



Liver shape analysis using partial least squares regression-based statistical shape model: application for understanding and staging of liver fibrosis

Mazen Soufi¹ · Yoshito Otake¹ · Masatoshi Hori² · Kazuya Moriguchi¹ · Yasuharu Imai³ · Yoshiyuki Sawai³ · Takashi Ota² · Noriyuki Tomiyama² · Yoshinobu Sato¹

Received: 31 January 2019 / Accepted: 21 October 2019

© CARS 2019

Abstract

Purpose Liver shape variations have been considered as feasible indicators of liver fibrosis. However, current statistical shape models (SSM) based on principal component analysis represent gross shape variations without considering the association with the fibrosis stage. Therefore, we aimed at the application of a statistical shape modelling approach using partial least squares regression (PLSR), which explicitly uses the stage as supervised information, for understanding the shape variations associated with the stage as well as predicting it in contrast-enhanced MR images.

Methods Contrast-enhanced MR images of 51 patients with fibrosis stages F0/1 ($n = 18$), F2 ($n = 15$), F3 ($n = 7$) and F4 ($n = 11$) were used. The livers were manually segmented from the images. An SSM was constructed using PLSR, by which shape variation modes (scores) that were explicitly associated with the reference pathological fibrosis stage were derived. The stage was predicted using a support vector machine (SVM) based on the PLSR scores. The performance was assessed using the area under receiver operating characteristic curve (AUC).

Results In addition to commonly known shape variations, such as enlargement of left lobe and shrinkage of right lobe, our model represented detailed variations, such as enlargement of caudate lobe and the posterior part of right lobe, and shrinkage in the anterior part of right lobe. These variations qualitatively agreed with localized volumetric variations reported in clinical studies. The accuracy (AUC) at classifications F0/1 versus F2–4 (significant fibrosis), F0–2 versus F3–4 and F0–3 versus F4 (cirrhosis) were 0.90 ± 0.03 , 0.80 ± 0.05 and 0.82 ± 0.05 , respectively.

Conclusions The proposed approach offered an explicit representation of commonly known as well as detailed shape variations associated with liver fibrosis stage. Thus, the application of PLSR-based SSM is feasible for understanding the shape variations associated with the liver fibrosis stage and predicting it.

Keywords Statistical shape model · Shape analysis · Liver fibrosis staging · Partial least squares regression

Introduction

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11548-019-02084-z>) contains supplementary material, which is available to authorized users.

✉ Yoshinobu Sato
yoshi@is.naist.jp

¹ Division of Information Science, Nara Institute of Science and Technology, 8916-5, Takayama-cho, Ikoma, Nara 630-0192, Japan

² Department of Diagnostic and Interventional Radiology, Osaka University Graduate School of Medicine, D1, 2-2 Yamadaoka, Suita, Osaka 565-0871, Japan

³ Department of Gastroenterology, Ikeda Municipal Hospital, 3-1-18, Jonan, Ikeda, Osaka 563-8510, Japan

Liver fibrosis is a consequence of a reversible wound-healing response mainly arising from viral infections (e.g., hepatitis B and C), or other chronic liver diseases [1]. It might ultimately progress into cirrhosis that is associated with prominent morbidity and mortality [2]. The diagnosis of liver fibrosis involves staging the degree of fibrosis with one of five stages (from F0 ‘no fibrosis’ to F4 ‘cirrhosis’) [3]. The early detection of liver fibrosis is important because it helps in reversing the fibrosis, for example, through the administration of antiviral treatments [4].

Biopsy is considered the gold standard in assessing liver fibrosis. However, it is invasive, has inherent risks [5] and is prone to sampling errors. Other methods include serum/blood

markers [6] and elastography imaging techniques, such as magnetic resonance elastography (MRE) [7]. However, these tests might yield false positives due to other complications, such as congestion or inflammations [8].

Diagnostic image-based features, such as texture features [9] or representations learned via deep learning by using convolutional neural networks (CNNs) [10, 11], were found feasible for staging of liver fibrosis. However, texture features are sensitive to image characteristics, e.g., number of quantization levels, which causes a limited reproducibility under changes in the imaging parameters [9, 12]. Choi et al. [11] developed a CNN-based approach using computed tomography (CT) images that yielded AUCs of 0.95–0.97. However, CNN-based approaches have a limited interpretability in terms of the derived representations [13]. Furthermore, CNN-based systems demand large datasets with annotated-labelled data for the training process, which is a challenging task.

