Classification of malignant and benign liver tumors using a radiomics approach

Starmans, Martijn P.A.; Miclea, Razvan L.; Van Der Voort, Sebastian R.; Niessen, Wiro; Thomeer, Maarten G.; Klein, Stefan

DOI
10.1117/12.2293609

Publication date
2018

Document Version
Final published version

Published in
Medical Imaging 2018

Classification of malignant and benign liver tumors using a radiomics approach

Martijn P. A. Starmans, Razvan L. Miclea, Sebastian R. van der Voort, Wiro J. Niessen, Maarten G. Thomeer, et al.

Martijn P. A. Starmans, Razvan L. Miclea, Sebastian R. van der Voort, Wiro J. Niessen, Maarten G. Thomeer, Stefan Klein, "Classification of malignant and benign liver tumors using a radiomics approach," Proc. SPIE 10574, Medical Imaging 2018: Image Processing, 105741D (2 March 2018); doi: 10.1117/12.2293609

SPIE.

Event: SPIE Medical Imaging, 2018, Houston, Texas, United States

Classification of malignant and benign liver tumors using a radiomics approach

Martijn P. A. Starmans[a, b], Razvan L. Miclea[a], Sebastian R. van der Voort[a, b], Wiro J. Niessen[a, b, c], Maarten G. Thomeer[a], and Stefan Klein[a, b]

[a]Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, the Netherlands
[b]Department of Medical Informatics, Erasmus MC, Rotterdam, the Netherlands
[c]Faculty of Applied Sciences, Delft University of Technology, the Netherlands

ABSTRACT

Correct diagnosis of the liver tumor phenotype is crucial for treatment planning, especially the distinction between malignant and benign lesions. Clinical practice includes manual scoring of the tumors on Magnetic Resonance (MR) images by a radiologist. As this is challenging and subjective, it is often followed by a biopsy. In this study, we propose a radiomics approach as an objective and non-invasive alternative for distinguishing between malignant and benign phenotypes. T2-weighted (T2w) MR sequences of 119 patients from multiple centers were collected. We developed an efficient semi-automatic segmentation method, which was used by a radiologist to delineate the tumors. Within these regions, features quantifying tumor shape, intensity, texture, heterogeneity and orientation were extracted. Patient characteristics and semantic features were added for a total of 424 features. Classification was performed using Support Vector Machines (SVMs). The performance was evaluated using internal random-split cross-validation. On the training set within each iteration, feature selection and hyperparameter optimization were performed. To this end, another cross validation was performed by splitting the training sets in training and validation parts. The optimal settings were evaluated on the independent test sets. Manual scoring by a radiologist was also performed. The radiomics approach resulted in 95% confidence intervals of the AUC of [0.75, 0.92], specificity [0.76, 0.96] and sensitivity [0.52, 0.82]. These approach the performance of the radiologist, which were an AUC of 0.93, specificity 0.70 and sensitivity 0.93. Hence, radiomics has the potential to predict the liver tumor benignity in an objective and non-invasive manner.

Keywords: hepatocellular carcinoma, computer-aided diagnosis, radiomics, classification, primary liver tumors

1. INTRODUCTION

Treatment planning in liver cancer is based on the tumor phenotype.[1,2] In practice, a first assessment of the phenotype is performed by a radiologist, based on MR images. Several guidelines for manual liver tumor scoring based on MR images have been developed.[3] However, the appearance of these lesions on the images may vary considerably due to differences in their histological features. A clear consensus on their exact definitions therefore does not exist, which can lead to subjective results.[3] Tissue biopsy is therefore often mandatory for suspected malignant tumors. This procedure is invasive, can be technically difficult and brings risks such as tumor seeding and haemorrhage to the patient.[4]

Malignant lesions which are identified early have multiple treatment options, thereby improving survival.[2] Early diagnosis is therefore crucial for treatment success. Presently, the primary task of a radiologist is to evaluate whether a tumor is malignant or not based on the MR images. Patients with suspected malignant lesions are followed or referred to tertiary medical centers, often receiving a biopsy for reasons mentioned above. Patients with suspected benign lesions generally receive no further treatment. False negatives, i.e. malignant lesions which are classified as benign, are therefore often lost in follow-up.

