PET Texture Analysis: Does it Have Clinical Significance in Locally Advanced Rectal Cancer?

P. Martin-Gonzalez [1], E. Gomez de Mariscal [1, 2], M.E. Martino [1,2], P.M. Gordaliza [1,2], J.L. Carreras [2, 3,4,] F.A. Calvo [2, 4, 5], J. Pascau [1, 2], M. Desco [1, 2, 6], A. Muñoz-Barrutia [1,2]

[1] Bioengineering and Aerospace Engineering Department, Universidad Carlos III de Madrid, Spain

[2] Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

[3] Department of Radiology and Medical Physics, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[4] School of Medicine, Universidad Complutense, Madrid, Spain

[5] Department of Oncology, Hospital General Universitario Gregorio Marañón, Madrid, Spain

[6] Centro de Investigación Biomédica en Red de Salud Mental (CIBERSAM), Madrid, Spain

Abstract

^{18}F-FDG PET/CT is a widely used tool in oncology for tumor staging and disease follow-up. The aim of this pilot study is to propose a method to automatically analyze heterogeneity from PET studies and evaluate its prognostic capacity. Thirty-seven patients diagnosed with locally advanced rectal cancer had an ^{18}F-FDG PET/CT study before undertaking chemoradiotherapy. From PET images, tumor region was segmented and metabolic parameters and texture parameters (global, local and regional) were extracted. Principal component analysis was performed to reduce the dimensionality. Multivariate binary logistic regression with cross-validation correction was applied to approximate the clinical relevance of the features. Areas under the ROC Curve (AUC) were used to compare the prediction obtained. Tumor regression grade was considered as the gold-standard. Principal components extracted from the standard parameters together with the textural features (explained variance of 92%) showed a poor assessment of response after cross-validation (AUC=0.525). The combination of individual metabolic and texture features that were statistically significant in the step-wise multivariate binary logistic regression raised the accuracy even after cross-validation correction (AUC=0.848). In conclusion, tumor texture as a measure of heterogeneity could be adding useful information to the standard parameters used routinely by the radiologists.

1. Introduction

Whereas updates in multimodality disease diagnosis and treatment have improved the outcome of cancer patients over the past decades, the election of an optimal personalized treatment for cancer patients remains a challenge [1]. The early prediction of tumor response would open the possibility of delivering tailored cancer treatments that could benefit the patients and optimize the organization in the clinic [2].

Many medical imaging procedures are routinely used in oncology for staging and follow-up. Thus, they represent a readily available set of data that can be used to study tumor response in a non-expensive manner using quantification algorithms. This approach, normally referred as radiomics, is currently in the spotlight of clinical research [3].

^{18}F-Fluorodeoxyglucose (FDG) PET imaging is an example of tool widely used for oncology purposes. PET analysis has mainly focused on the extraction of semi-quantitative features computed from the Standardized Uptake Value (SUV) that define tumor metabolism.

Recent research has proposed the combination of SUV parameters with volume descriptors (e.g., Metabolic Tumor Volume (MTV), Total Lesion Glycolysis (TLG)). Nevertheless, the prognosis capacity of all these parameters, even when combined, is very limited [4].

Thus, over the last years, radiomics approaches in PET have focused on the extraction of heterogeneity measurements using texture analysis. Since heterogeneity is related with angiogenesis, necrosis and even, proliferation, it can potentially provide valuable information related with tumor response to therapy [5].

In this context, the aim of the present work is to provide a complete open-source workflow to extract texture metrics that capture tumor heterogeneity and to study if they can predict response to treatment in patients with locally advanced rectal cancer.

2. Materials and Methods

2.1. Patients

All the patients in this study were diagnosed with locally advanced rectal cancer, either cT3-4 or cN+, according to the American Joint Committee on Cancer (AJCC). The inclusion criteria, staging and follow-up have been formerly described [6]. Patients underwent ^{18}F-FDG PET/CT imaging before starting their treatment. It consisted in neoadjuvant chemoradiotherapy, a surgery to remove the tumor, intraoperative radiotherapy and adjuvant chemotherapy. Detailed information on the treatment has been reported elsewhere [6].

