GBM heterogeneity characterization by radiomic analysis of phenotype anatomical planes

Ahmad Chaddad*, Christian Desrosiers, Matthew Toews
Laboratory for Imagery, Vision and Artificial Intelligence, Ecole de Technologie Superieure, Montreal, Qc, H3C1K3, Canada

ABSTRACT

Glioblastoma multiforme (GBM) is the most common malignant primary tumor of the central nervous system, characterized among other traits by rapid metastatis. Three tissue phenotypes closely associated with GBMs, namely, necrosis (N), contrast enhancement (CE), and edema/invasion (E), exhibit characteristic patterns of texture heterogeneity in magnetic resonance images (MRI). In this study, we propose a novel model to characterize GBM tissue phenotypes using gray level co-occurrence matrices (GLCM) in three anatomical planes. The GLCM encodes local image patches in terms of informative, orientation-invariant texture descriptors, which are used here to sub-classify GBM tissue phenotypes. Experiments demonstrate the model on MRI data of 41 GBM patients, obtained from the cancer genome atlas (TCGA). Intensity-based automatic image registration is applied to align corresponding pairs of fixed T1-weighted (T1-WI) post-contrast and fluid attenuated inversion recovery (FLAIR) images. GBM tissue regions are then segmented using the 3D Slicer tool. Texture features are computed from 12 quantifier functions operating on GLCM descriptors, that are generated from MRI intensities within segmented GBM tissue regions. Various classifier models are used to evaluate the effectiveness of texture features for discriminating between GBM phenotypes. Results based on T1-WI scans showed a phenotype classification accuracy of over 88.14%, a sensitivity of 85.37% and a specificity of 96.1%, using the linear discriminant analysis (LDA) classifier. This model has the potential to provide important characteristics of tumors, which can be used for the sub-classification of GBM phenotypes.

Keywords: Classification, GBM, GLCM, MRI, Texture feature.

1. INTRODUCTION

GBM is the most common malignant brain tumor, accounting for 16% of all primary brain tumors [1]. The median overall survival is 12−15 months [2], although some GBM patients might live more than 5 years [3]. MRI has been widely used for GBM imaging, providing in-vivo images with high tissue spatial resolution. Feature extraction from MRI, either by a trained radiologist or semi-automated systems, thus plays an important role in diagnosis and in selecting the best plan of treatment [4]. The process of GBM proliferation can be described in terms of multiple tissue phenotypes generated from the main GBM tumor, that can be spatially localized into multiple clusters of cells [5]. While the precise role of each phenotype in the GBM tumor is still unclear, the characterization of these phenotypes in terms of image texture features, in particular texture heterogeneity, may be potentially useful in surgical planning, computer-assisted diagnosis and prognosis, as well as understanding the underlying bio-anatomical mechanisms of the disease.

Specifically, this paper proposes to characterize GBM tissue phenotypes in terms of texture heterogeneity features. This model is well-suited for capturing cellular phenomena related to GBM malignancy, e.g. as regions of high cell density such as active tumor and edema, or low cell density such as necrotic regions [6–8]. Phenotype regions can be segmented manually or semi-automatically by image processing tools and multi-parametric MRI data. For example, within the context of this work, contrast enhanced (CE) and necrotic (N) tissue regions are identified from T1 weighted (T1-WI) MRI and edema (E) regions from fluid attenuation inversion recovery (FLAIR) sequences [9].
Automatic phenotype identification has clear potential clinical benefits, e.g. in computer-assisted diagnosis and prognosis, however few guidelines exist as to the most effective feature representation and machine learning frameworks. A number of image texture features exist and may be useful in quantifying links between image data and GBM tissue phenotypes, including Gaussian mixture model (GMM) features [10], histogram based features [11], wavelet based feature [12], and texture based features [13–16]. In particular, GMM based features have been successful in classifying between normal from abnormal brain heterogeneity [17]. Haralick features[18] represent an important general class of texture

features computed from the gray level co-occurrence matrix (GLCM), where a set of so-called *quantifier functions* are used to derive numerical values for texture properties such as entropy, homogeneity, etc. from GLCM descriptors. Such GLCM-based descriptors have previously been used to quantify tumor heterogeneity [13], potentially leading to standard reporting techniques in order to accurately characterize tumor heterogeneity [16]. A previous study, using texture features and decision trees to classify GBM phenotypes, showed an overall accuracy of 75.58% [19]. With respect to previous work, this paper presents the first study of texture features computed in different anatomical planes for the characterization of GBM phenotypes. The main hypothesis of our study is that texture features, which can reliably extracted in an automatic manner, can be extracted from standard anatomical planes and used to assess GBM phenotypes. Another contribution of this work is the comparison of GLCM features obtained from different image sequences, for the classification of GBM phenotypes.

