central nervous system tumors and about 80% of all malignant brain tumors [1], [2]. According to the World Health Organization (WHO 2016), gliomas can be classified from their molecular parameters in addition to histological properties into four grades (I, II, III or IV), grade I corresponding to non-invasive tumors, grade II/III to low/intermediate-grade gliomas, and grade IV to aggressive malignant tumors called glioblastoma multiforme (GBM) [3]. Combined grade II and III tumors are often referred to as lower-grade gliomas (LGG) [4]. LGGs typically occur in the cerebral hemispheres of young adults (between 20 and 50 years of age) and often evolve into more aggressive tumors like GBMs. Even with treatment including surgical resection and radiation therapy, the long-term outcome for this type of tumor remains poor [5].

Over the years, various studies have investigated the relationship between the genotype of LGG tumor cells and outcome. These studies have identified several genes associated with the development of gliomas, for example, the isocitrate dehydrogenase (IDH) gene [6]. It was shown that patients with the IDH1 mutation have a better prognosis compared to the patients with wild-type [7]. Since the ratio of patients with the IDH1 mutation is much greater for LGG than GBM (i.e., 70% compared to 10%), this mutation is considered a good genetic biomarker to predict the survival outcome of glioma patients [8], [9]. Moreover, LGG patients with chromosome 1p/19q co-deletion typically exhibit a longer overall

The associate editor coordinating the review of this manuscript and approving it for publication was Shubhajit Roy Chowdhury.

survival and a better response to chemotherapy compared to those with non-co-deleted counterparts [10]. TP53 mutation has also been associated with shorter survival in LGG patients [11]. Although genetic markers have proven useful for prognosis in patients with gliomas, their use in practice is impeded by important drawbacks. Thus, they require an invasive procedure to acquire physical samples from the tumor. The prognostic value of such genetic markers is also affected by the cellular heterogeneity occurring within (intra-tumor heterogeneity) and between (inter-tumor heterogeneity) tumors [12].

Image analysis techniques using MRI, CT or PET scans have recently emerged as a non-invasive alternative to derive prognostic biomarkers, detection and diagnosis for brain cancer [13]–[16]. Multiparametric MRI (e.g., T1-weighted, T2-weighted, post contrast T1 weighted, FLAIR, etc is widely used to identify brain tumors and monitor treatment response. Recent studies have also demonstrated the usefulness of image features encoding tumor shape and texture to quantify the phenotype of GBM tumors [17]–[21] and predict the survival outcome of GBM patients [22], [23]. In the case of low-grade tumors, the work in [4] showed that features extracted from the smooth non-enhancing margin of tumors could help detect molecular aberrations like IDH1 mutation and 1p/19q codeletion, or predict the histological grade and progression of these tumors.

Most image-based techniques for automated prognosis in brain tumor patients employ a radiomic analysis in which a wide range of features are computed from segmented tumor regions and used as input to a classifier. Radiomic analysis approaches typically consider available image modalities independently while extracting features and, therefore, do not make full use of information encoding relationships between these modalities. In [21], Li *et al*. proposed to learn image features from multi-sequence MRI data using a deep convolutional neural network (CNN), for the prediction of IDH1 mutations in LGG patients. Although CNNs enable the joint analysis of multiple image modalities, they typically require large training sets which are not available for survival prediction.

Recently, we have proposed multimodal radiomic features based on joint intensity matrices (JIMs) to predict the survival outcome of GBM patients [24]. JIMs extend the well-known grey-level co-occurrence matrix (GLCM) textures in single images by considering the relationship between intensity values across multi-parametric images. While our previous study showed JIMs to provide a more accurate prediction of GBM survival than single-modality radiomic features, this study also had limitations which required further investigation. First, the link between radiomic features, computed from different modalities, and gene mutations was not explored. Unlike GBM tumors, for which genomic markers like gene alterations are poorly predictive of survival, LGG patients with different gene status exhibit significantly different survival profiles. Additionally, in our previous study, spatial information was not fully exploited when computing texture features across diverse image modalities. Thus, we only considered the joint distribution of intensities in spatially-aligned voxels. In this work, we hypothesize that spatial relationship (i.e., relative position of voxels) encodes valuable information on tumor texture inside an image, but also across images of different modalities. This modification is also essential for this work on LGG, where regions of interest (ROIs) may not have corresponding voxels in different modalities.

The contributions of this work are as follows:

- We extend the JIM model of [24] to consider the relative position of voxels in multi-modal images. This alleviates the problem of non-corresponding ROI voxels across modalities, and allows capturing spatial information in multi-modal textures;
- While JIMs have been used to predict survival in GBM patients, this study is the first to evaluate their usefulness for lower grade gliomas (LGG) tumors;
- The current study proposes a comprehensive analysis, based on a cohort of 107 TCIA subjects, evaluating the discriminative power of JIMs to predict LGG patient survival, *as well as* to predict the gene status of these patients.

The rest of this paper is organized as follows. Section II describes the data used in this study as well as the proposed radiomic analysis pipeline. We then present the experimental setup and results in Section III and discuss our main findings in Sections IV.

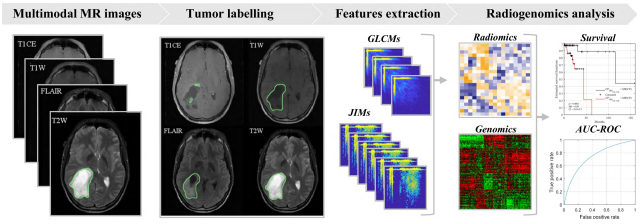

FIGURE 1. Proposed workflow. (1) Multimodal MR scans. (2) Slice-by-slice labeling of tumors (green mask). (3) Extraction and encoding of multimodal grey-level co-occurrence matrix (GLCM) and 3D-JIM features. (4) Statistical analyses (radiogenomics), based on Wilcoxon tests, log-rank tests and random forest classifiers, to identify relevant features for predicting the gene status and survival outcome of LGG patients.

II. MATERIALS AND METHODS

Figure 1 shows the workflow employed in this study. Four MRI sequences, i.e. T1-weighted (T1W), contrast-enhanced T1-weighted (T1CE), T2-weighted (T2W), and Fluid Attenuation Inversion Recovery (FLAIR) images, are first acquired for LGG patients. Tumor regions of interest (ROI) are then labeled manually in each scan. Afterwards, a 64×64 grey levels matrix is extracted automatically from each ROI and encoded using 19 standard quantifier functions. Various analyses are conducted to assess the usefulness of these features to predict gene status and outcome. In a first analysis, a Wilcoxon rank sum test is used to compare features in LGG patients grouped based on gene status (i.e., mutation vs wild). We then consider these features to train a random forest (RF)

model for the prediction of gene status. This model is also employed to identify features with high predictive value for discriminating between mutant and wild types of four genes (i.e., IDH1, ATRX, 1p/19q codeletion and TP53) relevant to LGG.