The protocol followed the recommendations of the Helsinki declaration, the institutional Ethics committee approved it, signed informed consent from all patients was obtained and all images were correctly anonymized.

2.2. Histopathology analysis

Response was evaluated following the protocol by Quirke *et al*. [7]. Resected specimens were examined and post-neoadjuvant (NAT) changes were classified according to the sixth edition of the American Joint Committee on Cancer (AJCC) sorting (ypTNM) by an expertise pathologist. The abovementioned changes of the tumor after NAT were organized according to the tumor regression grade (TRG) [8]: TRG 0. no response; TRG 1, residual cancer outgrowing fibrosis; TRG 2, fibrosis outgrowing residual cancer cells; TRG 3, presence of residual cancer cells; and TRG 4, complete histopathological response. Tumors with TRG 0-2 were considered NAT non-responders and tumors with TRG 3-4 were considered responders.

2.3. PET imaging and processing

All images were obtained using a Philips Gemini TF PET/CT with axial field of view of 18 cm and spatial resolution of 4.7 mm full-width half maximum. All patients had a 3D PET study with 16-slices CT imaging (slice thickness of 3 mm and reconstruction slice thickness of 1 mm with interval of 1.5 mm) that comprised the pelvis (from the anal verge to the iliac crests). Axial images were reconstructed in the coronal and sagittal planes for multi-planar viewing on a review workstation. Data was normalized and corrected for attenuation, scatter radiation, random coincidences, dead time and decay. Normalization was also applied to the blood glucose level measured before FDG administration.

Studies were reconstructed by means of normalization/attenuation weighted ordered subsets expectation maximization (two iterations and 16 subsets) and a post-smoothing step using a 0.5 Hanning filter.

The PET scan was obtained 45 minutes after the intravenous injection of 5MBq/kg of FDG. Patients rested and were orally hydrated after the injection. Using a movable alignment system, the PET scan started with the patients positioned as for the radiotherapy treatment. Patients' preparation included fasting for at least six hours before the study and they were given a cleansing enema on the morning of the scan.

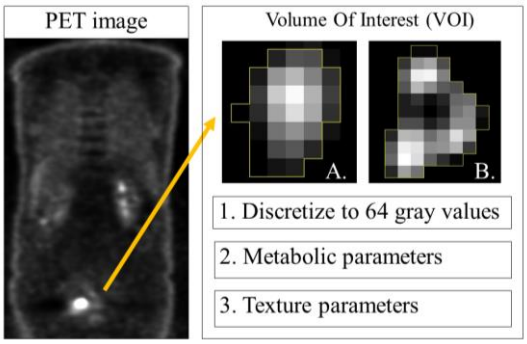

Figure 1: Summary of the workflow implemented for the analysis of ^{18}F-FDG PET-CT studies. A and B show an example of homogeneous and heterogeneous tumors respectively.

2.4. PET scan analysis

Figure 1 summarizes the whole work-flow implemented in this step. Tumors were segmented to obtain the Volume of Interest (VOI) by means of a 40% maximum activity threshold under the consideration of a nuclear clinician with expertise in the field. Following previous literature in the field [5], VOIs containing the tumor region were quantized to 64 gray values to allow comparison among patients and reduce noise in the measures extracted.

The 3-D Slicer open-source software (Version 4.0.0. Harvard University, Cambridge, MA) and the PET-IndiC module (Ethan Ulrich, University of Iowa) were used to compute the metabolic characteristics of each VOI. A complete list of all the features calculated is given in Annex A. Their definition has been reported elsewhere [9].

