HARALICK’S TEXTURE ANALYSIS APPLIED TO COLORECTAL T2-WEIGHTED MRI: A PRELIMINARY STUDY OF SIGNIFICANCE FOR CANCER EVOLUTION

Mumtaz Hussain Soomro¹, Gaetano Giunta¹, Andrea Laghi², Damiano Caruso², Maria Ciolina², Cristiano De Marchis¹, Silvia Conforto¹, Maurizio Schmid¹

¹Department of Engineering, University of Roma Tre, Rome, Italy
²Department of Radiological, Oncological and Pathological Sciences, University of Rome La Sapienza, Rome, Italy

{mumtazhussain.soomro, gaetano.giunta, cristiano.demarchis, silvia.conforto, maurizio.schmid}@uniroma3.it,
{andrea.laghi, damiano.caruso, maria.ciolina}@uniroma1.it

ABSTRACT
Haralick’s features have been extensively used in texture analysis of medical images. In this contribution, we have applied Haralick’s to T2-weighted colorectal MRI for a possible cancer evaluation. In particular, the T2-MRI images of 8 patients with colorectal pathology were identified as early stage malignant and later stage malignant using the whole amount of follow-up exams by radiologists. 192 Haralick’s textural features were computed from normalized gray level co-occurrence matrix with respect to four different directions. Mean and standard deviation were also calculated for the extracted features to assess the statistical significance of results. Among all the extracted features, only 5 from 14 Haralick’s textural features (viz. energy, contrast, correlation, entropy and inverse difference moment (IDM)) were found as significant for colorectal cancer evaluation. In future research, these five Haralick’s textural features may be useful to detect and evaluate colorectal cancer as well as constitute a basis for predicting the prognostic trend of the disease.

KEY WORDS
Medical imaging, T2-weighted MRI, Colorectal cancer evaluation, Haralick’s texture analysis.

1. Introduction
Colon/rectum or colorectal is considered as fundamental part of large intestine which starts from small intestine to anus. It helps to break large molecules into ions and nutrients and to absorb the water [1-2]. About half of million people die due to colorectal cancer every year. Colorectal cancer is diagnosed as the third most common cancer in men and the second most common in women [3]. The risk of getting colorectal cancer increases in people above 50 years age. There are others risk factors are also for colorectal cancer including heavy intake of fats and red meat, chain smoking, a family history of colorectal cancer and colon polyps [4].

Early detection of colorectal cancer is supreme interest make it guaranteed for treatment. Conventionally, microscopic histopathological colorectal samples were used to detect cancer and changes in tissues were observed visually to grade the colorectal cancer. This visual inspection of pathologists is restricted to quantitative measurements which are essential to detect and stage the cancer. Therefore, this visual inspection is time consuming and presents human eye error [5].

Taking into account the weaknesses in visual process, there exist a need of accurate automatic colorectal cancer diagnostics techniques [6]. In this regard, medical imaging techniques such as CT scan, PET, SPET are considered as valuable tools to detect and localize the colorectal cancer. Nevertheless, these tools are not able to detect early cancer at the cellular level [7-8].

Recent studies recommended that high resolution MRI can be suitable for diagnosis of colon cancer at early stage [9-10]. In this scenario, an automatic detection of colorectal cancer has four major directions: segmentation, feature extraction, feature selection and classification. In this work, an attempt is conducted to extract the features from colorectal MRI using Haralick’s textural analysis method [11] for colorectal cancer evaluation.

2. Material and Methods
2.1 Data Sets
T2-weighted MRI data have been collected from 8 patients with age 45-60 years. Such data are categorised into two aspects: either patient may undergo for pre-treatment or for surgical operation. We have equally divided data 4 patients with early stage malignancy and 4 with later stage malignancy and labeled them as GREEN and RED, respectively, as shown in figure 1.

2.2 Preprocessing
2.2.1 Intensity Non-uniformity Correction
There are many artifacts which may affect the MRI images quality, but the non-uniform intensity is the most
prominent artifact which lowers the performance of quantitative MRI analysis. Anatomically, intensity non-uniformity can be defined as irrelevant intensity variation throughout data [12]. We have used MIPAV software [13] to correct non-uniformity in our T2-MRI data.

2.2.2 Manual Segmentation
We are interested in only colon/rectum part and that part has been manually segmented from each MRI image. Haralick’s texture statistics have been computed on the segmented colon/rectum part including the tumoral tissue (see again figure 1).