2. MATERIALS AND METHODS

This study uses MRI data of 41 GBM subjects, publically available through The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA). We used intensity-based automatic image registration to align fixed T1-WI post contrast images with their corresponding FLAIR images. We then manually segmented the regions corresponding to necrosis, contrast enhancement and edema with the 3D Slicer tool, and extracted texture features in each of these regions from both T1-WI and FLAIR images. We repeated this technique for each of the three anatomical planes (i.e., axial, coronal and sagittal), and averaged the texture features over these three planes. Figure 1 shows the proposed workflow for GBM phenotypes characterization, classification and validation.

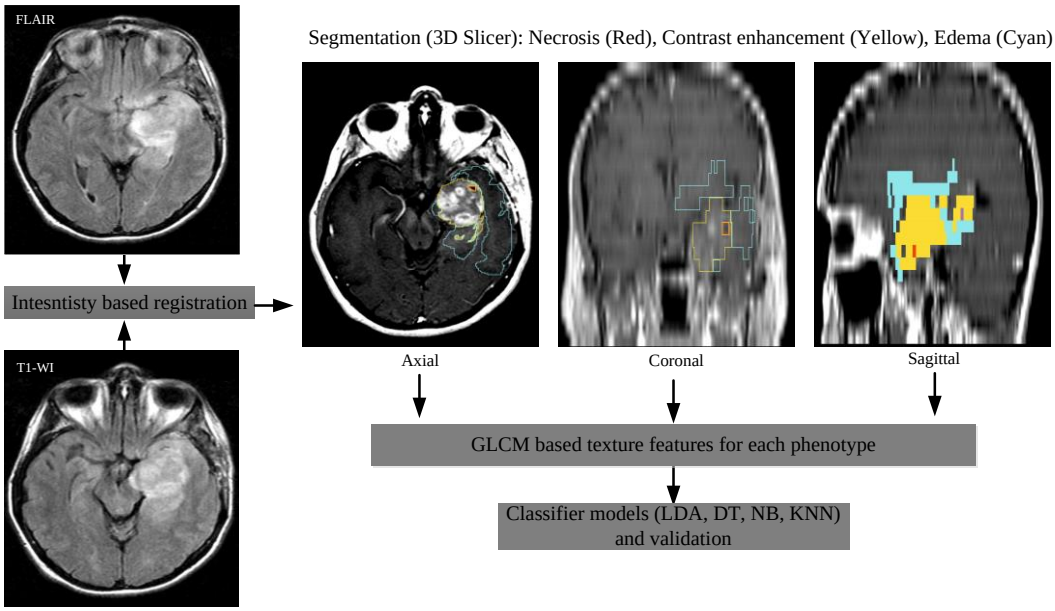


Figure 1. Overall schema of radiomic analysis for anatomical GBM phenotypes images.