Several texture parameters were extracted to perform the heterogeneity analysis. The description of all the parameters used has been previously reported [9] but a enumeration of the features is presented in Annex B. Namely, a set of 16 global texture features were obtain using a set of first and higher order statistics from the gray-level histogram. Moreover, patterns in specific neighborhoods are better captured by the usage of local texture parameters. In this study, Gray Level Co-occurrence Matrix (GLCM) was used to capture the local texture changes [10]. From it, six different statistics were extracted (See Annex B) as previous literature reported their ability to capture relevant texture information while keeping the number of features used relatively low [11,12,13]. Moreover, to capture real intensity changes rather than noise artifacts, the GLCM analysis was performed at five different distances (scales in the image processing nomenclature) [13]. The values implemented consisted in the odd distances from one to ten voxels. Lastly, regional texture analysis was performed to analyze intensity changes over different regions. Indeed, Gray Level Run Length Matrix [14] was used to study texture over different run-lengths for this project. Ten different statistics capturing regional texture measures were obtained from this matrix.

Global and regional texture features were extracted with the Heterogeneity-CAD module (Narayan, V. *et al*, Harvard Medical School) in the 3D Slicer open-source package. A built-in software in Python was implemented to extract local texture metrics. The code is available upon reasonable request.

2.5. Statistical analysis

Two different approaches were used to study the clinical significance of the parameters extracted. In all the cases, the histopathology analysis aggrupation of responders and non-responders was used as a gold-standard.

First, a model was built with reduced dimensionality applying Principal Component (PC) Analysis to the different sets of data. Those PCs with eigenvalue greater than one were included in each case. A multivariate binary

logistic regression was later performed to approximate the prognosis capacity of the PCs. Secondly, a step-wise binary logistic regression was performed with a 95% statistical significance threshold in the Wald test for each of the datasets.

An evaluation of the binary logistic models was performed using a k-fold cross-validation where 80% of the dataset was used as training data and the 20% remaining was used as the validation set in 1000 folds. Mean accuracy and mean Area Under the ROC Curve (AUC) with their respective 95% confidence intervals have been provided as a result to compare the accuracy in response prediction of all of the models.

3. Results

The prediction of response from the metabolic and texture features was studied using two different approaches. In all cases, three models were built: one with the metabolic features, one with the texture features and a third one, combining texture and metabolic metrics. Figure 2 summarizes the results obtained in each of the cases.

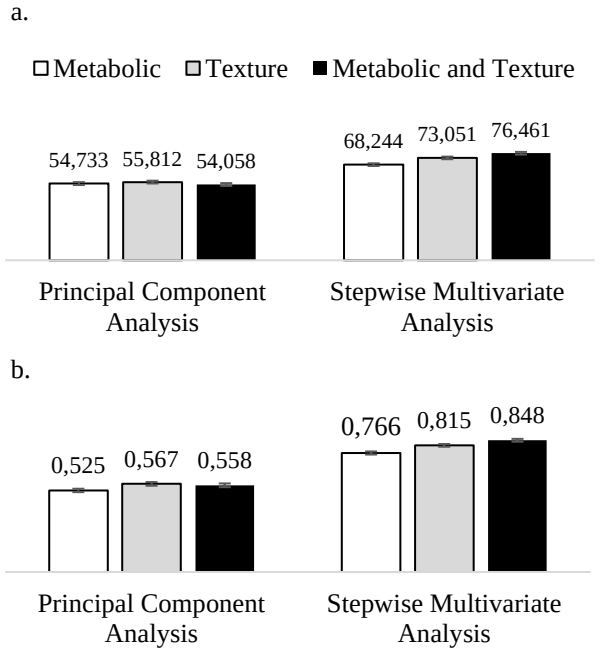

Figure 2: Results obtained from the response prediction. 2 (a) shows the mean accuracy of the models and 2 (b) shows the mean Area under the Curve (AUC). The error bars represent the 95% confidence interval in both cases.

3.1. Principal Components Analysis

The PCs extracted from the metabolic features (explained variance of 95,409%) yield an accurate result in only 54,733 ± 1,038 % of the cohort and an AUC of 0,525 ± 0,011.

The PCs extracted from the texture metrics database (explained variance of 91,161%) resulted in an accurate prediction in 55,182 ± 1,001% of the patients and an AUC of 0,567 ± 0,012.

