Pathological grading of Hepatocellular Carcinomas in MRI using a LASSO algorithm

Fei Gao¹,³, Bin Yan¹, Jian Chen¹, Minghui Wu² and Dapeng Shi²

¹National Digital Switching System Engineering & Technological R&D Center, ZhengZhou, Henan Province, China
²Department of Radiology, Henan General Hospital, ZhengZhou, Henan Province, China

³Email: gfflyfly@163.com

Abstract. To investigate the predictive ability of Radiomics signature for preoperative pathological grading of Hepatocellular Carcinomas (HCC), the no contrast MRI images were integrated and a comprehensive analysis was conducted to predict clinical outcomes using the radiomics features. Variable selection via LASSO and logistic regression were used to select the most-predictive Radiomics features for the pathological grading. Cross-Validation with receiver operating characteristic (ROC) analysis was performed and the area under the ROC curve (AUC) was employed as the prediction metric. Overall, the prediction performances by Radiomics features showed statistically significant correlations with pathological grading, however, improvement on the prediction performance by combining T1WI and T2WI data, classification performance obtained the AUC 0.829 in training dataset and the AUC 0.742 in validation dataset. This study consisted of 170 consecutive patients (training dataset: n=125; validation dataset, n=45). The results showed Radiomics signature was developed and validated to be a significant predictor for discrimination HCC pathological grading, which may serve as a complementary tool for the preoperative tumour grading in HCC.

1. Background
Hepatocellular Carcinomas (HCC) is the most frequently diagnosed cancer in the world, its incidence and mortality in China ranked the fourth and second respectively [1], a cancer usually treatable if an early diagnosis is possible. By assessing the characteristics of liver tumor noninvasively, MRI is often used in clinical practice for disease diagnosis and treatment guidance [2], MRI has great potential to guide therapy because it can provide a more comprehensive view of the entire tumor and it can be used on an ongoing basis to monitor the development and progression of the disease or its response to therapy. Further, MRI is noninvasive and is already often repeated during treatment in routine practice. But radiology doctors can't get more valuable information from MRI images directly, they must analysis images by the help of computer to find a further information for clinical outcomes.
Pathological grading in oncology is closely related to cancer diagnosis, prognosis, and treatment planning. For HCC, traditional pathological grading depends on aspiration biopsy and surgical pathology, this invasive method of understanding tumor heterogeneity is not suitable for all cancer patients before treatment. Lambin [3] proposed the concept of radiomics in 2012, it is an emerging field that converts image data into a high dimensional mineable feature space using a large number of automatically extracted data-characterization algorithms, these image features can capture the subtle differences in the image across different tumors, it has shown great advantages in phenotypic typing, treatment options and prognosis analysis, which is a research hotspot in clinical medicine and biomedical engineering [4-5]. Figure1 depicts the block diagram of the entire process.

More and more radiomics researches are used to study the phenotypic and classification of tumors. Kumar [6] expanded further and defined radiomics as “high throughput from CT, MRI, and PET (positron emission computed tomography), a large number of advanced quantitative image features are extracted and analyzed”, Doroshow [7] published articles in “Nature Reviews Clinical Oncology”,
pointing out that radiomics is one of the future directions of transformation medicine. In 2014, the articles of Aerts [2] published in “Nature Communication”, they extracted 440 quantitative image features from 1019 cases of lung cancer or head and neck cancer patients with CT data, including the gray distribution, shape and texture, these image features reflect the heterogeneity of tumor, and tumor pathological type, T stage, gene pattern of expression is related.

Image segmentation is a key step in radiomics. Precise segmentation of lesion area is very important for subsequent feature extraction and model construction. Many image segmentation algorithms are limited in practice because of the diversity and complexity of medical images in different tissues. There is no uniform and perfect segmentation algorithm in the field of medical image segmentation today. The main segmentation algorithms are manual segmentation, human-computer interactive segmentation and automatic segmentation. Although automatic segmentation can improve efficiency, but the effect of algorithm is not good at present. Manual segmentation is often used as the golden standard of image segmentation. However, manual segmentation has high inter observer bias, and manual segmentation is inefficient and reproducible.