For the survival analysis, we applied a log-rank test and Kaplan-Meier estimator to find features leading to significantly different survival curves when grouping patients based on the features' median value. Similarly, we employed a RF model to classify LGG patients into groups corresponding to short survival (i.e., below the median survival time) and long survival (i.e., above or equal to the median survival time). The statistical significance of resulting patient groupings was measured using the log-rank test and Kaplan-Meier estimator. All image manipulations, matrix calculations, significance tests and classifications were performed in MATLAB R2018b (Math Works Inc., Natick, MA, USA).

A. PATIENTS AND DATA

A total of 199 grade II and III glioma cases were retrospectively reviewed from The Cancer Imaging Archive (TCIA) [25], a publicly-available medical image repository. All cases have been previously de-identified by TCGA/TCIA, and corresponding datasets, clinical and relevant gene information are available for public download in the TCGA-LGG database. Moreover, variant annotation was taken from the publicly available TCGA whole exome sequencing (WES [26]) dataset (http://www.cbioportal.org/) downloaded on July 2017. Hence, no institutional review board or Health Insurance Portability and Accountability Act (HIPAA) approval were required for our study. Available information on the TCGA-LGG database is reported in [25], including TCIA institutional identifier, patient information (i.e., sex, age, weight), scanner information (i.e., manufacturer) and specific imaging volume information extracted from the DICOM headers (i.e., modality name, slice thickness, slice spacing, repetition time, echo time, inversion time, imaging frequency, flip angle, specific absorption rate, numbers of slices, pixel dimensions, etc. The TCGA-LGG database consists in multisite data that were acquired by potentially different MRI scanner models with different pixel spacing and slice thickness. To overcome these differences, all volume datasets were resampled to a common voxel resolution of $256 \times 256 \times v$ voxels (v is the slice number varying from one tumor-patient to another) and normalized to the [0, 1] range.

An oncologist with 25 years of experience reviewed axial T1CE, T1W, FLAIR and T2W images of available patients, in a blind fashion. Regions of interest (ROI) were delineated using 3D Slicer 3.6 (http://www.slicer.org/), by rigidly aligning MRI scans of different modalities and then manually labelling them slice by slice (e.g., Figure 1) in each of the four MRI sequences. Due to the complexity to find whole ROI, and to avoid the different acquisition parameters (e.g. sampling resolution) that make the registration between modalities prone to errors, LGG tumor labels were only obtained for a subset of 107 patients. From these cases, respectively 104, 105, 107 and 86 patients underwent T1W, T1CE, T2W and FLAIR scans. Additional characteristics of these patients are summarized in Table 1.

TABLE 1. Characteristics of study population.

Age	Average ± Stdev	44.53 ±13.53
Gender	Male/Female	57/50
OS	Median (censored)	460 (90)
Histology grade	Grade II/III	51/56
Radiation treatment	No/Yes/NA	11/49/47
Histological type	Oligodendroglioma/Astrocytoma	47/60
IDH1	Wild-type/Mutant/Unknown	30/77
ATRX		64/43
TP53		50/57
1p19q codeletion		78/27/2
Laterality	Left/Right/Midline	46/58/3

B. MULTIMODAL RADIOMIC FEATURE EXTRACTION

The proposed joint intensity matrix (JIM) features generalize the well-known grey-level co-occurrence matrices (GLCM) texture features to multiple image modalities [24], [27]–[29]. Given an intensity image with segmented ROI, GLCMs estimate the joint probability of observing a pair of intensity values in two ROI voxels whose relative position is defined by a distance d and angles (ϕ, θ). Specifically, GLCM is defined as a two-dimensional (2D) histogram of grey levels for a pair of voxels, which are separated by a fixed spatial relationship, specified in terms of distance and direction. In JIMs, the same strategy is used to compute the joint intensity distribution between 3D images of different modalities. Let I_1 and I_2 be two image modalities with segmented 3D tumor region R, and denote as i and j intensity values from the set $\{1, \ldots, N\}$. The JIM matrix corresponding to distance d and angle θ can be obtained as

$$J_{d,\phi,\theta}(i,j) = \sum_{\boldsymbol{x}\in R} \begin{cases} 1, & \text{if } I_1(\boldsymbol{x}) = i \text{ and } I_2\left(T_{d,\phi,\theta}(\boldsymbol{x})\right) = j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Here, $T_{d,\phi,\theta}(\boldsymbol{x})$ corresponds to the voxel located at distance d and angles ϕ, θ from $\boldsymbol{x}$. Typically, values used for the displacement d comprise an offset of one to six pixels in thirteen possible directions represented by angles (ϕ, θ), each ranging from 0° to 135° in 45° increments. Thus, JIMs are 2D matrices computed over 3D images. To obtain probability distributions, JIM matrices are then unit normalized:

$$\hat{J}_{d,\phi,\theta}(i,j) = \frac{J_{d,\phi,\theta}(i,j)}{\sum_{i,j=1}^{N} J_{d,\phi,\theta}(i,j)} \quad (2)$$

In this study, we considered $N = 64$ grey levels, giving JIM matrices of size $N^2 = 4096$. Compared to the model presented in [24], which only considered the joint distribution of intensities in spatially-aligned voxels, the JIM features

used in this work compare the intensity of a voxel in a specific 3D image with that of *all* voxels in the same ROI of different modality images. In addition to encoding spatial information in multi-modal textures, this modification is also necessary for cases where segmented ROIs are not aligned in different modalities.

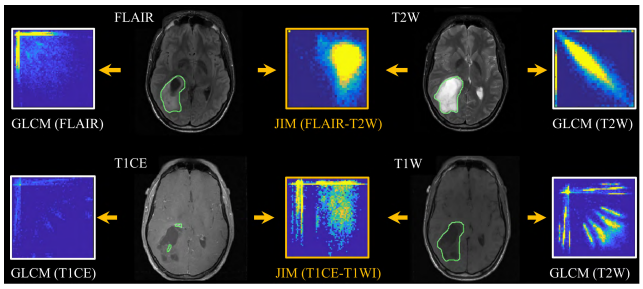

FIGURE 2. Example of GLCM and JIM co-occurrence matrices computed from segmented ROIs in FLAIR, T2W, T1CE and T1W images. GLCM features are based on a single MRI sequence (FLAIR, T2W, T1CE and T1W) while JIM is obtained from multi-parametric MRI data consisting of FLAIR-T2W and T1CE-T1W modalities.