Further author information: (Send correspondence to M.P.A. Starmans)
E-mail: m.starmans@erasmusmc.nl

Medical Imaging 2018: Image Processing, edited by Elsa D. Angelini, Bennett A. Landman, Proc. of SPIE
Vol. 10574, 105741D · © 2018 SPIE · CCC code: 1605-7422/18/$18 · doi: 10.1117/12.2293609

Classification of primary liver tumors based on MR images is complicated due to the varying appearance of lesions. Diagnosis of these patients in tertiary hospitals is often done by a radiologist specialized in abdominal imaging, thereby having a relatively high chance of successful classification. However, in primary or secondary healthcare centers, radiologists are frequently not specialized in this area. Hence, especially for the peripheral centers, there is a need for an improved method of characterizing primary liver tumors.

Radiomics can be used as a non-invasive alternative to determine the relation between quantitative medical image features and underlying biological phenomena.[5] The field of Radiomics is rapidly emerging and has shown promising results in areas such as lung cancer,[6–8] glioblastoma[9] and breast cancer.[10] Most studies focus on the use of computed tomography imaging.

The purpose of this study was to develop a primary liver tumor classification model which is able to distinguish between malignant and benign phenotypes using radiomics, based on their appearance on T2w MR images. Hepatocellular carcinoma (HCC), the most common primary liver tumor,[11] was included in the malignant class. The benign class consisted of Focular Nodular Hyperplasia (FNH) and Hepatocellular Adenoma (HCA), two of the most common benign solid hepatic liver tumors.[1] Lesions were semi-automatically segmented by an abdominal radiologist. Feature extraction, selection and classification were performed using a radiomics approach.

2. METHODS

2.1 DATASET

MR and pathology data for 119 patients were obtained from our tertiary referral medical center. As stated in the introduction, our clinical use case is the diagnosis of lesions based on the MR scan. Several cases are relatively simple for a radiologist to classify, hence an automatic method would have little added value. Inclusion and exclusion criteria were therefore set to exclude the less challenging cases.

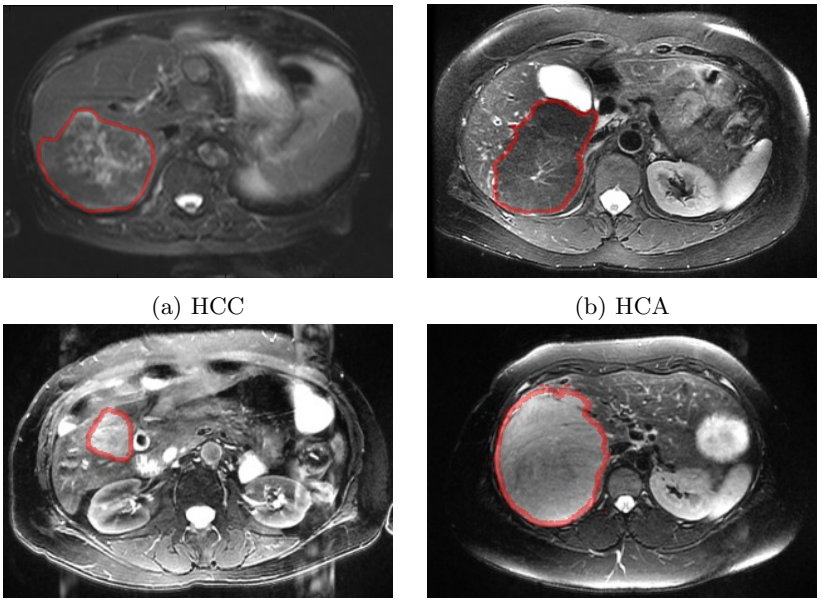
Lesions with a maximum diameter equal to or smaller than 3 cm were excluded, as these have a relatively low chance of being primal liver tumors.[12] Patients with cirrhosis were excluded, since the a priori chance of a liver tumor being HCC in these patients is by far the largest.[13] Diagnosis of cirrhosis was based on a combination of clinical and imaging findings. For the patients with a HCC, cirrhosis was always determined through a biopsy. In patients with multiple lesions, only the largest was included.

