Author’s response to reviews

Title: Blood Oxygen Level Dependent Magnetic Resonance Imaging for Detecting Pathological Patterns in Lupus Nephritis Patients: A Preliminary Study Using a Decision Tree Model

Authors:

Huilan Shi (shihuilan@vip.126.com)

Junya Jia (youshalei@126.com)

Dong Li (ldno3@163.com)

Li Wei (weilideemail@126.com)

Wenya Shang (wen_xiaoya.5151@163.com)

Zhenfeng Zheng (zhengzhenfeng@vip.126.com)

Version: 3 Date: 11 Jun 2017

Author’s response to reviews:

Reply to Reviewer 1:

1. It is not clarified in the paper which data were used as a basis to develop and train the MR image processing methods and statistical methods (training data), and which data are used to validate the methods (test data). There is some mention of cross validation (p7,63), but as far as I understand from the explanation that is only applied after fixing the methods, so the role of it is not clear to me. If the training data and the test data are the same, this is a serious limitation that must be acknowledged and properly taken into account in the interpretation. MRI images contain vast quantities of information and in principle almost unlimited numbers can be extracted from the data - increasing the probability that a good classification can be achieved through pure chance alone. This is all the more true because the number of cases is small and R2* is not very precise. Extrapolating from previous studies a 10-20% coefficient of variation for median R2* is not unreasonable, presumably significantly more for less precise parameters (skewness etc) and when physiological fluctuations are taken into account. This means that the probability of finding a good classification in N=12 training data due to chance alone may actually be quite high, but as far as I can see these effects are not accounted for in the analysis.

Response to comment 1:

We rechecked methods section of our manuscript, especially the Image Analysis sub-section. It was found that we did not explain clearly how to develop three predictive algorithm
classification models. After the total of 12 lupus nephritis patients were scanned by BOLD-MRI, we acquired bilateral renal BOLD-MRI information (left kidney and right kidney BOLD-MRI information) in each lupus nephritis patient. Subsequently, by FUNCTOOL program analysis, we gained three consecutive renal coronal anatomical planes of R2* maps in each side of kidney. In other words, we gained 6 R2* maps (3 R2* maps of left kidney and 3 R2* maps of right kidney) in each lupus nephritis patients. In each R2* map, the entire renal parenchyma, including both cortex and medulla, was selected as ROI. R2* values of each voxel of selected ROI were obtained by MATLAB software. Subsequently, a large R2* values data of each voxel from selected ROI were obtained. We randomly selected 100 groups of R2* values which included 100 consecutive voxels. Since each LN patient provided 3 R2* maps of left kidney and 3 R2* maps of right kidney, respectively. All of 12 LN patients provided 7200 groups of R2* data (including 3600 groups of R2* data from left kidneys and 3600 groups of R2* data from right kidneys). We used the 3600 groups of R2* data from left kidneys as training data to develop algorithm models. We used another 3600 groups of R2* data from right kidneys as testing data. See figure response 1.

Previous relevant renal BOLD-MRI studies showed definitely renal R2* values heterogeneity. In my opinion, renal R2* values variation was composed of spatial and horary heterogeneity. Firstly, renal tissue oxygenation distribution was not uniformity according to our BOLD-MRI investigation. For example, R2* values in superficial layer of renal cortex are lower than those in deep-seated layer of renal medulla. Moreover, R2* value changes from renal cortex to medulla is continuous and no obvious R2* values border is found between cortex and medulla. To validate renal R2* distribution mode, we selected one healthy volunteer as a testing sampling patient. Bilateral renal largest coronal anatomical planes were selected as testing planes. See Figure Suppl 1

Subsequently, we created 6 rectangle ROIs (wide × high= 6×30 pixels) in each renal R2* plane. Every rectangle ROI was placed along the direction from cortex to medulla. According to the rectangle ROIs high scale, we can measure the R2* value in different depth of renal tissue. See Figure Suppl 2.