We note that GLCMs can be seen as a special case of JIMs for which the same image is used for I_1 and I_2. Thus, JIM features extend GLCM by considering cross-modality dependencies between voxel intensities, in addition to their spatial relationship. Figure 2 shows an example of JIM and GLCM matrices computed from the segmented FLAIR-T2W and T1CE- T1W images of a patient. As shown, JIMs encode information that is complementary to GLCMs. To convert the 2D JIM matrices into feature vectors, we applied a set of standard quantifier functions [30]–[33]. In this study, we considered the average of the 19 functions (angular second moment, contrast, correlation, sum of squares variance, homogeneity, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, information correlation$_1$, information correlation$_2$, autocorrelation, dissimilarity, cluster shade, cluster prominence, maximum probability and inverse difference), which are presented in Table 1 of the Supplementary Materials. These functions measure various textural characteristics related to heterogeneity, which have been linked to tumor progression and outcome [23], [28]. Using this encoding, each tumor ROI is thus represented by a set of 190 features derived from 4 GLCMs (i.e., T1CE, T1W, FLAIR, T2W) and 6 JIMs (i.e., T1CE-T1W, T1CE-FLAIR, T1CE-T2W, T1W-FLAIR, T1W-T2W, FLAIR-T2W).

C. STATISTICAL ANALYSIS

Several analyses were performed to identify relevant features for the prediction of gene status and survival outcome. The Wilcoxon rank sum test was first used to compare the distribution of feature values across gene status types (i.e., mutation of four relevant genes: ATRX, IDH1, 1p19q-codeletion and TP53). For the survival analysis, we grouped the patients using the median of each feature into two groups: first group with feature value less than the median (cut-off), and second group with feature value above or equal to the median. We then applied the Kaplan-Meier estimator to derive the survival functions of these groups by considering their time-to-event (i.e., number of days from scan until death or last visit) distribution [34]. The log-rank test was employed to compare the resulting survival functions and determine whether their difference is statistically significant. To account for multiple comparisons (19 quantifiers functions $\times$ 10 GLCM/JIM co-occurrence matrices $=$ 190 tests), p-values obtained in these analyses were corrected using the Holm-Bonferroni procedure [35]. We considered each of features statistical significance at $p < 0.05$ following Holm-Bonferroni correction.

In another analysis, we used the same features as input to a random forest (RF) classifier with 1000 random trees for predicting the status (i.e., mutant vs. wild type) of IDH1, ATRX, 1p19q codeletion and TP53 genes, and for predicting the survival of LGG patients. As in most survival analysis studies, we considered the latter task as a binary classification problem, using the median survival time of uncensored patients (i.e., 460 days $\approx$ 15.33 months) to define two classes corresponding to short-term and long-term survival. Considering the median as separation threshold has the advantage of giving even-sized classes, which reduces size-related bias. Many classifier models could be used for the prediction. We considered the RF model because it limits overfitting (when the number of training samples is low) by using a bagging strategy which reduces errors due to sample variance, and can also be used to inspect the importance of features in the classification [36].

A 10-fold cross-validation was applied to obtain unbiased estimates of performance, in which training images are divided into 10 even-sized subsets and, in each fold, one subset is put aside for testing and the remaining 9 are used to train the classifier. Performance metrics is reported by computing the average area under the ROC curve (AUC) obtained from 10 folds. Since the RF model cannot be used directly on censored data (i.e., patients with no death event), we performed an imputation strategy where censored patients are given the average survival outcome of uncensored subjects with a time-to-death greater or equal to their own time of last visit. Other imputation strategies could also have been considered for this purpose [37]. To validate our prediction model and avoid the bias created by the imputation strategy, we applied the Kaplan-Meier estimator with log-rank test to compare the short-term and long-term survival that predicted by the RF model.

To compare the ability of models to predict gene status and survival outcome, we calculated significance using the chi-square test as previously described in [38]. Moreover, for identifying features with high predictive value, we measured the increase in RF error resulting from the permutation of feature values across out-of-bag observations. Importance value of features were computed for every RF tree and averaged over the entire ensemble. These values were then normalized by dividing them by the ensemble's standard deviation. Finally, the importance of each feature was obtained by averaging these normalized values across the 10 folds. A positive

value indicates that the feature is predictive, whereas negative values suggest that the feature has no predictive value.

III. EXPERIMENTAL RESULTS

We start by reporting results related to the prediction of gene status, and the present results obtained in the analysis of survival outcome.

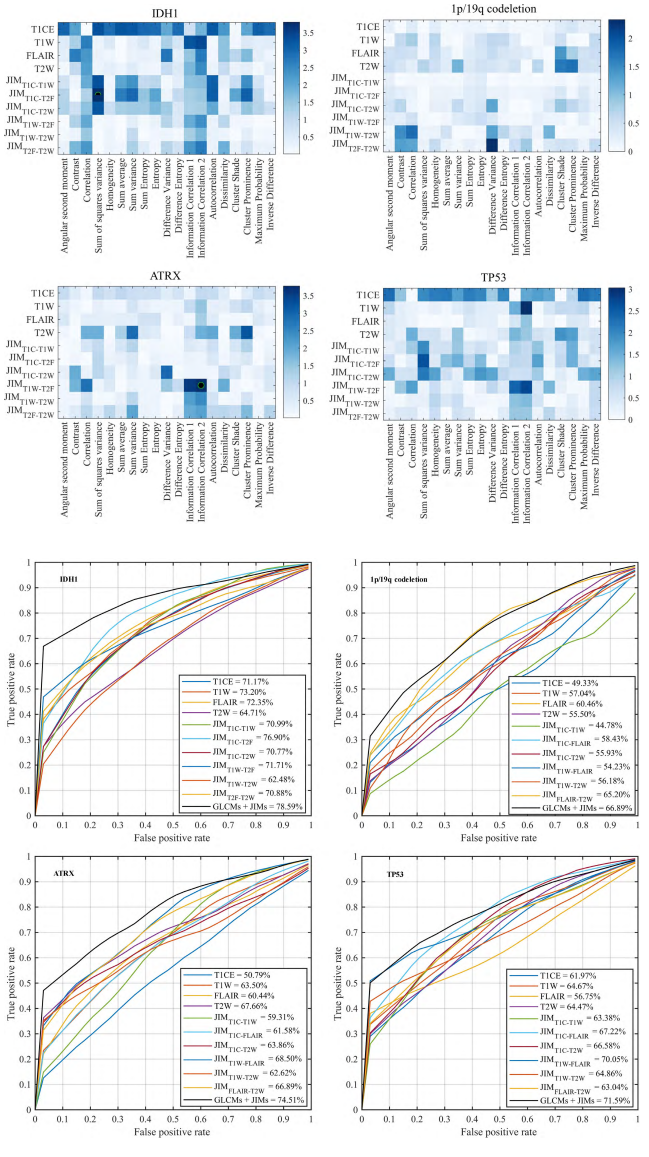

FIGURE 3. (Top) Heatmap of Wilcoxon test p-values (negative log10 scale) comparing the distribution of GLCM and JIM features in LGG patients with mutant or wild types for the IDH1, ATRX, 1p/19q codeletion and TP53 genes. Significant features (i.e., corrected $p < 0.05$) are indicated with a black-green circle. (Bottom) The area under the ROC curve (average AUC on 10 folds) obtained by the RF classifier with GLCM or JIM features, for predicting the gene status of LGG patients.