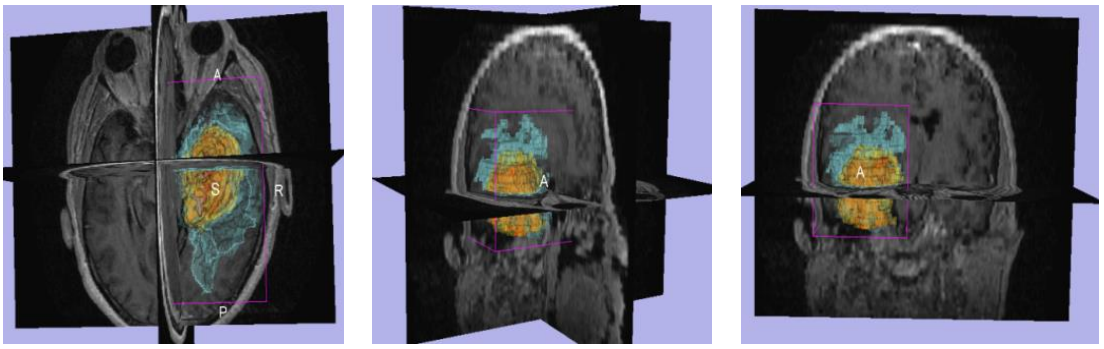
2.1 Data processing and acquisition

The data was acquired prior to any treatment, from subjects with brain tumors that were subsequently diagnosed as GBM. The diagnosis was based on histological examination. The data of each subject was then assessed visually using 3D Slicer, to determine if it had sufficient quality and contained the three phenotypes (N, CE and E).
The imaging protocol used whole-brain T1-WI and FLAIR sequences obtained from a 3T MRI scanner (GE-Healthcare). T1−WI scans were acquired based on the following parameters: slice thickness (ST) = 5mm, spatial resolution (SR) = 1.04mm, pixel spacing (PS) = 0.78 mm, repetition time (TR) = 650 ms, echo time (TE) = 9ms, and flip angle (FA) = 90°. Likewise, FLAIR scans were acquired using the following parameters: ST= 5mm, SR= 1.24mm, PS = 0.78 mm, TR = 10002 ms, TE = 147 ms, FA = 90° and acquisition time 10:24 min. Although other MRI sequences were available, only T1-WI and FLAIR were used for texture analysis, since these two sequences are most useful for highlighting GBM phenotypes. Acquired images had a resolution of 512 × 512 pixels and were converted to grayscale before further

processing. Note that the standard imaging parameters were used for each of the sequences as noted in the TCIA database. Subsequently, the registration, segmentation and texture feature extraction steps of the proposed workflow were applied to the images.

2.2 Segmentation

We segmented the phenotypes after employing intensity-based registration to align fixed T1-WI post contrast images with their corresponding FLAIR images. Note that this registration step is necessary since the T1-WI and FLAIR images are not aligned in the original data. Phenotype regions were then segmented and labeled manually in a slice-by-slice fashion, and used for texture feature extraction (see Figure 2). Texture features were computed on each of the phenotypes, from both T1-WI and FLAIR images.



(a) (b) (c)

Figure 2. Segmentation of GBM phenotypes where necrosis, contrast enhancement and edema are labeled by the red, yellow and cyan colors respectively. (a) Axial, (b) Sagittal and (c) Coronal anatomical plane.

2.3 Texture feature extraction

Gray-level co-occurrence matrix (GLCM) texture features were extracted from GBM phenotype regions in each of the three anatomical planes (axial, coronal and sagittal). GLCM measures second order statistics based on the joint co-occurrence of intensity values for pairs of pixels at a given distance d (or offset) and direction θ [18,20]. Typically, the co-occurrence matrix $P_{d,\theta}(i,j)$ of a two-dimensional (2D) image I of size $N \times N$ is defined as

$$P_{d,\theta}(i,j) = \sum_{x=1}^{N}\sum_{y=1}^{N}\begin{cases}1, & if\ I(x,y) = i\ \wedge\ I(x+dx,\ y+dy) = j \\ 0 & otherwise\end{cases} \quad (1)$$

where dx and dy specify the distance between the pixel of interest and its neighbor, along the x and y axis respectively. Resulting matrices have a size of $Ng \times Ng$, where Ng represents the number of gray levels in the image (Ng = 8 was used in this work).

In this study, we considered an offset of a single pixel and 8 different angles (0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°), leading to 8 GLCM matrices for each phenotype sample. These matrices were then summarized using 12 quantifier functions, as proposed by Haralick[18]: energy (f_1), entropy (f_2), contrast (f_3), homogeneity (f_4), correlation (f_5), cluster shade (f_6), variance (f_7), average (f_8), inertia (f_9), cluster tendency (f_{10}), maximum probability (f_{11}), and inverse variance (f_{12}). Applying these functions on each GLCM matrix, we obtained a total of 96 texture features for each phenotype region. This methodology was repeated for each anatomical plane, and the final texture features were obtained by averaging the features of the three planes. The GLCM texture features can be defined as the following vector

$$v_{GLCM} = [\, f_1\ f_2\ f_3\ f_4\ f_5\ f_6\ f_7 f_8\ f_9\ f_{10}\ f_{11}\ f_{12}] \quad (2)$$

2.4 Phenotype classification

The GLCM features extracted from each region were used to train classifiers for differentiating between GBM phenotypes.