A T2w sequence was selected for each patient, as abdominal radiologists initially use the T2w sequence for assessment in clinical practice. If available, the T2w sequence with Fat-Sat was used. The T2w Fat-Sat sequences in our database are generally multi-shot scans, thereby having a high Signal to Noise Ratio (SNR) and thus high quality for assessment. If this sequence was not available, the T2w sequence without Fat-Sat was used, as this sequence is most similar. These are mostly single-shot sequences in our database. Reversion to the T2w sequence without Fat-Sat occurred in 23 patients (19 %)

The resulting dataset consisted of 50 HCC, 46 Adenoma, 23 FNH patients, resulting in 50 (42%) malignant and 69 (58%) benign tumors. The data originates from 42 different institutes, translating to heterogeneity in the imaging protocols. This is for example reflected in the used magnetic field strength (105 images with 1.5 T, 14 with 3T), slice thickness, repetition time and echo time, see Table 1.

Setting (unit)	Mean	Std	Min	Max
Slice thickness (mm)	7	1	5	10
Repetition time (ms)	4531	3754	337	13846
Echo time (ms)	93	19	55	180

Table 1: Several properties of the imaging protocols used for the T2w (Fat-Sat) MR sequences in this study.



(a) HCC (b) HCA

(c) HCC: Pitfall (d) HCA: Pitfall

Figure 1: 2-D coupes from T2w (Fat-Sat) MR sequences showing segmented malignant (left) and benign (right) liver tumors. Two typical classification examples (top) and two pitfalls (i.e. cases which where confidently misclassified by both our model and the radiologist) (bottom) are depicted.

2.2 SEGMENTATION

Regions of interest (ROIs) corresponding to the tumor were semi-automatically extracted by an abdominal radiologist, see Figure 1. To this end, we propose an efficient segmentation method using image registration. As all ROIs are interactively checked, the result is similar to manual annotation.

First, the user delineated the lesion on a single slice. Next, image registration was used to warp the segmented slice to an adjacent one. An automatic intensity based registration method was used, starting with an affine transformation model, followed by a deformable registration with a B-spline transformation model. Mean squared difference was used as similarity metric with adaptive stochastic gradient descent[14] as optimization strategy. Registration was implemented using SimpleElastix.[15] Next, the user corrected the transformed contour if required. The procedure was repeated until all tumor slices were contoured.

2.3 RADIOMICS

State-of-the-art radiomics features[16] quantifying shape, texture, heterogeneity (Co-occurrence of Local Anisotropic Gradient Orientations (CoLlAGe)[17]) and intensity were extracted from the ROI. Texture features were constructed using the Gray Level Co-occurence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), Gabor filters and Local Binary Patterns (LBP).[18] A 3-D ellipsoid was fitted to the tumor contour to provide three orientation features. Patient age, sex and three semantic features (liver segment, maximum diameter, atoll sign[19]) were added. The semantic features were manually scored by a radiologist.

Accurate extraction of quantitative 3-D tumor characteristics is hampered by a large slice thickness. Furthermore, the varying slice thickness in our dataset could influence the features. Hence, in order to create a robust representation, features were extracted per 2-D tumor slice in the main scan direction (e.g. axial). The feature values were aggregated over all slices, after which first order statistics of the resulting values were computed. The slice thickness was added as a separate feature to aid the classification. In total, the total number of features per patient was 424.