A. GENE STATUS ANALYSIS

Figure 3 shows the heatmap of p-values (in negative log10 scale) obtained by the Wilcoxon test comparing the feature value distributions of patients with mutant or wild-type for the IDH1, ATRX, 1p/19q codeletion and TP53 genes. For all the studied genes, we observe various features exhibiting high difference between mutant and wild types. However, only two features are significant with corrected $p < 0.05$: JIMT1C-T2F sum of square variance for the IDH1 gene, and JIMT1W-T2F information correlation$_2$ for ATRX. Comparing genes, we see that GLCM and proposed JIM features have the highest differences for the IDH1 gene, in particular features extracted from T1CE images.

In the multivariate prediction analysis based on the RF classifier, for all genes, we obtained the highest average AUC ($p<0.05$) values using JIM features: 76.90% for IDH1 using JIMT1CE-FLAIR, 68.50% for ATRX using JIMT1W-FLAIR, 65.20% for 1p/19q codeletion using JIMFLAIR-T2W, and 70.05% for TP53 using JIMT1W-FLAIR. Notably, comparison of AUC obtained from GLCMs and JIMs reached significance in a chi-square test ($p=0.05$; Supplementary Figure 1). The modality combinations leading to the highest prediction score for each gene are consistent with results of the Wilcoxon test. For example, in the case of IDH1, the feature with greatest Wilcoxon test significance was also observed for the T1CE-FLAIR JIM combination (sum of square variance). Moreover, considering together features derived from GLCMs and JIMs improves performance by 1% to 6%, the largest improvement observed for ATRX.

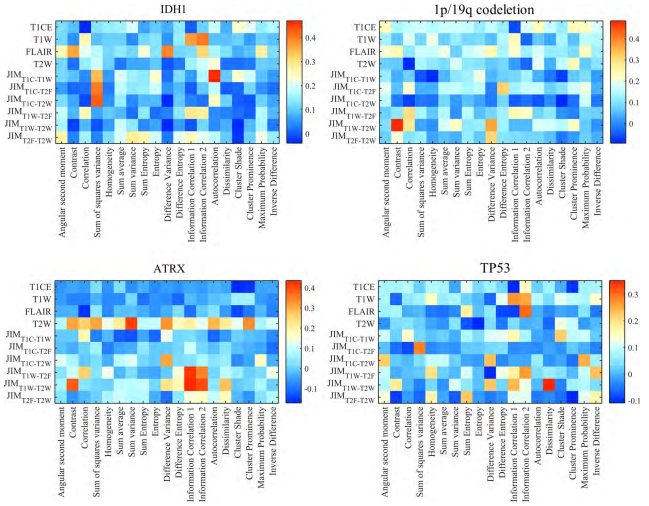

FIGURE 4. Importance of individual features for predicting the status (i.e., mutant or wild-type) of IDH1, ATRX, 1p/19q codeletion and TP53 genes using the RF classifier. Reported values correspond to the average increase in prediction error obtained by permuting the values of individual features across out-of-bag observations [36]. Positive and negative values correspond respectively to predictive and non-predictive features.

The importance of individual features for predicting gene status, as measured by the RF classifier, is shown in Figure 4. Once again, we observe various GLCM and JIM features with predictive value (i.e., importance score > 0). Several of these features (e.g., the sum of squares variance in IDH1 and information correlation 2 in ATRX) were previously found to be significant in the Wilcoxon analysis (Figure 3). Inconsistencies between the Wilcoxon significance and RF importance

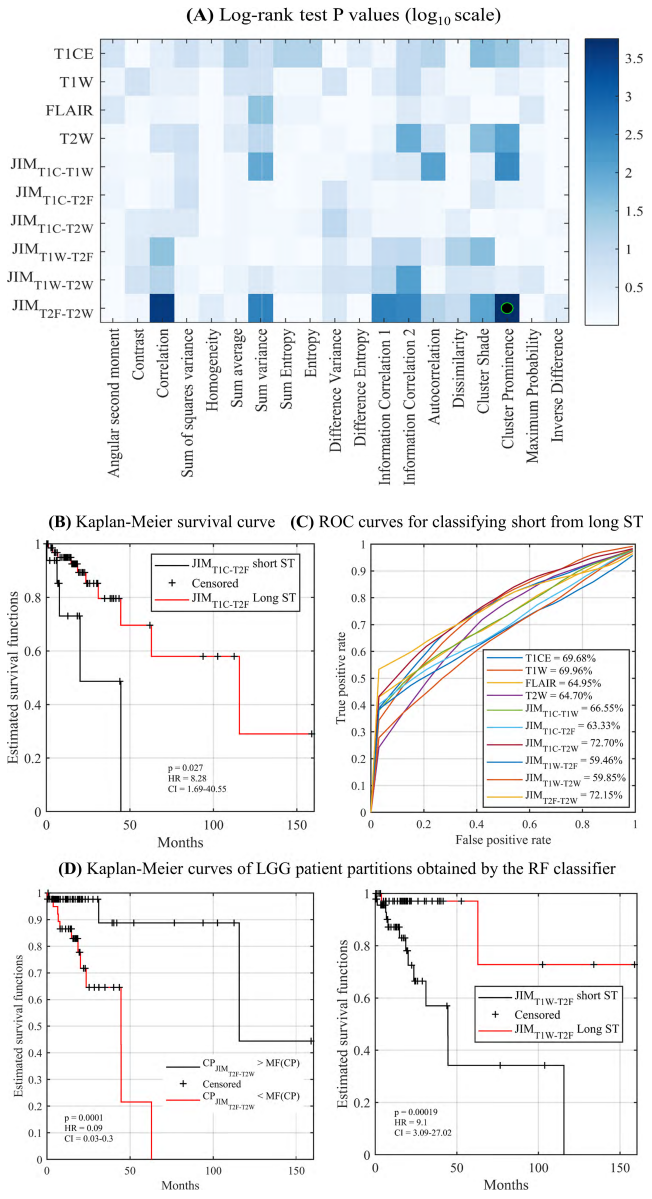

FIGURE 5. **(A) Heatmap of log-rank test p-values (negative log10 scale) comparing the survival function of patients divided by the median value of each feature. Significant features (i.e., corrected $p < 0.05$) are indicated with a black-green circle. (B) Kaplan-Meier survival curves obtained for the only significant feature: JIMT2F-T2W cluster prominence. MF refers to the median feature value. (C) The area under the ROC curves (average AUC on 10 folds) obtained by the RF classifier using GLCM and JIM features, for predicting patients with below-median or above-median survival. (D) Kaplan-Meier curves of LGG patients that significantly partitioned by RF classifier model using the following modality combinations: T1C-FLAIR and T1W-FLAIR. Red curves correspond to the long survival group and black curves to the short survival group.**

could be explained by the randomness of feature selection in RF, where discriminative features can be ignored in favor of more predictive ones.