We now have acquired 12 ROIs pictures each with 6 pixels wide and 30 pixels high. By using MATLAB software, every colorful pixel in ROI pictures is transformed into R2* values. The following picture showed the relationship of R2* values and the depth of renal tissue. Data is from two of ROI pictures. Along the direction from cortex to medulla, renal R2* values increase gradually. However, no obvious R2* values border between cortex and medulla is found. See Figure Suppl 3.
In developing decision tree model period, the training sample R2* data came from entire renal parenchyma, including renal cortex and medulla area. Similarly, the testing sample R2* date were extracted from renal cortex and medulla. Under this circumstance, our decision tree model could overcome R2* values spatial heterogeneity and avoid influence of discrepancy sample voxel location because of chance. On the other hand, when we predicted renal pathological patterns by using decision tree models, we speculated final results on the basis of multiple R2* sampling data, rather than one or two R2* samples. For predicting renal pathological classification of each lupus nephritis, 300 groups R2* values were used by decision tree model. For example, in the Table 4, the renal biopsy result of lupus nephritis patient Case 1 was IV type, predictive results of decision tree model show that proportion of III type and IV type were 25% and 75%, respectively. The final predictive result was IV type. In this way, we also eliminated the adverse impact because of chance. Consequently, on the basis of our study, it was no evidence that the good classification capability of decision tree models due to sample chance because of renal R2* spatial heterogeneity.

Secondly, another reason of R2* values variation was horary heterogeneity. It is believed that there are no two duplicate renal R2* maps which were scanned by BOLD-MRI in two different time points. To our knowledge, numerous factors can affect renal R2* values such as, hydration status, sodium balance, ARB, ACEI, diuretic and anti-oxidant therapy etc. Unfortunately, our study did not provide multiple time points of renal BOLD-MRI scanning for each LN patient. Because of this study limitation, we could not verify whether horary heterogeneity could impact on classification capability of decision tree.

2. There are key elements of the MR image processing aspects that are not sufficiently clearly explained, even though they are quite crucial to the conclusions. I am referring specifically to the aspects mentioned in p6, line 23 onwards referring to the subdivision of each ROI in 100 randomly selected groups of 100 consecutive voxels. It is not clear at all to me how this is done exactly, what data are extracted and how they are used to generate probabilities and classifications. The authors return to this part of the methods in the discussion section (p10,51 and onwards) explaining that this is a new methodology developed for this project with some more details on the approach - though it remains unclear to me what they have done exactly. They also attempt a justification in terms of heterogeneity but the explanation is confused and unfortunately doesn't help me to gain additional insight in what this does or why.

Response to comment 2:
In the methods section, we introduced a new renal R2* data extraction method from a R2* map. In order to explain this method more clearly, we represented this new methodology step by step.

Step one:

One renal coronal anatomical plane section was acquired from ADW 4.5 Workstation using FUNCTOOL program.

Step two:

To draw a large ROI including renal cortex and medulla manually, the renal collecting system was excluded.

Step three:

All of voxels of the ROI picture was transformed into digital according to RGB (Red-Green-Blue) encoding system. In the RGB type picture, RGB color of each voxel is encoded as a positive integral number from 0 to 255. By using MATLAB software, the ROI picture was transformed into a group of numerous numbers. The following picture show a small part of the entire transformed ROI picture.

Step four:

We selected 100 groups of consecutive numbers in the entire transformed numbers randomly. Each group composed 100 consecutive numbers which represented 100 consecutive voxel in ROI picture. The following figure showed one group of selected 100 consecutive numbers.

Step five:

According to the R2* value measuring scale, we can calculate the R2* value of each voxel. The following figure showed one group of R2* value.

Step 1 to 5 showed the whole flow path of the MR image processing.