Four classifier techniques from Matlab's *Statistics and Machine Learning Toolbox* were investigated, namely, linear discriminant analysis (LDA) [21], naïve Bayes (NB) [22], decision trees (DT) [23], and nearest neighbors (KNN) [24]. Classifier-

specific implementation details are as follows. In LDA, the probability distribution functions (PDF) of classes are assumed to be multivariate Gaussian with different mean but same covariance matrix, and Bayes classification is used to select the class with maximum probability for each test sample. Naïve Bayes considers features as independent from one another, given their class, and computes the PDF parameters of these features for each class. A univariate Gaussian PDF is assumed for all features. DT classification splits the set of training samples recursively, by applying a threshold on selected features, until all leaf nodes are sufficiently pure (i.e., they contain samples of the same class) or a maximum number of levels is attained. In this work, the Gini index was used as measure of purity. Finally, KNN finds the k training samples nearest to a given test sample, based on the Euclidean distance, and assigns the test sample to the most frequent class of its nearest neighbors. Based on prior testing, a value of k=10 was used in our experiments.

Since the number of samples is limited (41 GBM subjects), a leave-one-out cross validation methodology was adopted to obtain unbiased estimates of classification performance. This performance was measured using the accuracy, sensitivity and specificity metrics, according to the following equations:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

where TP and TN are the number of correctly classified positive (true positive) and negative (true negative) samples, and FP and FN correspond to the number of samples incorrectly classified as positive (false positive) and negative (false negative). Sensitivity quantifies the capability of a classifier to recognize positive samples, and is computed as

$$Sensitivity = \frac{TP}{TP+FN} \quad (4)$$

Likewise, specificity quantifies the capability of a classifier to recognize negative samples, as is computed as

$$Specificity = \frac{TN}{TN+FP} \quad (5)$$

Finally, confusion matrices and area under the receiver operating characteristic curve (AUC) values were computed to further quantify the accuracy of our methods. Note that a one-versus-all setting was used, where the samples of each class/phenotype (positives) are compared against the samples of the two other classes/phenotypes (negatives).

3. EXPERIMENTAL RESULT

The proposed technique was evaluated on the classification of regions into three GBM phenotypes: necrosis (N), contrast enhancement (CE) and edema (E). Table 1 gives the average accuracy, sensitivity and specificity obtained by the four classifiers in the leave-one-out cross-validation. The highest classification results were obtained by LDA using texture features derived from T1-WI images, with an accuracy of 88.14%, a sensitivity of 85.37 %, and a specificity of 96.10%. In comparison, DT, KNN and NB achieved accuracies of 79.66%, 73.73% and 70.34%, respectively, using the same images. Overall, GLCM features extracted from T1-WI images lead to a higher classification performance than FLAIR images.

Table 2 shows the confusion matrix obtained by the four classifiers using T1-WI and FLAIR images. For necrosis and edema, the best results were obtained using LDA and T1-WI images, with 35 of the 36 of the necrosis samples, and 34 of the 41 edema samples correctly classified. On the other hand, KNN obtained the best performance for contrast enhancement, with 36 out of 41 correctly classified samples. For T1-WI images, necrosis samples obtained the highest classification rates (97.3% with LDA) and most errors occurred between edema and contrast enhancement samples.

Table 1: Average classification performance obtained by the tested classifiers using texture features computed on T1-WI and FLAIR images.

