Computer Methods and Programs in Biomedicine 140 (2017) 249–257

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Learning MRI-based classification models for MGMT methylation status prediction in glioblastoma

Vasileios G. Kanas[a,b], Evangelia I. Zacharaki[b,c,*], Ginu A. Thomas[d], Pascal O. Zinn[e], Vasileios Megalooikonomou[b], Rivka R. Colen[d]

[a] Department of Electrical and Computer Engineering, University of Patras, Patras, Greece
[b] Department of Computer Engineering and Informatics, University of Patras, Patras, Greece
[c] Center for Visual Computing (CVC), CentraleSupélec, INRIA, Université Paris-Saclay, France
[d] Department of Diagnostic Radiology, University of Texas MD Anderson Cancer Center, Houston, TX, USA
[e] Department of Neurosurgery, Baylor College of Medicine, Houston, TX, USA

ARTICLE INFO

Article history:
Received 28 April 2016
Revised 14 December 2016
Accepted 29 December 2016

Keywords:
MGMT promoter methylation
Glioblastoma
Feature extraction
Multivariate analysis
Prediction model

ABSTRACT

Background and Objective: The O^6-methylguanine-DNA-methyltransferase (*MGMT*) promoter methylation has been shown to be associated with improved outcomes in patients with glioblastoma (GBM) and may be a predictive marker of sensitivity to chemotherapy. However, determination of the *MGMT* promoter methylation status requires tissue obtained via surgical resection or biopsy. The aim of this study was to assess the ability of quantitative and qualitative imaging variables in predicting *MGMT* methylation status noninvasively.

Methods: A retrospective analysis of MR images from GBM patients was conducted. Multivariate prediction models were obtained by machine-learning methods and tested on data from The Cancer Genome Atlas (TCGA) database.

Results: The status of *MGMT* promoter methylation was predicted with an accuracy of up to 73.6%. Experimental analysis showed that the edema/necrosis volume ratio, tumor/necrosis volume ratio, edema volume, and tumor location and enhancement characteristics were the most significant variables in respect to the status of *MGMT* promoter methylation in GBM.

Conclusions: The obtained results provide further evidence of an association between standard preoperative MRI variables and *MGMT* methylation status in GBM.

1. Introduction

High-grade brain neoplasms, such as glioblastoma (GBM), are among the most aggressive and difficult-to-treat forms of cancer. Temozolomide, a DNA alkylating agent, is the standard chemotherapeutic agent approved by the U.S. Food and Drug Administration for first-line treatment of GBM. Temozolomide causes alkylation at the O^6 guanine position of DNA and induces cytotoxic effects and apoptosis in cancer cells. Recent molecular studies have demonstrated that the methylation status of the O^6 methylguanine-DNA methyltransferase (MGMT) gene promoter is a predictor of response to temozolomide and a prognostic indicator of survival time in patients with temozolomide-treated GBM [1-6]. *MGMT* codes for a DNA repair protein that removes methyl groups from the O^6 guanine of DNA, thereby preventing degradation of DNA and cytotoxic effects induced by temozolomide [7].

Determination of the *MGMT* promoter methylation status requires obtaining a tissue sample via surgical resection or biopsy. Though biopsy is the gold-standard assessment, there are some limitations. Accurate characterization of methylation status is often not feasible with small tissue specimens. Moreover, assessment of surgical biopsy samples is prone to inter- and intra-rater variability due to GBM heterogeneity, which often results in undergrading of tumors [8]. Consequently, a noninvasive and reliable surrogate method of determining *MGMT* status could serve as an alternative (or a complement) to biopsy. Analysis of MRI tumor characteristics is one such potential alternative. Moreover, in machine learning approaches, it is possible, during the learning phase of the prediction models, to carefully select the training examples or minimize discrepancies by sampling the heterogeneous tumors at multiple locations, and thus avoiding errors in genetic profiling. Upon learning, in the testing phase, genetic data are not anymore required.

* Corresponding author.
E-mail address: evangelia.zacharaki@centralesupelec.fr (E.I. Zacharaki).