Radiomics analysis requires accurate feature extraction technology to extract all kinds of features, and then bioinformatics method is associated with various clinical phenotypes of tumor. We can obtain massive and high dimensional features by automatic image feature extraction algorithm, but in data fitting, when sample size is small, it will cause data dimension disaster when fitting. When building model, it will lead to over-fitting, the results can’t be applied to other data samples, the generalization ability is very poor. Therefore, it is necessary to discuss the appropriate dimensionality reduction methods to retain the characteristics of maximum correlation and minimum redundancy for subsequent feature selection. Fan [8] proposed variable selection should meet the following requirements: (1) the accuracy of prediction model; (2) interpretability of the model; (3) the stability of the model, that is small changes in data set do not lead to larger changes in the model; (4) to avoid bias in hypothesis testing; (5) we should try to control the computational complexity. At present, there are many ways to reduce dimension, such as principal component analysis (PCA), random forest (RF), support vector machine (SVM), clustering, etc. But these methods can only reach some of targets, although ridge regression can better deal with multiple collinearity among variables, it can’t provide a sparse model because it can’t reduce the dimension. LASSO is also the initials for Least Absolute Shrinkage and Selection Operator, this method is proposed by Tibshirani [9], it uses the absolute value function of the model coefficients as a penalty strategy to compress the model coefficients, therefore, LASSO can provide a sparse solution. Compared with the traditional method of variable selection, the LASSO method overcame the shortcomings of the traditional methods. With the increase of computing power, LASSO has been used more and more in high-throughput data field, such as gene, protein and image.

LASSO is an innovative variable selection method for regression. Variable selection in regression is extremely important when we have available a large collection of possible covariates from which
we hope to select a parsimonious set for the efficient prediction of a response variable. LASSO minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. LASSO not only helps to improve the prediction accuracy when dealing with multi-collinearity data, but also carries several nice properties such as interpretability and numerical stability.

2. A novel method for liver MR image pathological grading based on LASSO

The proposed pathological grading method consists of five steps: image acquisition and preprocessing, feature extraction, feature selection using LASSO, classification based Rad-score, and classification based clinical model combined the radiomics signature. This method is described in the following sections.

Analysis Workflow 1. Pathological grading based on LASSO.

Step1: Image acquisition;
Step2: Image preprocessing by Laplacian of Gaussian spatial band-pass filter;
Step3: Extracting the lesion area(ROI) from the MR images by manual segmentation;
Step4: Extracting the high dimension features of ROI;
Step5: All the samples will be divided into two parts: training dataset and test dataset;
Step6: Feature selection based on LASSO regression algorithm on training dataset;
Step7: Constructing a classification model by using selected features in step6 to make up signature;
Step8: Verifying the performance of classification mode in step7 on test dataset.

2.1. Image preprocessing

A process was applied to selectively extract features of diverse sizes and intensity variations, a series of images ranging from fine to coarse texture that highlighted and enhanced tumor features at different anatomic spatial scales, was derived from a MRI slice by a Laplacian of Gaussian spatial band-pass filter ($\nabla^2G$) with five filter values (0, no filtration; 1.0, fine textures; 1.5and 2.0, medium textures; 2.5, coarse textures) [10]. The distribution of the $\nabla^2G$ is given by the following the mathematical expression:

$$\nabla^2G(x,y) = \frac{-1}{\pi\sigma^4}[1 - (x^2 + y^2)/2\sigma^2]e^{-\frac{x^2+y^2}{2\sigma^2}}$$

(1)

Where x, y are the spatial coordinates of a pixel and $\sigma$ is the value of the filter parameter.

2.2. Feature extraction

In this stage, many significant features from the MR images were extracted to be subjected to classification. As previously discussed, the image was processed slice-by-slice, feature extraction was performed using a voxel overlapping.

In this work, We evaluated a total number of 328 MR image features, which are divided in four groups as follows:(I) tumor intensity, (II) shape, (III) texture and (IV)wavelet features. The first group quantified tumor intensity characteristics using first-order statistics, calculated from the histogram of all tumor voxel intensity values. Group2 consists of features based on the shape of the tumour (for example, sphericity or compactness of the tumor). Group3 consists of textual features that are able to quantify intratumour heterogeneity differences in the texture that is observable within the tumor volume. These features are calculated in all three dimensional directions within the tumor volume, thereby taking the spatial location of each voxel compared with the surrounding voxels into account.