Classifiers	% Accuracy		% Sensitivity		% Specificity	
	T1-WI	FLAIR	T1-WI	FLAIR	T1-WI	FLAIR
LDA	**88.14**	62.71	85.37	56.10	**96.10**	71.43
NB	70.34	55.93	78.05	58.54	70.13	68.83
DT	79.66	61.86	78.05	53.66	83.12	68.83
KNN	73.73	54.24	**87.80**	53.66	66.23	64.94

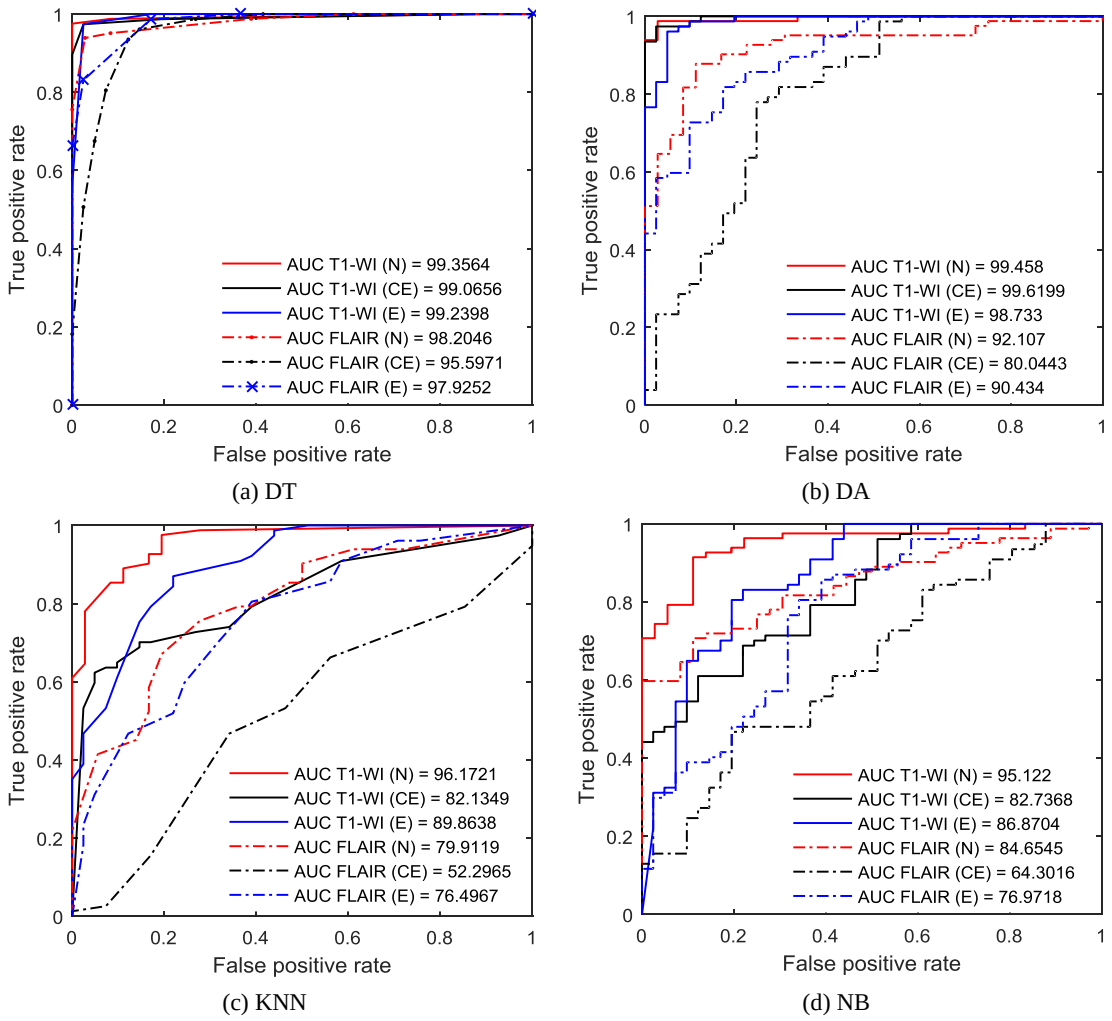


Figure 3. ROC curves and AUC values for classifiers (DT, LDA, KNN, NB), tissue phenotypes (N, CE, E) and MR sequences (T1-W, FLAIR).