http://dx.doi.org/10.1016/j.cmpb.2016.12.018
0169-2607/

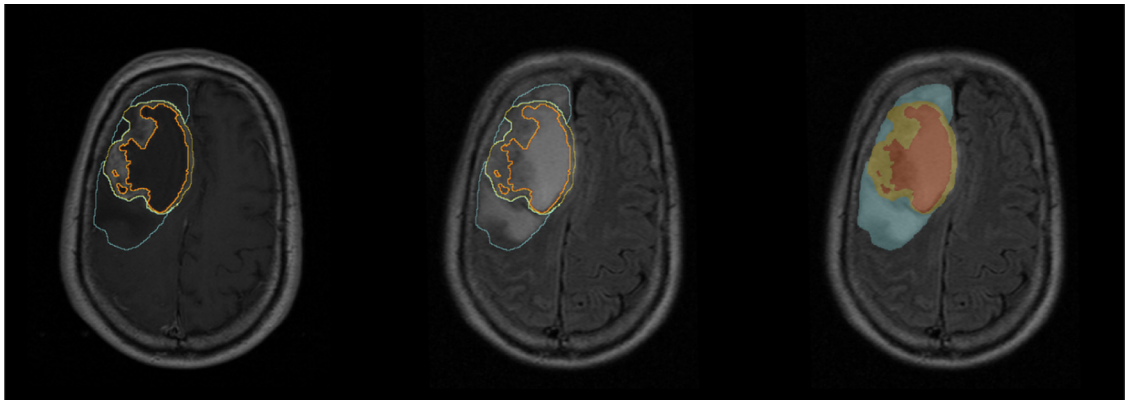

Fig. 1. MR images for a 59-year-old male patient with a right frontal lobe GBM. Left: Edema/invasion region (in blue) as the region of FLAIR hyperintensity in an axial image; Middle: enhancement (yellow) and necrosis (red) in a T1 post-contrast image; Right: label map of edema/invasion (blue), enhancement (orange), and necrosis (red), obtained by overlaying a registered FLAIR image on a T1 post-contrast image (base). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

One of the first studies to correlate MRI-based volumetric phenotypes with large-scale gene and microRNA expression profiles was conducted by Zinn et al. [9]. The data showed that *POSTN* (periostin) was upregulated and microRNA 219 (predicted to bind to periostin) was downregulated in GBMs whose signal intensity was high in fluid-attenuated inversion recovery (FLAIR) images. The findings also suggested that microRNA 219 and periostin are regulators of cellular invasion/edema. In another study, Zinn et al. [10] proposed a robust GBM classification system, incorporating preoperative MRI-based tumor volume along with patient age and Karnofsky score [11], for predicting survival before performing surgery or another invasive procedure.

MRI tumor variables such as edema/invasion volume, necrosis volume, enhancement, cysts, multifocality, and location, as visually assessed by expert radiologists, have been used in various studies [12-14] to detect GBM phenotypic signatures associated with molecular profiles and patient survival. In particular, associations between *MGMT* methylation status and imaging parameters have been reported [15,16]. Eoli et al. [15] suggested that *MGMT* methylation status is associated with specific tumoral locations and enhancement patterns (as assessed by MRI). In a similar study of high-grade (grade 3 and grade 4) gliomas, Moon et al. [16] correlated *MGMT* methylation status with variables measured by computed tomography and advanced MRI techniques like diffusion tensor imaging and T2* dynamic susceptibility contrast-enhanced perfusion-weighted imaging.

There have been only few reports on predicting *MGMT* methylation status in GBM patients from MRI data. In both studies [17,58] authors used a space frequency texture analysis to predict *MGMT* methylation status from MR images. To the best of our knowledge, that were the first investigations of the potential of MR 3D volumetric to predict *MGMT* methylation status. In a more recent study, Korfiatis et al. [57] extracted run length matrix (RLM)-based texture features to capture the variability of image intensity utilizing a dataset of 155 preoperative MRI examinations of GBM patients. In clinical practice, volumetric analysis remains a difficult task, and clinically relevant MRI variables are typically assessed qualitatively. In the present study, we hypothesized that certain variables derived from standard MRI sequences reflect differences in *MGMT* promoter methylation status in GBM patients. We therefore sought to evaluate the significance of MR 3D volumetric and qualitative imaging variables for predicting *MGMT* methylation status in GBM variable.

2. Methods

2.1. Data collection

Eighty-six treatment-naive GBM patients (27 women and 59 men; mean age, 58 years; age range, 14–84 years) with gene expression data available in The Cancer Genome Atlas (TCGA) and corresponding pretreatment MRI data available in The Cancer Imaging Archive (TCIA) were included in the study [18]. The collection of data complied with all applicable laws [19], including the Health Insurance Portability and Accountability Act, and Institutional Review Board approval was obtained. Because TCGA and TCIA are publicly available databases that contain no linkage to patient identification, the requirement for informed consent was waived. The MR sequences analyzed in this study were axial T1-weighted, axial T2-weighted FLAIR, and axial T1-weighted contrast-enhanced images. Fig. 1 shows a sample data set.

Genomic data had been obtained with Illumina DNA methylation probes and Illumina HumanMethylation27 and HumanMethylation450 BeadChip platforms. *MGMT* methylation status was determined using level 3 beta values for Illumina DNA methylation probes MGMT_P272_R and MGMT_P281_F [10]. The median average value for the two probes was used as the cutoff for categorizing methylation status. For patients with HumanMethylation27 and HumanMethylation450 data, *MGMT* methylation status was determined using probe cg12434587 and cg12981137 values in Bady et al. [20].

2.2. Variable extraction

2.2.1. Qualitative variables

For each patient, three board-certified neuroradiologists (blinded to the MGMT status of the tumor) independently reviewed pre- and post-contrast axial T1-weighted MR images as well as axial T2-weighted FLAIR images. Images were analyzed using the ClearCanvas (Toronto, Canada) platform (http://www.clearcanvas.ca/), and each neuroradiologist recorded a set of mark-ups for qualitative imaging variables (based on the VASARI variable set for GBM) [21-23] describing the size, location, and morphology of the tumoral region. For each tumor, the qualitative imaging assessments by the three neuroradiologists were concatenated into a single set, and the most frequent category in each set (for each variable and patient) was recorded as the consensus value. The raters categorized 24 qualitative variables (Table 1) according to the scoring guidelines they were given.