Group4 calculates the intensity and textural features from wavelet decompositions of the original image, thereby focusing the features on different frequency ranges within the tumor volume (figure 2). All features shown in table 1.

2.3. LASSO logistic regression model

LASSO is fittingly being used as a metaphor of L1 constraint applied to linear model. Consider the
Figure 2. Schematic of the undecimated three dimensional wavelet transform applied to each MR image. The original image X is decomposed into 8 decompositions, by directional low-pass (i.e. a scaling) and high-pass (i.e. a wavelet) filtering: XLLL, XLLH, XLHL, XLHH, XHLL, XHLH, XHHL, XHHH.

Table 1. All features extracting from MR images.

| Group          | All Feature name                                                                 |
|----------------|----------------------------------------------------------------------------------|
| Group1. First order statistics | Energy, Entropy, Kurtosis, Maximum, Mean, Mean absolute deviation, Median, Minimum, Range, Root mean square, Skewness, Standard deviation, Uniformity, Variance |
| Group2. Shape and size based features | Compactness1, Compactness2, Disproportion, Sphericity |
| Group3. Textural features | Gray-Level Co-Occurrence Matrix(GLCM) based features: Energy, Contrast, Entropy, Homogeneity, Correlation, Sum average, Variance, Dissimilarity, Autocorrelation |
| Group4. Wavelet features | Gray-Level Run-Length Matrix(GLRL) based features: SRE, LRE, GLM, RLN, RP, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, GLV, RLV |

usual linear regression model with data \((x_{i1}, x_{i2}, \ldots, x_{ip}, y_i), i = 1, \ldots, n\), where the \(x_{ij}\)'s are the regressors and \(y_i\) is the response variable of the \(i\)th observation. The ordinary Least Squares (OLS) regression method finds the unbiased linear combination of the \(x_{ij}\)'s that minimizes the residual sum of squares. However, if \(p\) is large or the regression coefficients are highly correlated (multicolinearity), the OLS may yield estimates with large variance which reduces the accuracy of the prediction. A widely-known method to solve this problem is Ridge Regression and subset selection. As an alternative to these techniques, Robert Tibshirani presented “LASSO” which minimized the residual sum of squares subject to the sum of absolute values of the coefficient being less than a constant.

\[
\hat{\beta}^L = \arg \min \{\sum_{i=1}^{N} (y_i - \alpha - \sum_j \beta_j x_{ij})^2 \}
\]

Subject to

\[
\sum_{j=1}^{p} |\hat{\beta}_j| \leq t \quad \text{(Constant)}
\]

If \(t > \sum_{j=1}^{p} |\hat{\beta}_j|\), then the LASSO algorithm will yield the same estimate as OLS estimate. However, if \(0 < t < \sum_{j=1}^{p} |\hat{\beta}_j|\), then the problem is equivalent to

\[
\hat{\beta}^L = \arg \min \{\sum_{i=1}^{N} (y_i - \alpha - \sum_j \beta_j x_{ij})^2 + \lambda \sum_j |\beta_j| \}
\]
\( \lambda > 0 \). It will be shown later that the relation between \( \lambda \) and LASSO parameter \( t \) is one-to-one. Due to the nature of the constraint, LASSO tends to produce some coefficients to be exactly zero. Compared to the OLS, whose predicted coefficient \( \hat{\beta}^O \) is an unbiased estimator of \( \beta \), both ridge regression and LASSO sacrifice a little bias to reduce the variance of the predicted values and improve the overall prediction accuracy [11].