Table 2. Summary of confusion matrix for GBM phenotype classification

	LDA						NB					
	CE		E		N		CE		E		N	
	T	F	T	F	T	F	T	F	T	F	T	F
41CE	**35**	**23**	2	11	4	7	**32**	**24**	0	1	9	16
41 E	3	13	**34**	**25**	4	3	19	19	**19**	**11**	3	11
36 N	0	9	1	1	**35**	**26**	4	5	0	0	**32**	**31**
	DT						KNN					
	CE		E		N		CE		E		N	
	T	F	T	F	T	F	T	F	T	F	T	F
41 CE	**32**	**22**	6	10	3	9	**36**	**22**	2	8	3	11
41 E	9	16	**31**	**24**	1	1	19	16	**22**	**21**	0	4
36 N	4	8	1	1	**31**	**27**	7	11	0	4	**29**	**21**

CE: Contrast Enhancement/active tumor; E: Edema; N: Necrosis; T:T1-WI; F:FLAIR

A similar pattern is observed for FLAIR images. While FLAIR images are usually preferred for the visualization of edema, our results suggest that texture features obtained from T1-WI images are more useful for recognizing this phenotype. This observation is further validated in Figure 3, which shows the ROC curves and AUC values across

classifier types, texture features and GBM phenotypes. Once again, we see that features extracted T1-WI images provide a better performance for all phenotypes, including edema.

4. CONCLUSIONS

We presented a novel method for discriminating between GBM tissue phenotypes, that uses GLCM texture features extracted from both T1-WI and FLAIR images. Four classifier models, based on linear discriminant analysis (LDA), decision trees (DT), naïve Bayes (NB) and neighbor neighbors (KNN), were evaluated on the data of 41 GBM subjects, using a leave-one-out approach for performance evaluation. Preliminarily results demonstrate the usefulness of GLCM features for phenotype prediction, with an overall accuracy of 88.14% obtained by LDA. Our experiments also show the advantages of texture features extracted from T1-WI images, even for edema regions which are usually detected using FLAIR images.

This study is part of larger project focusing on radiomic analysis for assessing GBM tissue phenotypes in terms of texture features, particularly texture heterogeneity [19,25,26]. Future work will focus on using such features for analyzing the overall survival of GBM patients, and will investigate additional types of information, such as 3D gradient orientation histograms obtained from 3D SIFT-Rank features[27].

REFERENCES

[1] Greenlee, R. T., Murray, T., Bolden, S.., Wingo, P. A., "Cancer statistics, 2000," CA. Cancer J. Clin. **50**(1), 7–33 (2000).
[2] McGirt, M. J., Chaichana, K. L., Gathinji, M., Attenello, F. J., Than, K., Olivi, A., Weingart, J. D., Brem, H.., Quiñones-Hinojosa, A. R., "Independent association of extent of resection with survival in patients with malignant brain astrocytoma," J. Neurosurg. **110**(1), 156–162 (2009).
[3] Stupp, R., Hegi, M. E., Mason, W. P., van den Bent, M. J., Taphoorn, M. J. B., Janzer, R. C., Ludwin, S. K., Allgeier, A., Fisher, B., et al., "Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial," Lancet Oncol. **10**(5), 459–466 (2009).
[4] Leemput, K. V., "Brain MRI Segmentation Using an Expectation-Maximization Algorithm."
[5] Frieboes, H. B., Zheng, X., Sun, C.-H., Tromberg, B., Gatenby, R.., Cristini, V., "An integrated computational/experimental model of tumor invasion," Cancer Res. **66**(3), 1597–1604 (2006).
[6] Cancer Genome Atlas Research Network., "Comprehensive genomic characterization defines human glioblastoma genes and core pathways," Nature **455**(7216), 1061–1068 (2008).
[7] Stupp, R., Hegi, M. E., Bent, M. J. van den., Mason, W. P., Weller, M., Mirimanoff, R. O.., Cairncross, J. G., "Changing Paradigms—An Update on the Multidisciplinary Management of Malignant Glioma," The Oncologist **11**(2), 165–180 (2006).
[8] Verhaak, R. G. W., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., Miller, C. R., Ding, L., Golub, T., et al., "Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1," Cancer Cell **17**(1), 98–110 (2010).
[9] Chaddad, A., Zinn, P. O.., Colen, R. R., "Quantitative texture analysis for Glioblastoma phenotypes discrimination," 2014 Int. Conf. Control Decis. Inf. Technol. CoDIT, 605–608 (2014).
[10] Chaddad, A., "Automated Feature Extraction in Brain Tumor by Magnetic Resonance Imaging Using Gaussian Mixture Models," Int. J. Biomed. Imaging **2015**, e868031 (2015).
[11] Downey, K., Riches, S. F., Morgan, V. A., Giles, S. L., Attygalle, A. D., Ind, T. E., Barton, D. P. J., Shepherd, J. H.., deSouza, N. M., "Relationship between imaging biomarkers of stage I cervical cancer and poor-prognosis histologic features: quantitative histogram analysis of diffusion-weighted MR images," AJR Am. J. Roentgenol. **200**(2), 314–320 (2013).
[12] Levner, I., Drabycz, S., Roldan, G., De Robles, P., Cairncross, J. G.., Mitchell, R., "Predicting MGMT methylation status of glioblastomas from MRI texture," Med. Image Comput. Comput.-Assist. Interv. MICCAI Int. Conf. Med. Image Comput. Comput.-Assist. Interv. **12**(Pt 2), 522–530 (2009).
[13] Castellano, G., Bonilha, L., Li, L. M.., Cendes, F., "Texture analysis of medical images," Clin. Radiol. **59**(12), 1061–1069 (2004).

