Parietal team, Inria
CEA, Université Paris-Saclay
1 Rue Honoré d'Estienne d'Orves
91120 Palaiseau
France

February 1, 2017

Nicholas Ayache
Co-editor in Chief
Medical Image Analysis

Subject: MEDIA-D-16-00507 - Response to reviewers

Dear Sir or Madam,

We would like to thank the reviewers for their pertinent remarks that have increased the quality of the manuscript. Whenever possible, the comments of the reviewers have been addressed and highlighted in the article.

Please, find below a point-by-point response to each issues raised by the reviewers.

Reviewer #1:

1. **In a typical DCE based analysis, the contrast enhanced images are rigidly registered to T1w images for intensity normalization. Typically for multi-parametric MRI based CAD system for prostate cancer detection T1w images are present. This is useful in differentiating cancers from blood hemorrhage following prostate biopsy. Please discuss why a pre or a post contrast T1w image is missing from the clinical trial that will be using DCE images.**
   The MRI acquisition was performed prior to carry out the biopsy. Therefore, the risk of blood hemorrhage is discarded. Additionally, pre-contrast T1w images are available since 4 to 6 volumes are acquired previous to the bolus injection of the contrast media. Indeed, this pre-contrast data are used in some models (i.e., S_0).
2. **The other advantage of normalizing to T1w images are patient specific information will be preserved. Different patients may have different cancer lesions of different grades (having different kinetic expression in DCE) and normalizing all patients to the same space may result in loss of information. Please comment.**
   Our normalization is designed to align the *mean kinetic expression*. Therefore, it is true that patient-related information can be lost. However, lesion with different grades should more or less dispersed to the mean kinetic expression. This information remains available after applying our normalization.
3. **T1w normalization will be probably be more robust to change in magnet, machine, reconstruction software and acquisition parameters across different institutions as the objective is normalization to be patient and**

institution specific. Please comment how the presented model may be affected in a multi-institutional study.
Our method does not rely on any manufacturer/machine-specific settings. Subsequently, our approach should work on any type of devices, enabling multi-institutional studies. We included these remarks in our future work section in p. 23/l. 396.

4. **Please discuss how the improvement in prostate cancer detection in DCE-MRI impacts the current clinical CAD pipeline. According to PIRADS v2 guidelines more emphasis is laid upon T1w, T2w and DWI compared to DCE images. Experienced radiologists can detect prostate cancers in peripheral zone with a high degree of accuracy (higher than 0.6 AUC). The transition zone lesions are more difficult to distinguish. Please comment how the model performs in detection of transition zone lesions. Further please comment how this model adds value to already established clinical pipeline of detecting prostate cancer (PIRADS).**
A stratified analysis separating PZ and CG showed an increase of the classification performance in CG after applying our normalization method as depicted in Fig. 8 p. 22. The PIRADS v2 emphasized that DCE-MRI have to be included in mp-MRI acquisition to detect small lesion. However, it also warn that the benefit is modest. Therefore, the impact of an improvement of detection in DCE-MRI can only be evaluated in a fully mp-MRI CAD system. Indeed, we recently showed, — Lemaitre, Guillaume. Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging. Diss. Université de Bourgogne; Universitat de Girona, 2016; Chapt. 6/p. 150/Table 6.10 —, that DCE-MRI features are considered important and the combination with other modalities will lead to better classification results. However, in our humble opinion, including those additional results will clutter the paper.

5. **Comparing the results of semi-automatic approach and the presented approach of curve fitting using the entire curve the results are not significantly different (0.65 AUC vs 0.66 AUC). The presented model is however automatic which would reduce inter-rater variance in a semi-automatic method. Please present a discussion comparing these two results.**
It is important to notice that for both approaches, — semi-quantitative and entire enhanced signal —, the results are obtained using our normalization approach, showing the importance of this step. We added a discussion in p. 23/l. 388 regarding the semi-quantitative approach and entire enhanced signal classification. The semi-quantitative approach reduce the number of features to analyze, speeding up the classification. However, it relies on the curve fitting which is also time consuming and might fail.

6. **Use of single feature for a random forest classifier (Table 2) may be detrimental for the classifier as random sub-sampling of the features are not possible. Please present the comparison with some linear classifiers like linear SVM.**
A classification using a linear SVM has been added in additional material. Unlike a linear SVM, RF allows to find multiple thresholds, leading to a much complex decision boundary for this feature. Bootstrapping the samples for each tree allows to cope with the over-fitting of each decision trees. Therefore, we do not think that a linear SVM will substantially better insights as are showing the results provided in the additional material.