Classification was performed using Support Vector Machines (SVMs)[20] with a polynomial kernel. Cross-validation was performed by randomly splitting the data in 80% training and 20% test data among 100 iterations. Stratified splits were used to maintain the initial class balance. Within each cross validation, a 5-fold cross validation was performed on the training set to optimize several hyperparameters. The F1-score was used as optimization metric. Hence in each of these folds, part of the training set is used to train the model, while the other part serves as a validation set.

Within these cross validations, the features were first scaled using z-scoring fitted to the training set. Features with a small variance across training patients (< 0.01) were removed. Next, feature selection was performed in these cross-validations using an exhaustive search among all possible combinations of the feature groups. For each feature group (e.g. shape, histogram), a hyperparameter was included to define if the group was selected or not. As the texture feature group contains several different types, this hyperparameter could select either all, none or only a single group (e.g. LBP) of these features. Lastly, the polynomial degree, slack and kernel homogeneity parameters of the SVM were also included in the hyperparameters in this search. The best hyperparameter set was used to refit a classifier on the full training set in that cross validation iteration. This resulted in a single optimal hyperparameter setting for each cross validation iteration, which can vary per split.

3. RESULTS

Performance was evaluated using the Area Under the Curve (AUC) of the Receiver Operating Curve (ROC), sensitivity, specificity and the F1-score. Performance of the predictive model is given in 95% corrected confidence intervals (CIs). These were computed using the corrected resampled t-test, which takes into account that the cross-validation samples are not statistically independent.[21] The malignant class was defined as the positive class.

For comparison, the tumors were also scored manually, on a ten point scale, by an abdominal radiologist with over 15 years of experience, specialized in liver imaging. Next to the T2w MR sequence, the radiologist was also shown the age and gender of each patient. Based on the ten point manual scoring scale, different decision boundaries can be chosen. The performance of the radiologist was evaluated for two different decision boundaries. The first boundary was manually set by the radiologist, which corresponds to clinical practice and will therefore further be indicated as the "normal" boundary. The second boundary was set to optimize the F1-score of the radiogist, as this was also used in the hyperparameter and feature selection optimization in the radiomics approach. Both performances are shown in Table 1. The ROC curves are shown in Figure 2.

In terms of AUC, the radiologist does outperform our method. However, the maxima of the CIs of our method for all metrics are near the scores of the radiologist at the optimal F1-score. Both methods are skewed towards specificity, which could be caused by the imbalance in our dataset. In contrast, the normal performance of the radiologist is skewed towards sensitivity. Compared to the performance using this decision boundary, the radiologist outperforms the mean of our model only by a few percent based on the F1-score.

The results of the exhaustive feature selection search are shown in Figure 3. The patient features, which included the age and gender of the patient, were included in almost all of the optimal parameter settings (97%). These features are thus important for the classification, corresponding with findings by other studies.[2] The number of times the other features are selected is close to the expectation values determined by random chance. This is equal to half of the iterations or 50% times for the histogram, shape, orientation and CoLiAGe features, one sixth of the iterations or about 17% times for the texture feature options. The differences between the actual number of selections and these expected numbers are small. This could therefore be attributed to random incidence. Hence no clear pattern could be distinguished for feature selection in these groups.

4. NEW OR BREAKTHROUGH WORK TO BE PRESENTED

We developed a radiomics model as a non-invasive approach to classify primary liver tumors based on T2w MR images. Additionally, we proposed an efficient semi-automatic segmentation workflow. The performance of the

	radiomics	rad. F1	rad. normal
AUC	[0.75, 0.92]	0.93	0.93
F1-score	[0.70; 0.85]	0.87	0.79
Sensitivity	[0.52, 0.82]	0.80	0.93
Specificity	[0.76, 0.96]	0.97	0.70

Table 2: Classification results of the proposed radiomics setup and manual scoring by a radiologist. Results for the radiologist are shown for the decision boundaries optimal for the F1-score (rad. F1) and for normal clinical practice (rad. normal).