2.4. CV methods and estimate of the LASSO parameter.

The tuning parameter \( t \) is called LASSO parameter, which is also recognized as the absolute bound. \( \sum_{j=1}^{n} |\hat{\beta}_j| = t \). Here define another parameter, \( s \), as the relative bound.

\[
\begin{align*}
s &= \frac{\sum_{j=1}^{p} |\hat{\beta}_j|}{\sum_{j=1}^{p} |\beta_j|}, s \in [0,1] \\
\end{align*}
\]

The relative bound can be seen as a normalized version of LASSO parameter. N-fold Cross-validation algorithms [12] can be used to compute the best \( s \). Cross-validation is a general procedure that can be applied to estimate tuning parameters in a wide variety of problems. The bias in RSS is a result of using the same data for model fitting and model evaluation. CV can reduce the bias of RSS by splitting the whole data into two subsamples: a training (calibration) sample for model fitting and a test (validation) sample for model evaluation. The idea behind the cross-validation is to recycle data by switching the roles of training and test samples.

The optimal \( s \) can be denoted by \( \hat{s} \). Prediction error can be estimated for the LASSO procedure by ten-fold cross-validation. The LASSO is indexed in terms of \( s \), and the prediction error is estimated over a grid of values of \( s \) from 0 to 1 inclusive. We wish to predict with small variance, thus we wish to choose the constraint \( s \) as small as we can. The value \( \hat{s} \) which achieves the minimum predicted error of \( \hat{\eta}(x) \) is selected.

**Algorithm 1.** An N-fold cross-validation selects a model.

1. Split the whole data into \( N \) disjoint subsamples \( S_1 \cdots S_N \)
2. For \( i = 1, \cdots N \) fit model to the training sample \( \cup_{i \neq \nu} S_i \), and compute discrepancy, \( d_\nu(s) \) using the test sample \( S_\nu \). \( d_\nu(s) = |PE_{S_\nu} - PE_{\cup_{i \neq \nu} S_i}|^2 \).
3. Find the optimal \( \hat{s} \) as the minimizer of the overall discrepancy \( d(s) = \sum_{\nu=1}^{N} d_\nu(s) \).

3. Experiments and results

3.1. Data set

This was a retrospective study and the ethical approval was obtained. 170 consecutive patients with histologically confirmed HCC who underwent surgical resection at our institute were recruited retrospectively between February 2012 and December 2016. The inclusion criteria were: (1) operation-proven untreated HCC, (2) underwent liver MR examination within a week before operation and the image quality meet the requirement of analysis, (clearly, more than 3 slices on the axial image), (3) surgical resection with histopathological confirmation. In total, 170 patients were divided into two groups including training data set [125 patients] and the validation dataset [45 patients].

The histology grading data was retrieved from the archived clinical histology report, in which the histological grades of the HCCs were noted. Low-grade tumour corresponds to Edmondson grades I, II and II, and high-grade tumour corresponds to Edmondson grades II-III, III, III-IV and IV. Figure 3 shows the MR image of HCC.

3.2. Results

All feature algorithms were implemented in Matlab2014. Totally 328 features extracted from T1WI and 328 features extracted from T2WI. LASSO algorithm was used to select the most useful features. The Radiomics score (Rad-score) of each patient was calculated by using the linear combination of selected features multiplying their respective coefficients. 10-fold cross-validation was applied to
training dataset and selecting the optimal model of the pathological grading of HCC. The results of HCC histology grading in training dataset and validation (Test) dataset shown in Table 2. Features selection using the LASSO logistic model shown in figure 4. The predictive performance for the discrimination of HCC histology grading presented as ROC were described in figure 5. All of the LASSO programs are implemented with R language.

Figure 3. Axial slice of a patient with Hepatocellular Carcinoma in T1WI MRI, the left is high-grade tumor, the right is low-grade tumor.

Figure 4. Features selection using the LASSO logistic model. Using the LASSO logistic model, the AUC was drawn versus log ($\lambda$). The optimal values (the vertical lines) of the lasso tuning parameter ($\lambda$) can be chosen using either the 1SE criteria or the minimum criteria by generalized cross-validation. In this study, $\lambda$ was selected based on the ten-fold cross-validation used the minimum criteria. (A) 31 features were selected using the lasso logistic regression model from T2WI. (B) 17 features were selected using the lasso logistic regression model from T1WI. (C) 18 features were selected using the lasso logistic regression model from T1WI combined T2WI.
Figure 5. Receiver operating characteristic (ROC) curves of (A) T1WI; (B) T2WI; (C) T1WI combined T2WI. Figure 5a-5c, the left represents the ROC curves on training dataset, the right represents the ROC curves on test dataset.
### Table 2. Performance of the model.