[14] Al-Kadi, O. S.., Watson, D., "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," IEEE Trans. Biomed. Eng. **55**(7), 1822–1830 (2008).
[15] Brown, R. A.., Frayne, R., "A comparison of texture quantification techniques based on the Fourier and S transforms," Med. Phys. **35**(11), 4998–5008 (2008).
[16] Davnall, F., Yip, C. S. P., Ljungqvist, G., Selmi, M., Ng, F., Sanghera, B., Ganeshan, B., Miles, K. A., Cook, G. J., et al., "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?," Insights Imaging **3**(6), 573–589 (2012).
[17] Chaddad, A., Zinn, P. O.., Colen, R. R., "Brain tumor identification using Gaussian Mixture Model features and Decision Trees classifier," 2014 48th Annu. Conf. Inf. Sci. Syst. CISS, 1–4 (2014).
[18] Haralick, R. M., Shanmugam, K.., Dinstein, I., "Textural Features for Image Classification," IEEE Trans. Syst. Man Cybern. **SMC-3**(6), 610–621 (1973).
[19] Chaddad, A., Zinn, P. O.., Colen, R. R., "Radiomics texture feature extraction for characterizing GBM phenotypes using GLCM," Biomed. Imaging ISBI 2015 IEEE 12th Int. Symp. On, 84–87, IEEE (2015).
[20] Chaddad, A., Tanougast, C., Dandache, A.., Bouridane, A., "Extracted haralick's texture features and morphological parameters from segmented multispectrale texture bio-images for classification of colon cancer cells," WSEAS Trans. Biol. Biomed. **8**(2), 39–50 (2011).
[21] Guo, Y., Hastie, T.., Tibshirani, R., "Regularized linear discriminant analysis and its application in microarrays," Biostatistics **8**(1), 86–100 (2007).
[22] Aggarwal, C. C., Data Classification: Algorithms and Applications, CRC Press (2014).
[23] Rokach, L., Data Mining with Decision Trees: Theory and Applications, World Scientific (2007).
[24] Dasarathy, B. V., Nearest neighbor (NN) norms: nn pattern classification techniques, IEEE Computer Society Press (1991).
[25] Chaddad, A.., Tanougast, C., "High-Throughput Quantification of Phenotype Heterogeneity Using Statistical Features," Adv. Bioinforma. **2015**, e728164 (2015).
[26] Chaddad, A.., Colen, R. R., "Statistical feature selection for enhanced detection of brain tumor," SPIE Opt. Eng. Appl., 92170V – 92170V, International Society for Optics and Photonics (2014).
[27] Toews, M.., Wells III, W. M., "Efficient and robust model-to-image alignment using 3D scale-invariant features," Med. Image Anal. **17**(3), 271–282 (2013).