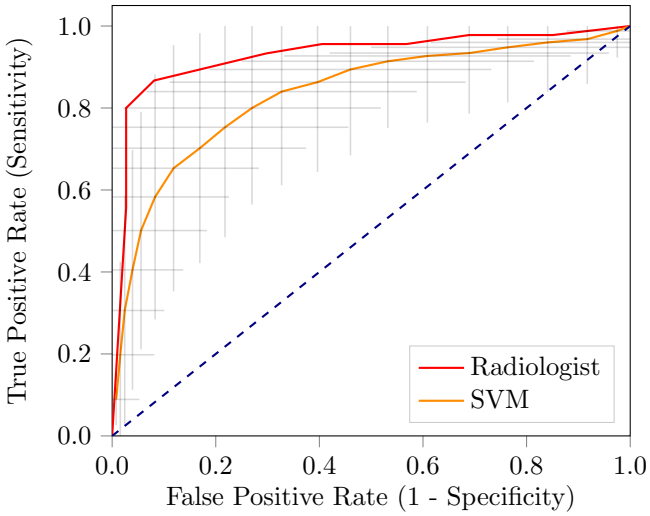


Figure 2: ROC curves showing the classification results by the radiologist (red, AUC = 0.93) and the CIs of the SVM (yellow, AUC = [0.75, 0.92]). The gray crosses give the false and true positive rate CIs.

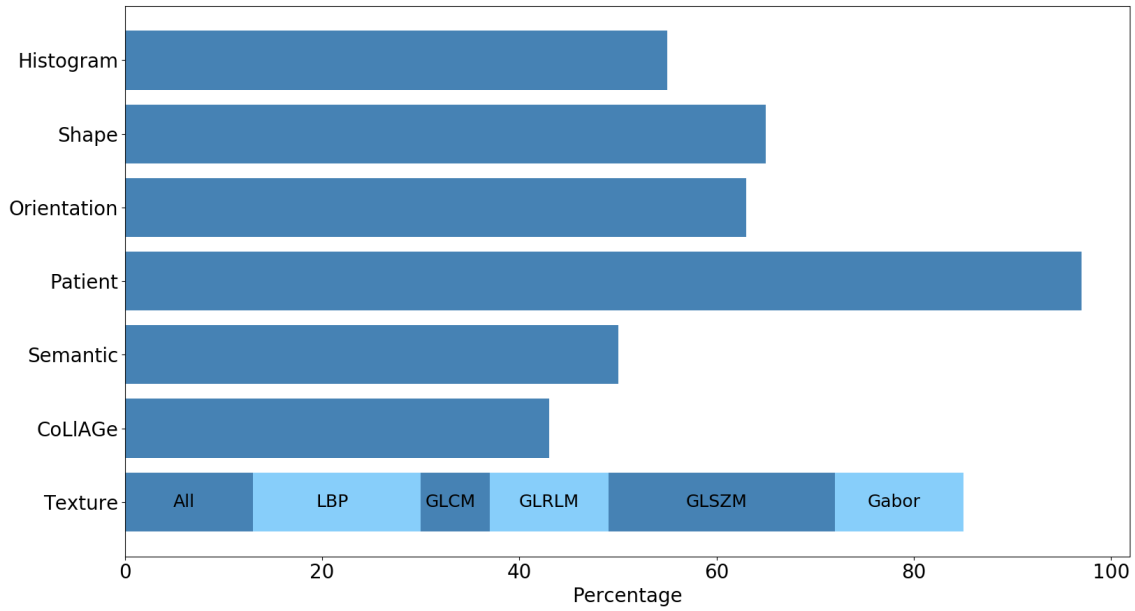


Figure 3: Percentage which a feature group (y-axis) was included in the optimal parameter settings in the best classifier for 100 cross-validation iterations.

resulting combination approaches scoring by an abdominal radiologist. Many radiomics studies only use local institutional data with small variations in the imaging protocols, thereby not hampering the feature extraction.[22] Instead, we showed that our methods work on a broad, multi-institutional dataset with a large heterogeneity in imaging protocol parameters.