|                  | Training dataset (n=125) | Test dataset (n=45) |
|------------------|--------------------------|---------------------|
|                  | Cutoff | Sensitivity | Specificity | Accuracy | AUC (95%CI) | P Value | Sensitivity | Specificity | Accuracy | AUC (95%CI) | P Value |
| Lasso (T1)       | 0.639  | 57.14%      | 85.45%      | 69.6%    | 0.812      | <0.0001 | 56%        | 70%        | 62.22%   | 0.712      | 0.0091  |
| (T1)             |        | (0.732 to 0.876) |           |          |           |         |            |            |          | (0.558 to 0.837) |         |
| Lasso (T2)       | 0.109  | 81.43%      | 63.64%      | 73.6%    | 0.804      | <0.0001 | 64%        | 70%        | 66.67%   | 0.722      | 0.0059  |
| (T2)             |        | (0.723 to 0.869) |           |          |           |         |            |            |          | (0.568 to 0.845) |         |
| Lasso (T1+T2)    | 0.220  | 81.44%      | 70.96%      | 76.8%    | 0.829      | <0.0001 | 60%        | 80%        | 68.89%   | 0.742      | 0.0015  |
| (T1+T2)          |        | (0.752 to 0.891) |           |          |           |         |            |            |          | (0.590 to 0.861) |         |

### 4. Discussion

The Radiomics model based on T1WI achieved an AUC of 0.812 in the training data-set and AUC of 0.712 in the validation data-set (figure 5a). The Radiomics model based on T2WI achieved an AUC of 0.804 in the training data-set and AUC of 0.722 in the validation data-set (figure 5b). The Radiomics model based on T1WI combined T2WI achieved the highest AUC of 0.829 in the training data-set and AUC of 0.742 in the validation data-set (figure 5c).

There’s significant difference between the median of Radiomics signature of high-stage patients and that of the low-stage patients, in both the training dataset (p < 0.0001) and the validation dataset (p < 0.01). The Radiomics signature both based on T1WI and T2WI presented good performance for the discrimination of high-stage patients and low-stage patients.

Compared with basic MR images traits, Radiomics features capture more information about intra-tumour heterogeneity objectively and quantitatively at low cost, correlate with underlying gene-expression patterns, and may help predict clinical outcomes. In our study, a few potential features were chosen from 328 candidate Radiomics features via the LASSO method to build a Radiomics signature (Rad-score), which proved to be an independent predictor for HCC histopathological grading.

### References

[1] Chen W, Zheng R and Baade P D 2016 Cancer statistics in China CA Cancer J. Clin 66 115
[2] Aerts H J, Velazquez ER and Leijenaar RT 2014 Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach Nat. Commun 5 4006
[3] Lambin P 2012 Radiomics: extracting more information from medical images using advanced feature analysis Eur. J. Cancer 48 441
[4] Ganeshan B, Skogen K, Pressney I, Coutroubis D and Miles K 2012 Tumour heterogeneity in oesophageal cancer assessed by CT texture analysis: preliminary evidence of an association with tumour metabolism, stage, and survival Clin. Radiol 67 157
[5] Gevaert O 2012 Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data: methods and preliminary results Radiology 264 387
[6] Kumar V 2012 Radiomics: the process and the challenges Magn.Reson.Imaging 30 1234
[7] Doroshow J and Kummer S 2014 Translational research in oncology 10 years of progress and future prospects Nat. Rev. Clin. Oncol 11 649
[8] Fan J Q and Li R Z 2001 Variable Selection via Penalized Likelihood J. A. Stat. Ass 96 1348
[9] Goeman J 2010 L1 Penalized Estimation in the Cox Proportional Hazards Model J. Biom 52 70
[10] Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S and Miles K 2012 Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival Eur. Radiol 22 796
[11] Friedman J, Hastie T and Tibshirani R 2010 Regularization paths for generalized linear models via coordinate descent J. Stat. Soft 33 1
[12] Fu W J 1998 Penalized Regressions: The Bridge Versus the Lasso J. Comp. and Graph. Stat 7 397