Computers in Biology and Medicine 73 (2016) 94–101

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Pseudo progression identification of glioblastoma with dictionary learning

Jian Zhang [a], Hengyong Yu [b], Xiaohua Qian [a], Keqin Liu [a], Hua Tan [a], Tielin Yang [c], Maode Wang [d], King Chuen Li [a], Michael D. Chan [e], Waldemar Debinski [e], Anna Paulsson [e], Ge Wang [f], Xiaobo Zhou [a,*]

[a] Department of Diagnostic Radiology, Wake Forest Medical School, Winston-Salem, NC 27103, USA
[b] Department of Electrical and Computer Engineering, University of Masachusetts Lowell, Lowell, MA 01854, USA
[c] School of Life Science, Xi'an Jiaotong University, Xi'an, Shanxi 710049, China
[d] The First Affiliated Hospital, Xi'an Jiaotong University, Xi'an, Shanxi 710049, China
[e] Department of Radiation Oncology, Wake Forest Medical School, Winston-Salem, NC 27103, USA
[f] Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

ARTICLE INFO

Article history:
Received 16 December 2015
Received in revised form
28 March 2016
Accepted 30 March 2016

Keywords:
Pseudo progression
Glioblastoma multiforme
Dictionary learning
Genomics analysis
Fractional anisotropy (FA)
Diffusion tensor imaging (DTI)

ABSTRACT

Objective: Although the use of temozolomide in chemoradiotherapy is effective, the challenging clinical problem of pseudo progression has been raised in brain tumor treatment. This study aims to distinguish pseudo progression from true progression.

Materials and Methods: Between 2000 and 2012, a total of 161 patients with glioblastoma multiforme (GBM) were treated with chemoradiotherapy at our hospital. Among the patients, 79 had their diffusion tensor imaging (DTI) data acquired at the earliest diagnosed date of pseudo progression or true progression, and 23 had both DTI data and genomic data. Clinical records of all patients were kept in good condition. Volumetric fractional anisotropy (FA) images obtained from the DTI data were decomposed into a sequence of sparse representations. Then, a feature selection algorithm was applied to extract the critical features from the feature matrix to reduce the size of the feature matrix and to improve the classification accuracy.

Results: The proposed approach was validated using the 79 samples with clinical DTI data. Satisfactory results were obtained under different experimental conditions. The area under the receiver operating characteristic (ROC) curve (AUC) was 0.87 for a given dictionary with 1024 atoms. For the subgroup of 23 samples, genomics data analysis was also performed. Results implied further perspective on pseudo progression classification.

Conclusions: The proposed method can determine pseudo progression and true progression with improved accuracy. Laboring segmentation is no longer necessary because this skillfully designed method is not sensitive to tumor location.

1. Introduction

Glioblastoma multiforme (GBM) is a highly invasive disease, and its recurrence at sites distant from the core of tumors is the main cause of its high mortality. The current standard of care is surgical resection, followed by radiation therapy (RT) and concomitant and adjuvant temozolomide (TMZ) chemotherapy [1].

Pseudo progression (PsP) is a phenomenon showing changes of subacute imaging data in human glioma subsequent to chemoradiotherapy suggestive of progression [2]. The incidence of tumor PsP ranges from 28% to 66% in all GBM patients undergoing chemoradiation and typically occurs within three months after the completion of concurrent radiation and TMZ chemotherapy [3]. PsP has been associated with a variety of nontumoral processes, such as treatment-related inflammation [1], postsurgical changes, ischemia, and radiation necrosis [4,5].

PsP is detected by magnetic resonance imaging (MRI) as contrast-enhanced regions that can mimic early tumor progression. However, PsP is difficult to identify by conventional MRI and morphologic MRI features [1,6,7]. Currently, the only method to distinguish PsP from early progression is by follow-up examinations on patients [2]. Differentiation between PsP and tumor progression is critical because PsP is an indication of that the current therapy is effective and should be continued, in which

* Corresponding author.
E-mail address: xizhou@wakehealth.edu (X. Zhou).

http://dx.doi.org/10.1016/j.compbiomed.2016.03.027
0010-4825/

group the patients have a better chance of survival, whereas tumor progression indicates that the current therapy is not effective and should be altered if possible. Developing advanced techniques and approaches to distinguish between PsP and true progression in GBM is important [2,6].

Advanced MRI techniques, such as perfusion imaging, diffusion tensor imaging (DTI), and magnetic resonance spectroscopy (MRS), have been introduced into the GBM treatment response analysis [8,9]. Different modalities of MRI can provide additional information and novel aspects to improve the distinction. For example, DTI is based on the movement of water molecules that diffuse more rapidly in the direction aligned with the internal structure. Different metrics, such as apparent diffusion coefficient (ADC) and fractional anisotropy (FA) [10], can be computed from DTI, which characterizes tissue integrity and describes the physical attributes of neuronal fibers in the brain. In [8], these values were observed to be significantly different in regions having or close to pathological changes. Therefore, these metrics can potentially be used for distinguishing PsP from early tumor recurrence.