The principle idea of our new methodology was to classify LN pathological type by using probability of algorithm model predictive classification. In other words, each sample of renal
R2* data will generate one predictive classification independently and “votes” for the corresponding class. The majority of the votes decided the overall prediction. This aggregate vote of multiple samples was inherently less noisy and less susceptible to chance than a single sample output. This methodological idea was also similar to the random forest (RF), which mitigate the volatility due to small data and improves the robustness of predictions. [L. Breiman, Bagging predictors, Mach. Learn. 24 (1996) 123–140] [L. Breiman, Random forest, Mach. Learn. 45 (1999) 1–35] [R.J. Marshall, The use of classification and regression trees in clinical epidemiology, J. Clin. Epidemiol. 54 (2001) 603–609]

3. What this study is also lacking is a clear hypotheses to justify why there would be a role for BOLD in LN. R2* is sensitive to blood oxygenation, but is there a reason why one would expect this to help in the subclassification of LN? What exactly is the role of blood oxygenation in LN pathogenesis? What changes would one expect and can these be related to differences seen on the histology? The only motivation the authors are giving for this study is that it hasn't been done (p4, 48). Another issue is that the authors discuss BOLD only in terms of oxygenation but the effect of microstructural changes in R2* are equally important and must be taken into account in the interpretation.

Response to comment 3:

Our study group also investigated renal R2* map texture characteristics of lupus nephritis patients. We found that there were difference of texture indexes among different renal pathological types. For example, we studied the GLCM indexes of LN renal R2* map and found the energy values differences, one of the GLCM index, among three renal pathological classification including III type, IV type and V type. Based on these results, we hypotheses that renal BOLD-MRI may show some kind of specific image patterns, which may corresponding to pathological types. Therefore, we not only measured the R2* value of each voxels from selected ROI, but also concerned the R2* values relationship between each voxels. The aim and motivation of our study were to explore relationship between renal R2* map characteristics and histological pattern.

Renal microstructural changes may impact on renal oxygenation. For example, our research group investigated the relationship of renal R2* values and renal pathological injury pattern by multiple corresponding analysis. We found that renal tubular atrophy correlated with lower R2* value (higher oxygenation) in proliferation LN patients. However, in non-proliferation LN patients, such as V class LN, well preserved renal tubular microstructure correlated with higher R2* value (lower oxygenation). We speculated the possible mechanism in connection with ionic transport pump distributed on renal tubular. Under normal physiological conditions, these ionic pumps consume a massive amount of oxygen. Tubule injury could damage these pumps resulting in a reduction in oxygen consumption. [Shi H, Yan T, Li D, Jia J, Shang W, Wei L, Zheng Z.

4. P5, MRI techniques: there are important details missing in the technical specification of the imaging methods. For instance is this done in free-breathing? Breath hold? Triggered? 2D or 3D sequence? What slice orientation? Is parallel imaging used etc.. Full details of the key sequences need to be provided.

Response to comment 4:

We added more BOLD MRI information into our paper:

BOLD-MRI were acquired using three consecutive parallel coronal slices for each side of kidney. The patients should be breath-hold of 20 seconds (in expiration) during MRI scanning.

Other information:

Due to breath-hold during patients’ BOLD-MRI scanning, we do not need triggered parameters. To our best knowledge, almost all of published renal BOLD-MRI studied focused on 2D R2* map of BOLD-MRI. We have not found any BOLD-MRI study related to 3D R2* map. Therefore, we think that we do not need to state 2D or 3D sequence in BOLD-MRI.

5. P6, Algorithm models: more details about the models need to be provided and references need to be given.

Response to comment 5:

Linear discriminant analysis (LDA) is a supervised categorical technique that maximizes group differences by creating a weighted linear combination of the discriminating variables. [Ref]. The original LDA has two derivation including fisher LDA (FLDA) and least square LDA (LSLDA). FLDA is based on Fisher-Rao’s criterion [Fisher RA. The use of multiple measurements in taxonomic problems. Annals of Eugenics 1936;7:179-188.] [Rao CR. The utilization of multiple measurements in problems of biological classification. Journal of the Royal Statistical Society Series B (Methodological) 1948;10:159-203], which is to find the projection w to maximize the objective function denoting the ratio of between-class to within-class variances.

The formula listed below indicates the discriminant functions.
Where $D$ discriminant function, $b$ the discriminant coefficient or weight for that variable, $X$ discriminating variables, $c$ a constant and $n$ the number of predictor variables.