2.3 Feature Extraction
We have tested the most popular Haralick’s textural analysis method [11] with our data. We have computed 14 different Haralick’s texture features including energy (F1), contrast (F2), sum of squares (F3), correlation (F4), sum average (F5), inverse difference moment or homogeneity (F6), entropy (F7), sum variance (F8), sum entropy (F9), difference variance (F10), difference entropy (F11), information measure correlation- 1 (F12), information measure correlation- 2 (F13) and maximum correlation (14). These statistics have been computed from normalized gray level co-occurrence matrix (GLC) of every segmented colon/rectum part including the tumoral tissue. GLC matrices have been computed with respect to four main directions (viz: 0, 45, 90 and 135 degrees). We have then considered a total set of 192 Haralick’s features which have been divided equally (after visual inspection) into GREEN and RED, i.e. 96 for each. The GLC matrices have been normalized by taking the neighbor pixel values and reference pixel values into account as suggested in [14]. The purpose of GLC matrix normalization has consisted of to paying attention to probability rather than just counting of co-occurrences.

2.4 Statistical Analysis
Mean and standard deviation have been used as statistical parameters to analyze the variation in all the extracted Haralick’s features.

3. Experimental Results
Our preliminary analysis on Haralick’s texture indices has shown that five (viz: energy, contrast, correlation, entropy, inverse difference moment) over 14 features could be beneficial in order to differentiate between the early stage malignant and the later stage malignant related texture in colorectal MRI. We have then focused on these five Haralick’s texture features. In particular, energy gives the information about uniformity of image. It assumes its value between 0 and 1. The highest value 1 denotes low variation in image with respect to intensity. Contrast measures local variation. Its large value indicates the higher intensity variation among pixels in the image. Correlation is able to capture the similarity among image gray levels. It assumes values between -1 and 1. Its higher value indicates that the image gray levels have higher

Figure 1  Flow chart of methodology, dividing the data into two categories by visual inspection of radiologists, manual segmentation and feature extraction.

17
linear dependencies. Entropy quantifies the randomness or disorder in image gray levels. It shows higher entropy values for more chaotic distribution of image gray levels. Finally, inverse difference moment (IDM) is also referred to as homogeneity, i.e. the inverse of contrast, and measures the similarity of co-occurrence gray levels.

Table 1 shows the five most significant Haralick’s features for early stage malignant MRI images while the results for later stage malignant MRI images are shown in Table 2. Table 1 reveals that energy, correlation and IDM have higher values while contrast and entropy have lower values. Conversely, Table 2 shows that entropy and contrast have higher values whereas rest have lower values. In fact, the average and the standard deviation over each of the four considered directions have been calculated for the above mentioned five features. Table 1 and 2 confirm that each computed Haralick’s feature has a similar behavior in all the directions (i.e. degree 0, degree 45, degree 90 and degree 135).

Finally, figure 2 shows the comparison between early stage malignancy (GREEN) and later stage malignancy (RED) after the average of the Haralick’s features over all the four directions.

4. Discussion

In fact, the T2-weighted colorectal MRI images with early stage malignancy (GREEN cases) have less disorder or randomness than images with high or later stage malignancy (RED cases). The results have shown that T2-MRI colorectal images with early stage malignancy (GREEN) have higher energy, IDM and correlation values, indicating uniformity in image, as well as lower entropy and contrast values, revealing that images have low randomness and dissimilarity in their grayscale levels. On the contrary, T2-MRI colorectal images with later stage malignancy (RED) have higher entropy and contrast, representing higher randomness or disorder and dissimilarity in image gray levels.

Our initial results have pointed out that energy, contrast, correlation, entropy and IDM are noteworthy Haralick’s features suited to characterize colorectal cancer with regard to its possible evolution. From our preliminary testing trials, we have found that energy and entropy are the most significant and powerful features among above mentioned five features, while energy can be significant if it is considered as a normalized parameter.

Nevertheless, the sample size used in this preliminary study has been limited to initially perform a training procedure to extract an appropriate number of features from MR images for perspective cancer evaluation. In future attempts of this work, we aim to validate the predictive model with a more large number of patients.

5. Conclusion

This contribution has presented a preliminary attempt on using Haralick’s textural features to possibly evaluate the prognostic trend of colorectal cancer by MRI semi-automatic image analysis. Haralick’s features are widely explored and commonly used in texture analysis of medical images. In this study of significance for cancer
evaluation, Haralick’s texture statistics were applied to T2-weighted colorectal MRI data collected from 8 patients. Total of 192 Haralick’s textural features were calculated from normalized gray level co-occurrence matrix with respect to four different directions. Moreover, in order to analyze the statistical variation in the extracted features, their mean and standard deviation were also computed. The preliminary results of this work using Haralick’s statistic measures motivate the authors to keep on toward improvements in current ongoing research in the next future.

References