Liver shape variations, such as irregular surface and volumetric shrinkage/enlargement, have been considered as feasible indicators of liver fibrosis [14–19]. Commonly known volumetric variations include enlargements in caudate lobe and lateral segment of left lobe along with shrinkage in right lobe [14]. Ozaki et al. [15] have further reported detailed variations, such as enlargement in the posterior part of right lobe and shrinkage in its anterior part in cirrhotic livers. Hunt et al. [16] also reported enlargements at caudate process of caudate lobe in cirrhotic livers. Nevertheless, modelling those shape variations associated with liver fibrosis remains a non-trivial task.

In order to model the shape variation modes, a dimensionality reduction approach can be employed. A comprehensive survey on dimensionality reduction approaches can be found in [20]. A commonly used approach is principal component analysis (PCA) [17–19]. Hori et al. [17] extracted shape features using a statistical shape model (SSM) that characterized the liver shape variations in CT images based on PCA. However, PCA considers the gross variations in the data without considering the relationship between the shape features (scores) and the response variable (fibrosis stage) [21], which is regarded as supervised information. Hence, PCA is considered an unsupervised dimensionality reduction approach [20]. Thus, the association between the shape variabilities and the liver fibrosis stage may not be represented in the SSM, which might affect the performance of subsequent classifiers using the derived shape features. Hori et al. [17] addressed this problem by additionally estimating fibrosis stage scores from the PCA-based shape features using a support vector regression technique.

On the other side, supervised dimensionality reduction approaches consider the class separability of the response variable while deriving the variation modes [20]. These approaches include least absolute shrinkage and selection operator (LASSO) [22], elastic net [23], partial least squares

regression (PLSR) [24–27] and linear discriminant analysis (LDA) [28, 29]. Among these approaches, PLSR and LDA have shown success in building SSMs [26–29]. However, LDA has a limited performance on small datasets and assumes the training data follows a normal distribution [28, 29]. PLSR has demonstrated feasibility in modelling disease-relevant variations from medical images [26, 27]. For instance, Lekadir et al. [26] developed a PLSR-based SSM for prediction of myocardial infarction. They achieved a prediction accuracy >90% using a small number of modes, even when applied to a small number of training datasets [26].

In this study, we aimed at application of an SSM approach based on PLSR to understand the shape variations associated with liver fibrosis stage and predict it. The contributions are twofold. First, we proposed to use the fibrosis stage as a supervised information in two steps, i.e., the derivation of the shape features (scores) by supervised dimensionality reduction, that is using PLSR, and the prediction of the fibrosis stage via a supervised classifier. Second, we provide an explicit representation for commonly known as well as detailed shape variations associated with liver fibrosis based on the SSM, which was, up to our knowledge, not provided in similar studies [9–11, 17].