Finally, the PCs extracted from the database that contained all the metabolic and texture features together (explained variance of 93,076%) produced an accurate prediction in 54,058 ± 1,030 of the cohort and an AUC of 0,558 ± 0,012.

3.2. Step-wise multivariate binary logistic regression

When using only the metabolic features dataset, three variables remained statistically significant. Namely, Glycolysis Q1, Glycolysis Q2 and Q1 distribution. The model built with these three variables produced an accurate prediction of response in 68,244 ± 0,965% of the cohort and an AUC of 0,766 ± 0,010.

When using the texture features dataset, four parameters related with the local texture features (GLCM) remained statistically significant (Energy at distances of three, five and nine voxels and Inverse Different Momentum Normalized (IDMN) at distance of seven voxels). This model performed an accurate prediction in 73,051 ± 0,922% of the patients and an AUC of 0,815 ± 0,009.

When metabolic and texture features were analyzed together, Glycolysis Q1, Glycolysis Q2, Energy from GLCM at distances of five and nine voxels and GLCM IDMN at distance seven voxel remained statistically significant. The results produced from this model were an accuracy of 76,461 ± 0,883% and an AUC of 0,848 ± 0,009.

4. Discussion

Heterogeneity has already been proposed as a prognosis factor in several cancer types such as breast, esophageal or lung cancers [4, 5]. In this study, we wanted to evaluate the relevance of heterogeneity in colorectal cancer.

A complete pipeline has been carefully revised to establish a feature extraction workflow. After metabolic and texture metrics were calculated, the prognosis capacity of these sets of variables had to be evaluated.

A first approximation to evaluate the clinical significance of texture analysis in PET was performed using Principal Component Analysis. Principal components extracted from metabolic and texture features showed a promising capacity to predict tumor response in our previous work [19]. Nonetheless, cross-validation evaluation showed how the principal component´s model was not robust, yielding a poor accuracy and AUC, when several validation and training cohorts were used. Therefore, after PCA, a more robust classification method was performed using a step-wise multivariate binary logistic regression at a 95% statistical significance. Surprisingly, the results from this model maintained a high accuracy and AUC even after cross-validation. This implies the robustness of the model built with these variables.

As a highlight, the better prediction was found when using texture and metabolic features together. This supports previous findings in the literature [5] where the clinical significance of texture seems to improve the routinely used metabolic features for long-term outcome prediction.

We acknowledge several limitations of this work since a relatively small number of patients has been used and there

is no validation with an independent dataset. Therefore, the clinical significance of heterogeneity in LARC presented in this study needs to be understood as a preliminary result that must lead to further investigation. Furthermore, updates in PET technologies over the past years have obtained a better resolution. Texture features extracted from images acquired using these novel technologies may be able to provide improved prognosis information over the one obtained from our dataset.

5. Conclusion

A model for prediction of response using pretreatment PET metabolic and texture features has been evaluated. After proposing a complete pipeline for this purpose, the main statistical analysis results showed how an automatic texture measurement could improve the available prognosis capacity of regular metabolic features and may become essential in routine oncology procedures.

A more advanced project with a larger patient cohort and images acquired with novel PET/CT technology is planned in order to increase the robustness of the findings from this study and extent clinical staging tools.

Acknowledgements

This work was partially supported by the Spanish Ministry of Economy and Competitiveness (TEC2013-48552-C2-1-R, TEC2015-73064-EXP, TEC2016–78052-R) (to AMB) and a Leonardo grant to Researchers and Cultural Creators 2017, BBVA Foundation (to AMB). PET dataset was financed in part by a Research Grant from Mutua Madrileña Biomedical Foundation study code CMF, FMM 06-02. PMG is supported by 'Beca de Colaboración' of the Spanish Ministry of Education, Culture and Sports. The CNIC is supported by the Ministry of Economy, Industry and Competitiveness (MEIC) and the Pro CNIC Foundation, and is a Severo Ochoa Center of Excellence (SEV-2015-0505).

References

[1] Smith N, Brown G. Preoperative staging of rectal cancer. Acta Oncol . 2008; 47(1):20-31.