5. DISCUSSION AND CONCLUSIONS

In this study, we have shown that radiomics is a promising, non-invasive approach for the classification of primary liver tumors. Performance of our model approaches that of an abdominal radiologist, showing the potential for usage in clinical practice.

Several limitations of this study should be mentioned. First off, in practice, the radiologist uses additional sequences for the tumor scoring, such as T1 weighted MR or diffusion weighted imaging. This might improve the performance of the radiologist. However, the addition of these sequences might also increase the performance of our model. Second, a single observer performed the segmentation, while radiomics features can be sensitive to the segmentation method.[23] Future work therefore includes the segmentations of a second observer. As the tumors show a clear edge on the images, we do however not expect a significant difference. Additionally, the performance of our model was in this study compared to scoring by a radiologist specialized in liver imaging with many years of experience. As our user case is focused at the peripheral radiologist, a comparison with manual scoring such a radiologist would be more interesting.

Lastly, the guidelines for liver tumor classification based on MR images lists several specific features which are important. However, in our feature selection optimization, besides the importance of the patient age and sex, no clear pattern in the feature selection could be found. A possible explanation for these could be the correlation between features. Many of these features and groups are able to quantify the same underlying information. Especially the texture feature groups have a large overlap in the type of patterns they are able to quantify. Hence, many of these features are highly correlated, causing a part to be redundant in the classification. In the SVM classification, it therefore does not matter which of these highly correlated features is selected, as long as the required information is present. Eliminating redundant features before the selection could give us more insight in the type of features that are important for classification.

Future work includes development of an automatic segmentation method, as semi-automatic segmentation still can be a tedious process. Additionally, more imaging sequences besides T2w MR will be added. Lastly, our radiomics approach will be tested on an external dataset.

6. ACKNOWLEDGEMENTS

This work is part of the research program STRaTeGy with project number 14929 - 14930, which is (partly) financed by the Netherlands Organisation for Scientific Research (NWO).

REFERENCES

[1] Grazioli, L., Bondioni, M. P., Haradome, H., Motosugi, U., Tinti, R., Frittoli, B., Gambarini, S., Donato, F., and Colagrande, S., "Hepatocellular adenoma and focal nodular hyperplasia: value of gadoxetic acidenhanced MR imaging in differential diagnosis," *Radiology* **262**(2), 520–529 (2012).

[2] Balogh, J., Victor, D., Asham, E. H., Burroughs, S. G., Boktour, M., Saharia, A., Li, X., Ghobrial, M., and Monsour, H., "Hepatocellular carcinoma: a review," *Journal of Hepatocellular Carcinoma* **Volume 3**, 41–53 (2016).

[3] Mitchell, D. G., Bruix, J., Sherman, M., and Sirlin, C. B., "LI-RADS (Liver Imaging Reporting and Data System): Summary, discussion, and consensus of the LI-RADS Management Working Group and future directions," *Hepatology* **61**(3), 1056–1065 (2015).

[4] Silva, M. A., Hegab, B., Hyde, C., Guo, B., Buckels, J. A. C., and Mirza, D. F., "Needle track seeding following biopsy of liver lesions in the diagnosis of hepatocellular cancer: a systematic review and meta-analysis," *Gut* **57**(11), 1592–1596 (2008).

[5] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A., and Aerts, H. J., "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer* **48**(4), 441–446 (2012).

[6] Scrivener, M., de Jong, E. E. C., van Timmeren, J. E., Pieters, T., Ghaye, B., and Geets, X., "Radiomics applied to lung cancer: a review," *Translational Cancer Research* **5**(4), 398–409 (2016).