B. SURVIVAL ANALYSIS

Figure 5A gives the heatmap of log-rank test p-values (in negative log10 scale) for groups of patients divided by the median value of features. In this analysis, a single feature appears to be statistically significant after p-value correction: cluster prominence derived from the JIMT2F-T2W. The Kaplan-Meier curves corresponding to this feature are depicted in Figure 5B. It can be seen that LGG patients with below-median value have a lower survival rate with a hazard ratio (HR) of 0.09 and a confidence interval (CI) of 0.03-0.3. Accordingly, these patients have a greater median survival of 15.70 months compared to 15.33 months for patients with above-median value. The prediction accuracy of the RF classifier using GLCM or JIM features as input is provided in Figure 5C. Overall, JIMT1C-T2W and JIMT2F-T2W features lead to the best prediction with an average AUC of 72.70% and 72.15%, respectively. These results support the findings of the univariate analysis in Figure 5A.

To further validate these results, we applied the Kaplan-Meier estimator and log-rank test on the partition obtained by the RF classifier (Figure 5D). We find that patient groups obtained using JIMT1C-T2F and JIMT1W-T2F features have significantly different survival profiles, with corrected log-rank $p < 0.05$ (Table 2). Conversely, GLCM features and JIM features derived from T1C-T1W, T1C-T2W, T1W-T2W and T2F-T2W modality combinations are not significant.

TABLE 2. **Summary of log-rank and Kaplan-Meier analysis for two predicted groups (short and long survival time) of LGG patients using random forest classifier model.**

	Features	MS (months)		HR	CI	p
		Short ST	Long ST			
GLCMs	T1CE	6.63	17.77	1.42	0.41-4.87	0.8123
	T1WI	16.55	15.35	1.16	0.39-3.47	0.9959
	FLAIR	11.73	20.60	1.41	0.46-4.27	0.7485
	T2W	12.37	15.63	1.26	0.41-3.83	0.9038
JIMs	T1C-1W	7.80	15.45	0.38	0.13-1.11	0.1365
	T1C-T2F	7.25	16.52	8.28	1.69-40.56	***0.0276**
	T1C-2W	7.80	16.23	0.41	0.14-1.19	0.1717
	T1W-T2F	14.00	17.77	9.15	3.10-27.02	***0.0002**
	T1W-2W	7.25	16.08	0.70	0.21-2.26	0.7647
	T2F-T2W	15.33	15.70	2.96	1.06-8.24	0.0700

* Represents the significant features with $p < 0.05$; MS: median survival; ST: survival time; HR: Hazard ratio; CI: confidential interval.

To evaluate the complementarity of JIMs to standard GLCMs and features encoding tumor volume, Figure 6 reports the prediction performance obtained with different combinations of these features. Combining GLCM and JIM features (i.e., a total of $19 \times 10 = 190$ features per patient), we get an improved average AUC of 83.42% (Figure 6A). Adding tumor volume to GLCM and JIM features further increases average AUC to 86.79%. In contrast, using tumor volume by itself leads to a poor prediction, with an average AUC of 46.01%. Notably, comparison of AUC obtained from volume, GLCMs and JIMs reached significance in a chi-square test ($p=0.05$; Supplementary Figure 2).

Moreover, Kaplan-Meier curves of patient groups predicted by the RF model are given in Figure 6B. Groupings corresponding to GLCM+JIM or GLCM+JIM+Volume combinations show significantly different survival profiles

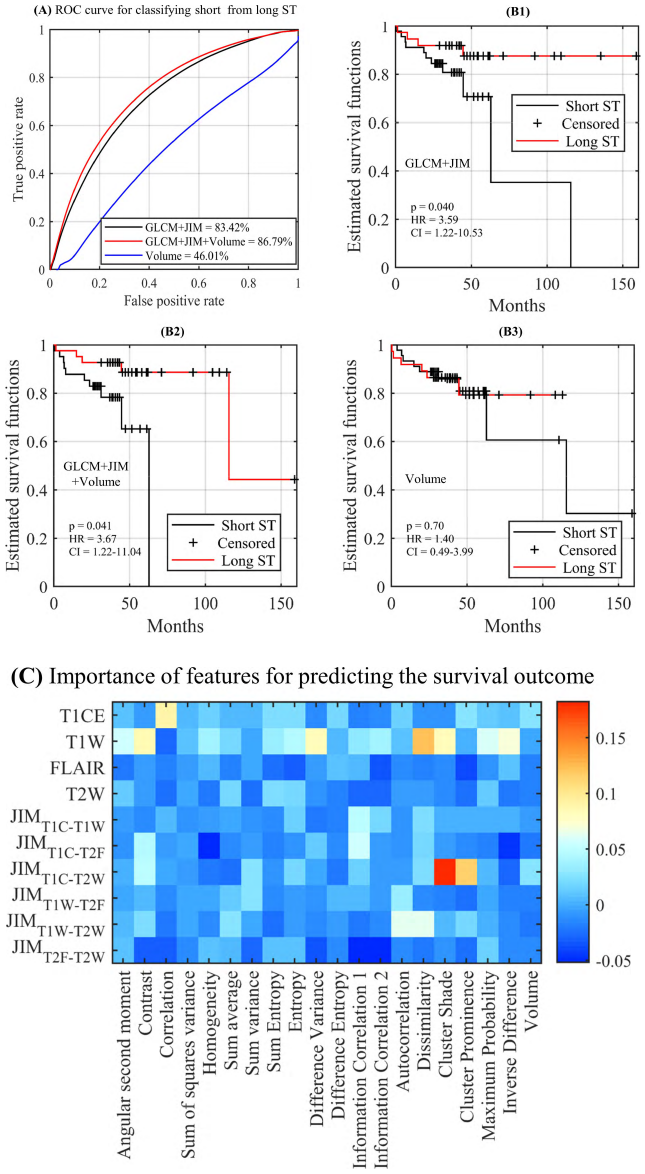

FIGURE 6. **(A) ROC curve (average AUC on 10 folds) for the prediction of short (i.e., below-median) versus long (i.e., above-median) survival patients using volume, combined GLCM and JIM features (i.e., GLCM+JIM), or all features (i.e., GLCM+JIM+Volume). (B1, B2, B3) Kaplan-Meier curves of the predicted short (black curve) and long (red curve) survival groups. (C) Importance of individual features for predicting the survival group with the RF classifier. Positive/negative values correspond to predictive/ non-predictive features.**

with log-rank $p < 0.05$. Following previous results, using tumor volume alone gives patient groupings with no statistically significant differences. Inspecting the importance of individual features in a RF classifier trained with combined features, we see that the most predictive features are from JIMT1C-T2W (i.e., cluster shape and cluster prominence).