The binary logistic regression (LR) model is used because of the response variable takes just two values. This model is primarily used to identify the relationship between more independent variables ($X_i$) and the dependent variable ($Y$) to predict the independent variables that are most influential on the dependent variable. The formulae listed below shows the relationship between response probability and the predictor variables.

Where, $\pi_i$ is the ratio of the probability of one of the classification, $\beta_0, \beta_i$ are parameters to be estimated, and $p_i$ is the response probability for $i$th group, $k$ is number of variables.[Alkarkhi AF, Easa AM. Comparing discriminant analysis and logistic regression model as a statistical assessment tools of arsenic and heavy metal contents in cockles. Journal of sustainable development 2008;1:102-106]

Decision tree is a simple algorithm technique to classify patterns in numerous categories. In this model, the relationship between data is represented in a tree structure, starting from a root node to different nodes via multiple branches and finally ending in some terminal nodes. In our current study, Chi-squared automatic interaction detection (CHAID) algorithm was used. By CHAID algorithm, the generated decision tree plots demonstrates relationship between split variables and associated related factors, which enables population subgroups with homogeneous to be revealed. Decision tree contains a group of multiple mutually exclusive pathway from root node to terminal nodes, which represents classification rules.[Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision tree: an overview and their use in medicine. J Med Syst. 2002;26:445-463.]

6. P7, 25. Why only 4 patients?

Response to comment 6:

In page 7, line 25, we did not find “4 patients”. Would you please give us more detailed information?
7. P7, 51. How were the user-specified levels chosen? This may relate to point 1)

Response to comment 7:

In page 7, line 51, we did not find “the user-specified levels chosen”. Would you please give us more detailed information?


Response to comment 8:

In Table 2, we added Definition of the activity and chronicity indexes:

Leucocyte exudation: exudation of more than two polymorphonuclear leucocytes per glomerulus

Glomerular cell proliferation: glomerular endocapillary hypercellularity leading to reduction of circulatory volume of glomerular capillary loops.

Karyorrhexis and fibrinoid necrosis: the presence of pyknotic and fragmented nuclei; the occurrence of intensely eosinophilic material within solidified segments of glomeruli.

Cellular crescents: proliferating extracapillary cells occupying one-fourth or more of the glomerular capsular circumference.

Hyaline deposits: eosinophilic material of a homogenous consistency along the circumference of the luminal surface of glomerular capillaries constituted the classical wire loop lesion.

Interstitial inflammation: infiltration of mononuclear cells into interstitial spaces.

Glomerular sclerosis: glomerular capillary collapse with attendant expansion of mesangial matrix material and subsequent solidification.

Fibrous crescents: structures composed predominantly or exclusively of fibrous tissue lining Bowman’s capsule in a circumferential pattern.

Tubular atrophy: thickening of tubular basement membranes, with or without tubular epithelial cell degeneration.

Interstitial fibrosis: the deposition of periglomerular and peritubular fibrous tissue.
In Table 3, we added definition of sensitivity, specificity and accuracy

Sensitivity: true positive rate = number of true positive / (number of true positive + number of false negative)

Specificity: true negative rate = number of true negative / (number of true negative + number of false positive)

Accuracy: (number of true positive + number of true negative) / (number of true positive + number of false positive + number of false negative + number of true negative)

In Table 5, we added definition of primary class, secondary class, homogeneity and heterogeneity

Primary class: principle histopathological manifestation

Secondary class: detailed histopathological manifestation

Homogeneity: without concomitant V type glomerular histopathological manifestation

Heterogeneity: with concomitant V type glomerular histopathological manifestation

Reply to Reviewer 2:

1. The major limitations of the study have been clearly delineated by the authors themselves. The number of subjects studied is small and the spectrum of histologic abnormalities studied is narrow. The study lacks class II, V or VI patients, most of their patients had a mixed membranous component to their underlying proliferative pathology, chronicity was limited and GFR was well preserved. Applicability to what clinicians encounter in general practice is thus limited. Any clinically-relevant predictive model would need to deal with a broader range of pathology comparable to what is seen in practice. The real challenge is greater than merely differentiating class III from class IV LN but also assessing the extent of disease activity versus chronicity. The lack of ADC in DWI MRI also limits interpretation of their results since these data would be expected to enhance the predictive power of noninvasive techniques.