Reviewer #2:

1. **The title is misleading and not very scientific and I would recommend to remove the second part and reword the first part to cover the contents of the paper.**
   We are not against changing the title but we would like to take advantage of this revision to get the opinion of all reviewers regarding this matter.
2. **The use of the English language needs some attention: get it all proof read.**
   As suggested, the manuscript has been corrected by an English editing service to ensure a good quality writing.
3. **What guided biopsy was used: US?**
   The guided biopsy was performed with a TRUS. We added this statement in p. 15/l. 275.
4. **The number of cases seems low. This is mentioned as possibly affecting the results and that the limitations might be based on the sample size. At the end of section 5 one could draw the conclusion that they should do a more extensive study.**
   The statement has been added in the discussions p. 23/l. 396.
5. **How would this translate to other types of MRI systems: it would be good to see experimental results on this or at least some discussion.**
   Refer to reviewer #1 / comment 3 since this is a shared remarks.
6. **The disappointing performance and large standard deviation of the PUN model should be investigated and explained.**
   A discussion is added in p. 17/l. 320. The original model lacks of degrees of freedom. Two additional parameters will allow for a better fit. However, we used the original formulation in our experiments.
7. **There should be a detailed discussion on the results to indicate which differences are statistically significant. "Significant" is mentioned in the conclusions, but there is no evidence for this.**
   We added a Wilcoxon significance test - see Fig. 7. Now, the term significant is only used for $p < 0.05$.
8. **The overall results seem poor and this could be further discussed.**
   It is difficult to compare with the state-of-the-art methods since the dataset used are different and the reported results in those studies are usually linked to a mp-MRI CAD instead of only DCE. Nevertheless, our results are in the same range than Litjens, Geert, et al. "Computer-aided detection of prostate cancer in MRI." IEEE transactions on medical imaging 33.5 (2014): 1083-1092 regarding the classification performance of the Tofts parameters.
9. **Sections 5 and 6 are both disappointingly short.** Due to additional analyses, the discussion became longer.
10. **Some journals references incomplete. Gliozzi et al. is not a single page. Check: Schabel, Shanbhag, Zhu.**
    The references have been corrected.

Reviewer #5:

1. **There are however several issues with the paper In the current state. Using quantitative parameters obtained from pharmacokinetic models as an endpoint comparison indeed may be important as the models provide physiological information related to capillary permeability, blood fraction and interstitial space volume, etc.. However I am not really convinced by the detailed results presented in Table 2, as no statistical test was performed when the classification performances are higher with normalized data, and for some of the models the results are worst. For me, the benefits of the method are not fully demonstrated in terms of classification performance using the models. How these parameters are modified with the normalization?**
   This is important to emphasize that our goal was to propose a normalization method to allow classification of the entire enhanced signal which should perform as good as the current models. We apply the same normalization on the different pharmacokinetic models and, as mentioned by the reviewer — the results are not significantly better. Indeed, these experiments were carried out to investigate the benefit or not of the normalization with the current models, without expecting the results to improve.

2. **Some of the benefits or improvements may be demonstrated before any further classification step. The method could be useful when performing population analysis with the raw data, previous to defining a model.**
   We do not really see which analysis can be performed to evaluate the benefits of our methods before the classification step. A study of the variances with and without normalization could be carried out but an evaluation after classification seems equivalent and a good manner to evaluate the benefit of the normalization. In fact, this is the results reported in Fig. 6(c).

3. **Figure 1 is not pertinent for the goal of this paper. It is not clear how is represented the interindividual variation in figure 2? In Fig 2a the PDFs are concatenated in time to obtain a heatmap representation, but in 2 b-c the differences across individuals does not appear.**
   Figure 1 is used to illustrate inter-patients variation in T2w-MRI and to highlight the need of a different normalization approach than the one needed in T2w. We improve the Fig. 2. Figure 2(a) presents only the process to build the heatmap. In Fig. 2(b)-2(c), we added parameters which highlight more clearly the variations and added their corresponding values in caption.

4. **It is not clear the rational of equation 1, as the parameters are rather empirical. There are no other ways to smooth the heatmap, for example through a diffusion process, solving a PDE equation? Or by registering through the series?**
   There is probably a confusion. We do not try to smooth the heatmap but we search a smooth estimate. Equation 1 represents the weight of the graph and is a standard way to formalize graph-based problem in image processing.

5. **The theoretical foundations of equation 3 are not clear. Other registration methods for time signas may have been included.**
   While other registrations could have been used, using a rigid registration directly affect the variations α_i and Δ_t. Applying more advanced registration such as elastic registration as in Lemaitre, Guillaume, et al. "Normalization of T2W-MRI prostate images using Rician a priori." SPIE Medical Imaging. International Society for Optics

and Photonics, 2016. would make it complicated to shift the data with only these two parameters. Furthermore, rigid registration will not change the variations of the contrast agent over time.

6. **Other question that may be raised is the influence of the equation parameters (tau, alfa..) in the pharmacokinetic models. Do they change with f(t)?**
   We think that this question is related to the results presented in Sect. 4.1

7. **A thorough review of the structure and in general an improvement of the paper in terms of writing may be helpful to fully understand the message, the rational and the proposed methodology.**
   As suggested, the manuscript has been corrected by an English editing service to ensure a good quality writing.

Thank you for your time and consideration.

I look forward to your reply.

Sincerely,

Guillaume Lemaître