IV. DISCUSSION

Radiomic analysis has recently emerged as a powerful technique for quantifying the intrinsic heterogeneity, genetic characteristics and other phenotypes of tumors [39], [40]. Several studies have also demonstrated the link between radiomic features and both clinical outcomes and gene-expression levels [4], [22], [23], [41]–[43]. In this work, we proposed using radiomic features, based on the joint distribution of intensities across four MRI sequences, to predict the gene status and survival outcome of LGG patients. Our results showed that proposed JIM features can help determine the mutant or wild-type status of relevant genes for LGG. Thus, we obtained average AUC values of 78.59% (IDH1), 74.51% (ATRX), 71.59% (TP53), and 66.89% (1p/19q codeletion), when combining JIMs with GLCM features, compared to highest values of 73.20%, 67.66%, 60.46% and 64.67% using single-modality GLCMs (Figure 3).

These results are comparable with those of recent studies employing imaging data to predict gene status in LGG [4], [21], [44] and GBM [45] tumors. For instance, Zhou *et al.* reported an AUC (632+ bootstrap technique) of 79% using texture features and random forest classifiers for predicting IDH1 mutation in 165 TCGA subjects with grade II or III gliomas. Prediction accuracy was increased to 86% when adding Visually Accessible Rembrandt Images (VASARI) annotations and clinical variables to texture features as input. In [21], a model based on convolutional neural networks (CNN) and Fisher vectors is proposed to predict IDH1 mutation in 151 low-grade patients of a private dataset. Authors report an AUC of 92.1% using single-modality data (T2 FLAIR) and an AUC of 95.2% with multi-modal data (T1 contrast and T2 FLAIR), compared to 85.7% when employing 671 radiomic features encoding tumor texture, shape and location [44]. However, these values were obtained on a highly-unbalanced dataset using only grade II patients. Our results also compare to those presented in [46], where an average AUC of 76.3% was obtained using 431 radiomic features for predicting TP53 status in 272 patients with primary grade II or III gliomas. Differences between our results and those reported in the literature for predicting gene status with imaging features could be due to several additional factors, including the number of samples, the quality of images and the tuning of classifier models.

The proposed analysis also demonstrated the advantage of multimodal JIM features over conventional GLCMs for predicting the survival outcome of LGG patients. JIM features based on T1C-FLAIR and T1W-FLAIR modality combinations were found to have significantly different distributions in patients with short versus long survival, and shown to be highly predictive of survival (Figure 5). Additionally, our analysis found JIM features to provide information that is complementary to conventional GLCMs and to improve prediction when combined with these features (Figure 6). These results are consistent with recent studies confirming the association between imaging features and LGG survival. For instance, the study in [47] showed that features derived from diffusion-weighted and perfusion-weighted images have prognostic value. Likewise, Zhou *et al.* [4] found radiomic features derived from multiparametric MRI images to be predictive of progression

in LGG patients. Unlike the features used in these studies, our JIM features can encode relationships between multiple modalities.

The proposed features measure intensity heterogeneity inside and across images of different modalities. We hypothesize that this heterogeneity is linked with histological patterns within LGG tumors that are indicative of progression and overall survival. However, explaining why certain features (e.g., JIMT1C-T2F sum of square variance and JIMT1W-T2F information correlation$_2$) are more predictive than others for a specific task is challenging. Increasing scan resolution could alleviate this problem by bridging the gap between the scale of image textures and the microstructure of tumor tissues.

Although promising, this study has some limitations that merit discussion. First, it is based on data that was retrospectively collected from the TCIA database. While the number of subjects (i.e., n=107) is relatively low, the primary goal of the study was to evaluate the performance of JIM features for gene status and survival prediction of LGG patients. Nonetheless, evaluating these features on an external dataset could help prospectively validate our findings. Second, images were collected from multiple sites, each one using different acquisition parameters. For instance, the slice thicknesses of scans used in our study varied from 1.5 to 5 mm. Since acquisition parameters may affect texture, future work could consider data from a single site with a common slice thickness. Moreover, extending our proposed model to consider additional imaging modalities like diffusion-weighted imaging (DWI) or dynamic contrast-enhanced (DCE) sequences could further enhance performance. Furthermore, as in other radiomics approaches, our analysis is based on handcrafted features that encode textural characteristics of tumor regions. Learning discriminative features directly from the data, for instance using convolutional neural networks, could potentially improve the prediction of gene status and survival outcome.

V. CONCLUSION

This study presented novel radiomic features, derived from an extended version of joint intensity matrices (JIMs), that compute the joint probability of intensity values across multiparametric MRI images. We investigated the usefulness of these features for comparing and predicting the gene status (IDH1, ATRX, TP53 and 1p/19q codeletion) and survival outcome of LGG patients. Results demonstrate the advantages over standard GLCM features of our extension of JIMs, and suggest that the proposed features could be used as biomarkers for predicting IDH1 and ATRX gene status in LGG tumors, as well as the survival outcome of LGG patients.

REFERENCES

[1] M. L. Goodenberger and R. B. Jenkins, "Genetics of adult glioma," *Cancer Genet.*, vol. 205, no. 12, pp. 613–621, Dec. 2012.

[2] T. Komori, "The 2016 WHO classification of tumours of the central nervous system: The major points of revision," *Neurologia Medico-Chirurgica*, vol. 57, no. 7, pp. 301–311, Jul. 2017.

[3] The Cancer Genome Atlas Research Network, "Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas," *New England J. Med.*, vol. 372, no. 26, pp. 2481–2498, Jun. 2015.

[4] H. Zhou, M. Vallières, H. X. Bai, C. Su, H. Tang, D. Oldridge, Z. Zhang, B. Xiao, W. Liao, Y. Tao, and J. Zhou, "MRI features predict survival and molecular markers in diffuse lower-grade gliomas," *Neuro-Oncol.*, vol. 19, no. 6, pp. 862–870, Jan. 2017.