Response to Comment 1:
Our primary exploration study did have obvious limitations such as small sample size and narrow spectrum of clinical and histologic characteristics. Firstly, the study lacked the class II and VI LN patients because pathological injuries were too mild or extensive to undergo renal biopsy. The lack of class V LN patients data lead to our decision tree model without predictive capability in non-proliferative LN patients. The including patients were not representative of total population because of well-preserved GFR. It is well known that the eGFR will be lower when the renal pathological injury was extensive and active. Sometimes, some LN patients needed to undergo hemopurification therapy. Secondly, the R2* map may not evaluate the extent of histopathological injury. We found that lower R2* values were detected in extensive proliferative LN patients’ renal tissues. However, lower R2* values were also detected in control group (healthy volunteers group). [Huilan Shi, Tiekun Yan, Dong Li, Junya Jia, Wenya Shang, Li Wei, Zhenfeng Zheng. Detection of renal hypoxia in lupus nepritis using blood oxygen level-dependent MR imaging: a multiple correspondence analysis. Kidney Blood Press Res 2017;42:123-135.]. Thirdly, we did not investigate other MRI index such as ADC, which was deemed to useful factor for prediction renal tissue fibrosis.[Inoue T, Kozawa E, Okada H, Inukai K, Watanabe S, Kikuta T, Watanabe Y, Takenaka T, Katayama S, Tanaka J. Noninvasive evaluation of kidney hypoxia and fibrosis using magnetic resonance imaging. JASN 2011;22(8):1429-1434]  

2. I am skeptical that the specificity/sensitivity of the best model is sufficient to support clinical decision-making and thus obviate the need for kidney biopsy. 

Response to Comment 2:  

Our decision tree model was the best specificity/sensitivity model in three predictive algorithm models. The specificity/sensitivity indexes were used to evaluate predictive capability in one single testing sample. However, we predicted the patient’s renal pathological diagnosis dependent on multiple samples detection results. In other words, the final diagnosis came from the predictive result of majority proportion samples. This principle idea of prediction was similar to decision tree & random forest. By using multiple samples detection idea, we may increase the accuracy of prediction. Therefore, our decision tree model plus multiple samples detection may promise to support clinical decision-making. Due to our study limitation, our present decision tree model still did not substitute the kidney biopsy. For example, if we include more indexes from multiple functional MRI technique such as DTI-MRI, ASL-MRI, DWI-MRI and DKI-MRI etc into our decision tree model, it is hopeful that our decision tree model may have comparable accuracy of diagnosis with renal biopsy inspection.
3. A major additional limitation is the lack of a standardized pre-study protocol. Hydration status, sodium balance, and ARB, ACEI, diuretic and anti-oxidant therapy were not standardized. All these factors can influence R2*.

Response to Comment 3:

The lack of a standardized pre-study protocol was definitely our major limitation. We have already considered the multiple factors which can impact on renal R2* values. On one hand, not only we should consider many influenced factors as possible as we can, but also we should clarify the importance degree of each factor affecting on renal R2* values in order to map out reasonable standardized protocol. To our best knowledge, these possible influenced factors have not studied clearly in the published literature. On the other hand, lupus nephritis is not a common glomerulonephritis like IgA nephropathy. Standardizing lupus nephritis pathophysiological condition may decrease the qualified study samples. One of the feasible way was to analyze these factors by using well-designed research approach and data mining technique such as canonical correlation analysis and structural equation model (SEM).

Minor

1. The discussion section should be shortened

Response to Comment 1:

We eliminated some parts of text in the discussion section.

2. The first line of the Background section of the Abstract is difficult to understand

Response to Comment 2:

Precise renal histopathological diagnosis will guide therapy strategy in patients with lupus nephritis.