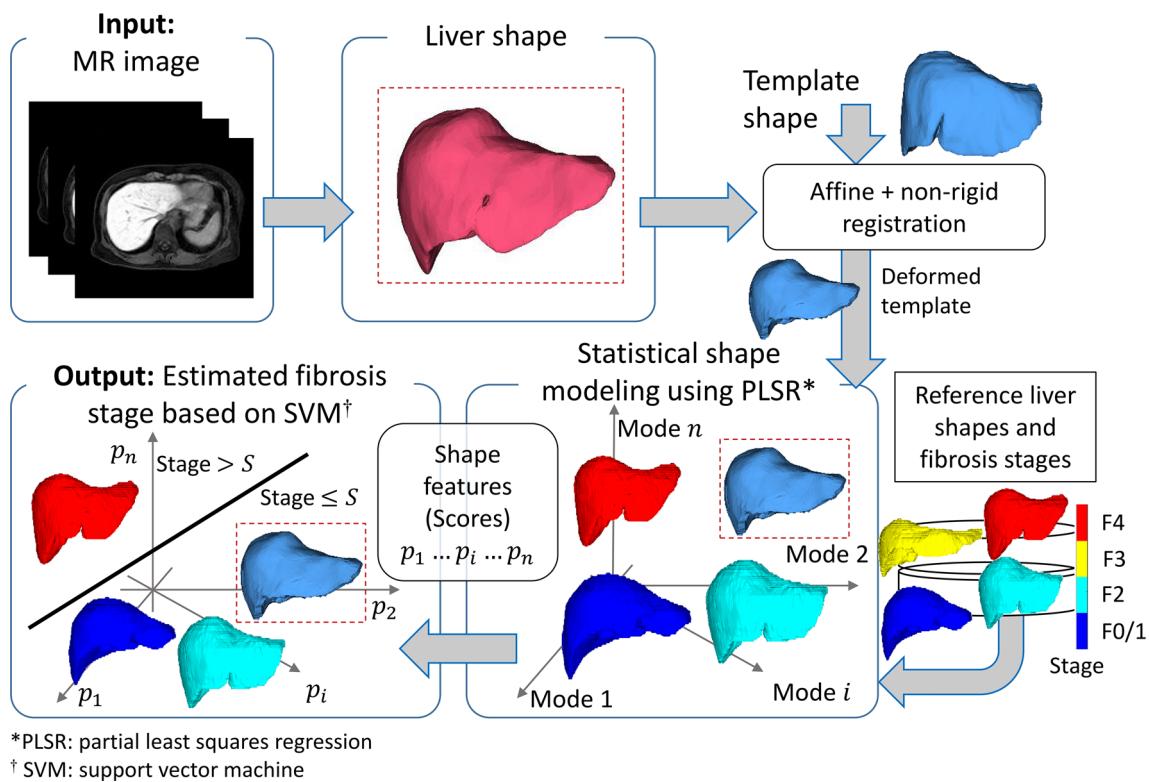
Methods

Overall scheme

Figure 1 shows the overall scheme for modelling the shape variations associated with liver fibrosis stage and predicting it. First, the liver shape was segmented from the 3D MR images. A smoothed polygon model was reconstructed from the segmented liver shape. A template shape was defined by arbitrarily selecting one of the shapes in the dataset and was registered to each target shape in the dataset using affine and non-rigid registrations. An SSM was constructed from the deformed shapes based on PLSR, thereby yielding the scores representing the shape variations. The scores were used to predict the liver fibrosis stage by using support vector machine (SVM).

Liver segmentation

In this study, the performance of the proposed approach was tested on liver shapes segmented based on either manual or automatic segmentations. The manual segmentation was performed by tracing the liver in the images using an image workstation (Vincent, Fujifilm, Tokyo, Japan). For the automatic segmentation, a CNN (3D U-Net [30]) was used.



*PLSR: partial least squares regression

† SVM: support vector machine

Fig. 1 Overall scheme for modelling the shape variations associated with liver fibrosis stage and predicting it from MR images by using PLSR-based shape features (scores)

Construction of statistical shape model based on PLSR

The SSM was constructed by using PLSR. The concept is to compute the scores representing the shape variations by iteratively maximizing the covariance between liver shape vectors and the corresponding fibrosis stage, as a supervised information, based on linear regression [24, 25]. In order to obtain the shape vectors, deformed template shapes with a point-to-point correspondence were constructed as follows [31]. The liver shapes were transformed into smooth surface polygon models. The number of vertices was reduced to 4000 vertices following Hori et al. [17]. One shape was arbitrarily selected as a template shape. The shapes in the dataset (target shapes) were normalized based on the bounding box of an arbitrarily selected shape [32]. The deformed shapes were obtained by registering the template to the target shapes via affine and multi-resolution non-rigid registrations [31] using the Image Registration Toolkit (IRTK) library [33, 34]. In order to reduce the bias induced by the selected template, the point-to-point correspondence was established by using a bootstrapping approach [35] repeated for four iterations.

Next, PLSR was performed using the SIMPLS algorithm [36], by which a score vector for each deformed shape was obtained. The score vector was standardized, and the scores

corresponding with the modes summarizing 99.5% of cumulative variation in the training dataset were selected.

Prediction of fibrosis stage

The fibrosis stage was predicted from the score vector by using a linear SVM classifier following previous studies [17, 18]. The SVM was trained and validated through a five-fold cross-validation. In order to avoid the bias in the SVM caused by the random selection in the cross-validation, the optimal SVM hyper-parameters were obtained using a nested leave-one-patient-out cross-validation applied to the training dataset at each fold [37]. The SVM model that achieved the largest AUC on the training dataset in each fold was selected.

Results

Materials

Institutional review board approvals for this retrospective study were obtained from Ikeda Municipal Hospital (approval ID: 3253) and Osaka University Hospital (approval ID: 14251). The requirement for informed consent was waived. Consecutive 74 patients, from whom a liver spec-