[5] S. Venneti and J. T. Huse, "The evolving molecular genetics of low-grade glioma," *Adv. Anatomic Pathol.*, vol. 22, no. 2, p. 94, Mar. 2015.

[6] A. L. Cohen, S. L. Holmen, and H. Colman, "IDH1 and IDH2 mutations in gliomas," *Current Neurol. Neurosci. Rep.*, vol. 13, no. 5, p. 345, May 2013.

[7] C. Hartmann, B. Hentschel, W. Wick, D. Capper, J. Felsberg, M. Simon, M. Westphal, G. Schackert, R. Meyermann, T. Pietsch, G. Reifenberger, M. Weller, M. Loeffler, and A. von Deimling, "Patients with IDH1 wild type anaplastic astrocytomas exhibit worse prognosis than IDH1-mutated glioblastomas, and IDH1 mutation status accounts for the unfavorable prognostic effect of higher age: Implications for classification of gliomas," *Acta Neuropathologica*, vol. 120, no. 6, pp. 707–718, Dec. 2010.

[8] R. J. Molenaar, T. Radivoyevitch, J. P. Maciejewski, C. J. F. van Noorden, and F. E. Bleeker, "The driver and passenger effects of isocitrate dehydrogenase 1 and 2 mutations in oncogenesis and survival prolongation," *Biochimica et Biophysica Acta (BBA)-Rev. Cancer*, vol. 1846, no. 2, pp. 326–341, Dec. 2014.

[9] C. Houillier, X. Wang, G. Kaloshi, K. Mokhtari, R. Guillevin, J. Laffaire, S. Paris, B. Boisselier, A. Idbaih, F. Laigle-Donadey, K. Hoang-Xuan, M. Sanson, and J.-Y. Delattre, "IDH1 or IDH2 mutations predict longer survival and response to temozolomide in low-grade gliomas," *Neurology*, vol. 75, no. 17, pp. 1560–1566, Oct. 2010.

[10] G. Kaloshi *et al.*, "Temozolomide for low-grade gliomas predictive impact of 1p/19q loss on response and outcome," *Neurology*, vol. 68, no. 21, pp. 1831–1836, May 2007.

[11] A. K. Chan, Y. Mao, and H.-K. Ng, "TP53 and histone H3.3 mutations in triple-negative lower-grade gliomas," *New England J. Med.*, vol. 375, no. 22, pp. 2206–2208, Dec. 2016.

[12] X. Fan, Y. Wang, C. Zhang, L. Liu, S. Yang, Y. Wang, X. Liu, Z. Qian, S. Fang, H. Qiao, and T. Jiang, "ADAM9 expression is associate with glioma tumor grade and histological type, and acts as a prognostic factor in lower-grade gliomas," *Int. J. Mol. Sci.*, vol. 17, no. 9, p. 1276, Aug. 2016.

[13] G. Wu, Y. Chen, Y. Wang, J. Yu, X. Lv, X. Ju, Z. Shi, L. Chen, and Z. Chen, "Sparse representation-based radiomics for the diagnosis of brain tumors," *IEEE Trans. Med. Imag.*, vol. 37, no. 4, pp. 893–905, Apr. 2018.

[14] A. Arnaud, F. Forbes, N. Coquery, N. Collomb, B. Lemasson, and E. L. Barbier, "Fully automatic lesion localization and characterization: Application to brain tumors using multiparametric quantitative MRI data," *IEEE Trans. Med. Imag.*, vol. 37, no. 7, pp. 1678–1689, Jul. 2018.

[15] C. Ma, G. Luo, and K. Wang, "Concatenated and connected random forests with multiscale patch driven active contour model for automated brain tumor segmentation of MR images," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1943–1954, Aug. 2018.

[16] M. Lê, H. Delingette, J. Kalpathy-Cramer, E. R. Gerstner, T. Batchelor, J. Unkelbach, and N. Ayache, "MRI based Bayesian personalization of a tumor growth model," *IEEE Trans. Med. Imag.*, vol. 35, no. 10, pp. 2329–2339, Oct. 2016.

[17] A. Chaddad, C. Desrosiers, and M. Toews, "GBM heterogeneity characterization by radiomic analysis of phenotype anatomical planes," *Proc. SPIE, Int. Soc. Opt. Photon.*, vol. 9784, Mar. 2016, Art. no. 978424.

[18] A. Chaddad, C. Desrosiers, and M. Toews, "Radiomic analysis of multicontrast brain MRI for the prediction of survival in patients with glioblastoma multiforme," in *Proc. 38th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2016, pp. 4035–4038.

[19] A. Chaddad, P. O. Zinn, and R. R. Colen, "Radiomics texture feature extraction for characterizing GBM phenotypes using GLCM," in *Proc. IEEE 12th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2015, pp. 84–87.

[20] G. Lee, H. Y. Lee, E. S. Ko, and W. K. Jeong, "Radiomics and imaging genomics in precision medicine," *Precis. Future Med.*, vol. 1, no. 1, pp. 10–31, Mar. 2017.

[21] Z. Li, Y. Wang, J. Yu, Y. Guo, and W. Cao, "Deep learning based radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma," *Sci. Rep.*, vol. 7, no. 1, Jul. 2017, Art. no. 5467.

[22] A. Chaddad, C. Desrosiers, L. Hassan, and C. Tanougast, "A quantitative study of shape descriptors from glioblastoma multiforme phenotypes for predicting survival outcome," *Brit. J. Radiol.*, vol. 89, no. 1068, Jun. 2016, Art. no. 20160575.