However, previous studies mostly focused on extracting biofeatures or morphologic features from the scans and identifying the presence of PsP by comparing the value of the features with certain discriminative criteria [1,11]. The performance of such approaches cannot meet the precision required for the clinical management of patients. The authors were perplexed about the tumor progression and stated that advanced techniques are necessary.

Consequently, researchers have started to apply machine learning approaches [12,13], such as support vector machine (SVM) [9], to improve the analysis of perfusion images. Eight parameters derived from multiple MRI sequences are used to compose the feature vector for classification. This approach directly uses the extracted morphological features without further analysis. The region-of-interest (ROI) should be segmented for each set of MRI sequence beforehand, but this step is labor intensive and introduces much ambiguity. A more complicated tumor growth model was developed as the metric of therapy response [14]. A score called Days Gained was computed by this model to distinguish PsP from true progression. This approach requires follow up of image data of patients. Segmentation cannot be avoided either. Generally, improvements to these approaches are greatly needed, particularly with the use of advanced analysis.

This study focuses on DTI sequence and uses extracted volumetric FA images as the key measurement. The dictionary learning (DL) approach is applied to the image sequence to obtain sparse representations. DL is a powerful image analysis method based on sparse representation techniques. DL methods have various forms derived from different application problems, such as image denoising [15], image in-painting [16], and image compression [17,18]. However, most of these attempts are based on two-dimensional (2D) images. In the present study, we extend this approach to discriminate volumetric subjects. ROI segmentation is no longer needed in the DL approach. Therefore, the proposed method is more efficient and accurate. Then, the feature matrix is composed by the sparse representation with a specially designed permutation, which could solve the location divergence of tumors. Finally, a small group of features is selected from the feature matrix, and the subjects are classified using the SVM approach.

2. Material and methods

Between 2000 and 2012, a total of 161 patients [19] with GBM were treated with chemoradiotherapy after surgical resection at Wake Forest Baptist Medical Center. Along with chemoradiotherapy, the patients underwent MRI/DTI scans every two or three months after each stage of therapy for monitoring. MRI/DTI data; treatment schemes, such as date of surgery, type of surgery, RT dose, concurrent chemotherapy, concurrent chemotherapy agents, and patient information, such as date of birth, date of death, ethnicity, and gender, were fully available. Different treatment schemes could lead to different therapy results, including PsP, which could be identified by the proposed image analysis approach. However, the image analysis approach itself is essentially irrelevant to the treatment schemes. The doctor's diagnosis of PsP was made on the basis of clinical and imaging data and the clinical experience of the treating physician. Biopsy could be performed when necessary. For each patient, we selected the DTI sequence that was obtained when the patient was first diagnosed with PsP or true progression. Both PsP and true progression were assumed to have just occurred, and the time was the earliest for diagnosis by using imaging. Under this condition, 79 samples with qualified DTI data were selected and used, including 23 pseudo and 56 true progression samples, which were all collected after 2005. Among 79 samples, 23 samples possessed genomic data. The composition of the two datasets is shown in Table 1. The study protocol was approved by the Wake Forest Baptist Medical Center review board.

Table 1
Structure of the two datasets. Dataset 1 consists of patients with both genomics and imaging data. Dataset 2 consists of patients with imaging data. Dataset 1 is a subset of Dataset 2.

	Dataset 1	Dataset 2
Number of samples	23	79
Number of pseudo progression	9	23
Number of true progression	14	56

2.1. Pre-processing

The DTI data sequences were processed with MRIcron [20] and FSL [21,22]. Each DTI sequence of the DICOM images was transformed into a 4D image with the gradient information using the MRIcron. Then, the FSL software was used to correct for eddy current-induced distortions and subject movements. Afterwards, non-brain tissue such as skin was removed [23]. The FA value of each voxel was computed by the FSL. Each reconstructed brain volume was linearly registered to the standard brain template FMRIB58 [24]. As a result, all the registered volumetric FA images had the same number of slices and the same resolution. Thus, the same voxels across different patients were properly registered to one another. Fig. 1 shows two randomly selected samples with one slice of the DTI raw data, the FA image, and the registered FA image, respectively. The figure also shows that the position and head shape are registered accurately.

2.2. Dictionary learning and sparse representation

A dictionary is a matrix $D \in R^{N\times K}$, where R is the real space, and N and K are integers. One column of D is called an atom, which is an N dimensional vector, denoted as $d_k \in R^{N\times 1}$ $(k = 1, \ldots, K)$. The dictionary is redundant or over-complete if $N \ll K$. Let us express an $\sqrt{N} \times \sqrt{N}$ image patch by a vector $x \in R^{N\times 1}$, where each element is equal to the gray value of one pixel of the patch. Given a dictionary D, we suppose a sparse linear combination of the atoms d_k that can represent the image patch x. Hence, we have

$$\| x - D\alpha \|_2^2 \leq \varepsilon \quad (1)$$

where $\varepsilon > 0$ is a small error bound. $\alpha \in R^{K\times 1}$ is the representation vector and possesses few nonzero entries. We have $\| \alpha \|_0 \ll N \ll K$,