[2] Blomqvist L, Glimelius B. The "good", the "bad", and the "ugly"rectal cancers. Acta Oncol. 2008; 47:5-8.

[3] Hatt M, Tixier F, Visvikis D, Cheze Le Rest C. Radiomics in PET/CT: More Than Meets the Eye? J Nucl Med. 2017;58(3):365-366.

[4] Altini C, Niccoli Asabella A, De Luca R, et al. Comparison of 18F-FDG PET/CT methods of analysis for predicting response to neoadjuvant chemoradiation therapy in patients with locally advanced low rectal cancer. Abdom Imaging. 2015;40(5):1190-1202.

[5] Hatt M, Tixier F, Pierce L, Kinahan PE, Le Rest CC, Visvikis D. Characterization of PET/CT images using texture analysis: the past, the present{…} any future? Eur J Nucl Med Mol Imaging.2017;44(1):151-165.

[6] Calvo FA, Sole CV, de la Mata D, Cabezón L, Gómez-Espí M, Alvarez E, Madariaga P, Carreras JL. ^{18}F-FDG PET/CT-based treatment response evaluation in locally advanced rectal cancer: a prospective validation of long-term outcomes. Eur J Nucl Med Mol Imaging. 2013;40(5):657-67

[7] Quirke P, Dixon MF, Durdey P, Williams NS. Local recurrence of rectal adenocarcinoma due to inadequate surgical resection: histopathological study of lateral tumor spread and surgical excision. Lancet. 1986;328(8514):996-999.

[8] Rödel C, Martus P, Papadoupolos T, et al. Prognostic Significance of Tumor Regression After Preoperative Chemoradiotherapy for Rectal Cancer. J Clin Oncol. 2005;23(34):8688-8696.

[9] P. Martin-Gonzalez. Clinical Significance of Heterogeneity and Protein Expression in Locally Advanced Rectal Cancer. (Unpublished final degree disertation). Universidad Carlos III de Madrid, 2017

[10] Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern.1973;SMC-3(6):610- 621.

[11] Conners RW, Harlow CA. A Theoretical Comparison of Texture Algorithms. IEEE Trans Pattern Anal Mach Intell.1980;PAMI-2(3):204- 222.

[12] Segmentation of a high-resolution urban scene using texture operators. Comput Vision, Graph Image Process.1984;25(3):273-310.

[13] Liu S, Cai W, Wen L, Feng DD. Multiscale and multiorientation feature extraction with degenerative patterns for 3D neuroimaging retrieval. In: 2012 19th IEEE International Conference on Image Processing. IEEE; 2012:1249-1252.

[14] Galloway MM. Texture analysis using gray level run lengths. Comput Graph Image Process. 1975;4(2):172-179.

Annex A: Metabolic parameters

The metabolic parameters used in this study were: SUVmean, SUVpeak, SUVmin, SUVmax, Metabolic Tumor Volume (MTV), Total Lesion Glycolysis (TLG), Standardized Metabolic Activity (SAM), Median, First quartile, Upper Adjacent, Glycolysis in first, second, third and forth quatiles and Distribution in first, second, third and forth quatiles.

Annex B: Texture parameters

The global texture features used were: Voxel Count, Energy, Entropy, Minimum Intensity, Maximum Intensity, Mean Intensity, Median Intensity, Range, Mean Deviation, Root Mean Square, Standard Deviation, Skewness, Kurtosis, Variance, Uniformity and Coefficient of Variation.

The local texture features extracted from the GLCM were: Energy, Contrast, Entropy, Cluster Prominence, Cluster Shade and Inverse Difference Momentum Normalized.

The regional texture features extracted from the GLRLM were: Short-Run Emphasis, High Gray-Level Run Emphasis, Long-Run Emphasis, Short-Run Low Gray Level Emphasis, Gray-Level Non-Uniformity, Short-Run High Gray Level Emphasis, Run-Length Non-Uniformity, Long-Run Low Gray Level Emphasis, Run Percentage, Low Gray-Level Run Emphasis and Long-Run High Gray Level Emphasis.