[7] Gevaert, O., Xu, J., Hoang, C. D., Leung, A. N., Xu, Y., Quon, A., Rubin, D. L., Napel, S., and Plevritis, S. K., "Nonsmall cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray datamethods and preliminary results," *Radiology* **264**(2), 387–396 (2012).

[8] Aerts, H. J. W. L., Velazquez, E. R., Leijenaar, R. T. H., Parmar, C., Grossmann, P., Cavalho, S., Bussink, J., Monshouwer, R., Haibe-Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M. M., Leemans, C. R., Dekker, A., Quackenbush, J., Gillies, R. J., and Lambin, P., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications* **5** (2014).

[9] Gevaert, O., Mitchell, L. A., Achrol, A. S., Xu, J., Echegaray, S., Steinberg, G. K., Cheshier, S. H., Napel, S., Zaharchuk, G., and Plevritis, S. K., "Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features," *Radiology* **273**(1), 168–174 (2014).

[10] Jerez, J. M., Molina, I., Garca-Laencina, P. J., Alba, E., Ribelles, N., Martn, M., and Franco, L., "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial Intelligence in Medicine* **50**(2), 105–115 (2010).

[11] Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A., "Global cancer statistics, 2012," *CA: A Cancer Journal for Clinicians* **65**(2), 87–108 (2015).

[12] Befeler, A. S. and di Bisceglie, A. M., "Hepatocellular carcinoma: Diagnosis and treatment," *Gastroenterology* **122**(6), 1609–1619 (2002).

[13] Oka, H., Kurioka, N., Kim, K., Kanno, T., Kuroki, T., Mizoguchi, Y., and Kobayashi, K., "Prospective study of early detection of hepatocellular carcinoma in patients with cirrhosis," *Hepatology* **12**(4), 680–687 (1990).

[14] Klein, S., Pluim, J. P. W., Staring, M., and Viergever, M. A., "Adaptive Stochastic Gradient Descent Optimisation for Image Registration," *International Journal of Computer Vision* **81**(3), 227–239 (2009).

[15] Marstal, K., Berendsen, F., Staring, M., and Klein, S., "SimpleElastix: A user-friendly, multi-lingual library for medical image registration," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*], 134–142 (2016).

[16] Parekh, V. and Jacobs, M. A., "Radiomics: a new application from established techniques," *Expert Rev Precis Med Drug Dev* **1**(2), 207–226 (2016).

[17] Prasanna, P., Tiwari, P., and Madabhushi, A., "Co-occurrence of Local Anisotropic Gradient Orientations (CoLlAGe): A new radiomics descriptor," *Sci Rep* **6** (2016).

[18] Ojala, T., Pietikainen, M., and Maenpaa, T., "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7), 971–987 (2002).

[19] Thomeer, M. G., Broker, M. E. E., Lussanet, Q. d., Biermann, K., and Dwarkasing, R. S., "Genotype-phenotype correlations in hepatocellular adenoma: an update of MRI findings," *Diagnostic and Interventional Radiology* **20**(3), 193–199 (2014).

[20] Cortes, C. and Vapnik, V., "Support-vector networks," *Machine learning* **20**(3), 273–297 (1995).

[21] Nadeau, C. and Bengio, Y., "Inference for the generalization error," in [*Advances in neural information processing systems*], 307–313 (2000).

[22] Bai, H. X., Lee, A. M., Yang, L., Zhang, P., Davatzikos, C., Maris, J. M., and Diskin, S. J., "Imaging genomics in cancer research: limitations and promises," *The British Journal of Radiology* **89**(1061), 20151030 (2016).

[23] Parmar, C., Rios Velazquez, E., Leijenaar, R., Jermoumi, M., Carvalho, S., Mak, R. H., Mitra, S., Shankar, B. U., Kikinis, R., Haibe-Kains, B., Lambin, P., and Aerts, H. J. W. L., "Robust Radiomics Feature Quantification Using Semiautomatic Volumetric Segmentation," *PLoS ONE* **9**(7), e102107 (2014).