[23] A. Chaddad and C. Tanougast, "Extracted magnetic resonance texture features discriminate between phenotypes and are associated with overall survival in glioblastoma multiforme patients," *Med. Biol. Eng. Comput.*, vol. 54, no. 11, pp. 1707–1718, Nov. 2016.
[24] A. Chaddad, P. Daniel, C. Desrosiers, M. Toews, and B. Abdulkarim, "Novel radiomic features based on joint intensity matrices for predicting glioblastoma patient survival time," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 795–804, Mar. 2019.
[25] F. W. Prior, K. Clark, P. Commean, J. Freymann, C. Jaffe, J. Kirby, S. Moore, K. Smith, L. Tarbox, B. Vendt, and G. Marquez, "TCIA: An information resource to enable open science," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 1282–1285.
[26] A. Warr, C. Robert, D. Hume, A. Archibald, N. Deeb, and M. Watson, "Exome sequencing: Current and future perspectives," *G3, Genes, Genomes, Genet.*, vol. 5, no. 8, pp. 1543–1550, Aug. 2015.
[27] T. Torheim, E. Malinen, K. Kvaal, H. Lyng, U. G. Indahl, E. K. F. Andersen, and C. M. Futsæther, "Classification of dynamic contrast enhanced MR images of cervical cancers using texture analysis and support vector machines," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1648–1656, Aug. 2014.
[28] W. Gomez, W. C. A. Pereira, and A. F. C. Infantosi, "Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound," *IEEE Trans. Med. Imag.*, vol. 31, no. 10, pp. 1889–1899, Oct. 2012.
[29] M.-C. Yang, W. K. Moon, Y.-U. F. Wang, M. S. Bae, C.-S. Huang, J.-H. Chen, and R.-F. Chang, "Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis," *IEEE Trans. Med. Imag.*, vol. 32, no. 12, pp. 2262–2273, Dec. 2013.
[30] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, May 1979.
[31] L.-K. Soh and C. Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 2, pp. 780–795, Mar. 1999.
[32] M. Bevk and I. Kononenko, "A statistical approach to texture description of medical images: A preliminary study," in *Proc. 15th IEEE Symp. Comput.-Based Med. Syst.*, Jun. 2002, pp. 239–244.
[33] D. A. Clausi, "An analysis of co-occurrence texture statistics as a function of grey level quantization," *Can. J. Remote Sens.*, vol. 28, no. 1, pp. 45–62, 2002.
[34] D. G. Kleinbaum and M. Klein, "Kaplan-Meier survival curves and the log-rank test," in *Survival Analysis*, New York, NY, USA: Springer, 2012, pp. 55–96.
[35] S. Holm, "A simple sequentially rejective multiple test procedure," *Scand. J. Statist.*, vol. 6, no. 2, pp. 65–70, 1979.
[36] K. J. Archer and R. V. Kimes, "Empirical characterization of random forest variable importance measures," *Comput. Statist. Data Anal.*, vol. 52, no. 4, pp. 2249–2260, Jan. 2008.
[37] J. P. T. Higgins, I. R. White, and A. M. Wood, "Imputation methods for missing outcome data in meta-analysis of clinical trials," *Clin. Trials*, vol. 5, no. 3, pp. 225–239, 2008.
[38] N. S. Sumi, M. A. Islam, and M. A. Hossain, "Evaluation and computation of diagnostic tests: A simple alternative," *Bull. Malaysian Math. Sci. Soc.*, vol. 37, no. 2, pp. 411–423, 2014.
[39] S. Mitra and B. U. Shankar, "Integrating radio imaging with gene expressions toward a personalized management of cancer," *IEEE Trans. Human–Mach. Syst.*, vol. 44, no. 5, pp. 664–677, Oct. 2014.
[40] V. Parekh and M. A. Jacobs, "Radiomics: A new application from established techniques," *Expert Rev. Precis. Med. Drug Develop.*, vol. 1, no. 2, pp. 207–226, Mar. 2016.
[41] B. Ganeshan, S. Abaleke, R. C. D. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: Initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imag.*, vol. 10, no. 1, pp. 137–143, Jul. 2010.
[42] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine learning methods for quantitative radiomic biomarkers," *Sci. Rep.*, vol. 5, Aug. 2015, Art. no. 13087.
[43] C. Parmar, R. T. H. Leijenaar, P. Grossmann, E. R. Velazquez, J. Bussink, D. Rietveld, M. M. Rietbergen, B. Haibe-Kains, P. Lambin, and H. J. W. L. Aerts, "Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer," *Sci. Rep.*, vol. 5, Jun. 2015, Art. no. 11044.
[44] J. Yu, Z. Shi, Y. Lian, Z. Li, T. Liu, Y. Gao, Y. Wang, L. Chen, and Y. Mao, "Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma," *Eur. Radiol.*, vol. 27, no. 8, pp. 3509–3522, Aug. 2017.
[45] K. L.-C. Hsieh, C.-Y. Chen, and C.-M. Lo, "Radiomic model for predicting mutations in the isocitrate dehydrogenase gene in glioblastomas," *Oncotarget*, vol. 8, no. 28, pp. 45888–45897, May 2017.
[46] Y. Li, Z. Qian, K. Xu, K. Wang, X. Fan, S. Li, T. Jiang, X. Liu, and Y. Wang, "MRI features predict p53 status in lower-grade gliomas via a machine-learning approach," *NeuroImage, Clin.*, vol. 17, pp. 306–311, Jan. 2018.
[47] V. Cuccarini, A. Erbetta, M. Farinotti, L. Cuppini, F. Ghielmetti, B. Pollo, F. Di Meco, M. Grisoli, G. Filippini, G. Finocchiaro, M. G. Bruzzone, and M. Eoli, "Advanced MRI may complement histological diagnosis of lower grade gliomas and help in predicting survival," *J. Neuro-Oncol.*, vol. 126, no. 2, pp. 279–288, Jan. 2016.

AHMAD CHADDAD received the Ph.D. degree in systems engineering from the University of Lorraine, in 2012. His previous experience includes working as a Postdoctoral Research Fellow with the University of Texas MD Anderson Cancer Center (2013–2015) and McGill University (2016–2017). He joined the Lady Davis Institute for Medical Research, McGill University. His current research interests include biomedical imaging, signal and image processing, pattern recognition, and microelectronics circuits. He is a member of The Laboratory of Design, Optimization and Modelling of Systems (LCOMS).

CHRISTIAN DESROSIERS received the Ph.D. degree in computer engineering from Polytechnique Montreal, in 2008. He was a Postdoctoral Researcher on the topic of machine learning with the University of Minnesota. In 2009, he joined ETS, University of Quebec, as a Professor with the Department of Software and IT Engineering. His main research interests focus on machine learning, image processing, computer vision, and medical imaging. He is the co-director of the Laboratoire d'imagerie, de vision et d'intelligence artificielle (LIVIA) and is a member of the REPARTI Research Network.

BASSAM ABDULKARIM received the medical degree from Paris XI University and completed his residency training in radiation oncology with the Gustave Roussy Cancer Center, and the Ph.D. degree in radiation biology from Paris XI University, in 2002. He is currently a Professor of oncology with McGill University and a Principal Investigator with the Research Institute of McGill University Health Center (MUHC), where he has been serving as the Director of the McGill Division of Radiation Oncology, since 2011. His current research program focuses on investigating molecular mechanisms underlying response to different radiation regimens in glioblastoma.

TAMIM NIAZI received the Doctor of Medicine and Master of Surgery (MDCM) designation from McGill University, in 2001. He continued his internship and residency with the McGill University Health Centre. From 2006 to 2007, he was the Clinical Trials Fellow with the National Cancer Institute of Canada Clinical Trials Group (NCIC-CTG). He is currently an Adjoint Professor with the Department of Oncology, McGill University. He is also the fellowship co-director of the Division of Radiation Oncology, Department of Oncology, McGill University. He has presented and published over 50 peer reviewed academic works.