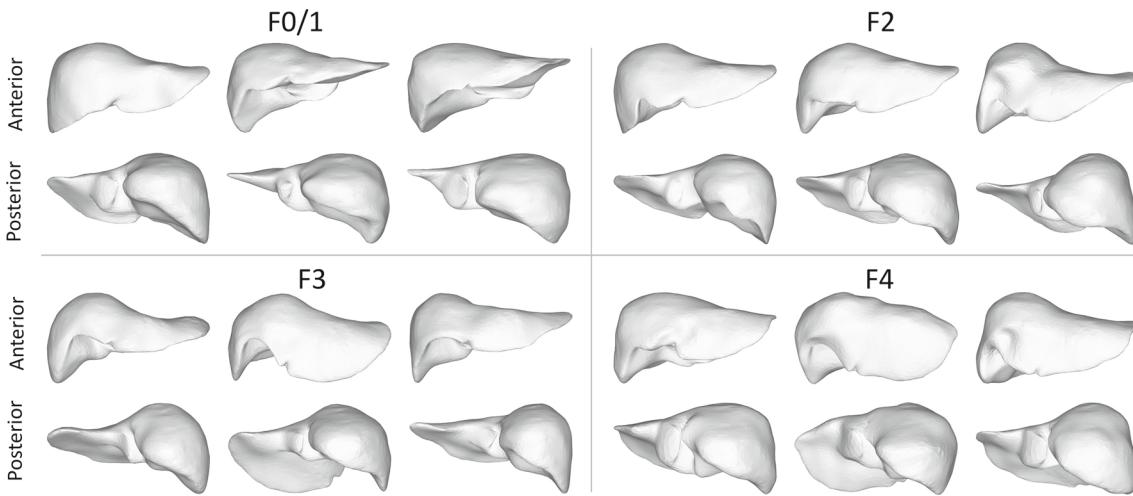


Fig. 2 Representative cases for the shape variations in the liver at different fibrosis stages

imen was acquired by either a percutaneous liver biopsy or a surgical resection, and who underwent an MR elastography at Ikeda Municipal Hospital were initially evaluated. Exclusion criteria were (a) no contrast-enhanced MR imaging ($n = 7$) and (b) focal liver lesions >2 cm or a treatment history for liver tumors ($n = 16$), resulting in a dataset of 51 patients (male: 24, female: 27; age: 37–88 year). Contrast-enhanced MR imaging and MR elastography were undergone at the same time. The dataset included 50 patients who had chronic liver diseases such as alcoholic liver disease ($n = 7$), autoimmune hepatitis ($n = 1$), hepatitis B virus infection (HBV; $n = 1$), hepatitis C virus infection (HCV; $n = 40$, HCV and alcoholic fatty liver disease $n = 1$) and non-alcoholic fatty liver disease ($n = 2$). The remaining one patient did not have a chronic liver disease. Liver tumors were identified in six patients (liver metastasis $n = 1$; hepatocellular carcinoma $n = 5$). The liver fibrosis stage was pathologically determined by the percutaneous liver biopsy ($n = 48$) or the surgical resection ($n = 3$). New Inuyama classification was used for evaluating the fibrosis stage [3]. The distribution of the fibrosis stage in the dataset was: F0/1 ($n = 18$), F2 ($n = 15$), F3 ($n = 7$) and F4 ($n = 11$). Figure 2 shows representative cases for the shape variations in the liver observed at the different fibrosis stages.

Gadolinium ethoxybenzyl diethylenetriamine pentaacetic acid-enhanced (Gd-EOB-DTPA; EOB-Primovist; Bayer Yakuhin, Ltd, Osaka, Japan) fat-suppressed T1-weighted MR images were acquired by a 1.5 Tesla scanner (Signa HDxt 1.5T, GE Medical Systems, Milwaukee, WI) with an 8-ch body coil using a Liver Acquisition with Volume Acquisition (LAVA) sequence. The images were acquired within a field-of-view of 380×380 mm 2 , with repetition time of 4.3 ms, echo time of 2.0 ms, flip angle of 12°, an in-plane matrix size of 320×192 pixels, and slice thickness/interval of 5/2.5 mm.

The performance of the SSM-based scores was compared to the measurements of serum/blood tests and MRE images. The serum/blood measurements were based on the Fibrosis-4 index (FIB-4) fibrosis marker [38]. The MRE images were acquired by using a 2D gradient-echo MRE sequence with a repetition time of 50 ms, echo time of 22 ms, flip angle of 30°, field-of-view of 400×240 mm 2 , matrix size of 256×64 pixels, slice thickness of 8 mm, acoustic vibration of 60 Hz and a motion encoding gradient (MEG) of 60 Hz.

Evaluation metrics

The automated segmentation accuracy was evaluated by using Dice coefficient ($DC \in [0, 1]$) [39]. A larger DC indicates a larger overlap between the ground-truth and the segmented livers. In order to visualize the shape variations associated with the fibrosis stage, the signed surface distance [40] was computed between each point in the reference shape (Shape 1) and its corresponding point at the deformed shape. This yielded positive distances at enlargement regions, and negative distances at shrinkage regions.

We noticed that the caudate lobe was translated due to its relatively small volume under the overall enlargement of the left lobe and shrinkage of the right lobe. Therefore, the shape variations at the caudate lobe were analyzed using point-wise displacement vectors obtained via iterative closest point (ICP) registration between Shape 3 (see Fig. 3) and Shape 1 using the IGT Module in 3D Slicer Ver.4.4.0 [41].

The fibrosis staging was assessed based on the following binary classifications for diagnosing different grades of the disease: F0/1 versus F2–4 (clinically significant fibrosis), F0–2 versus F3–4 (advanced fibrosis), and F0–3 versus F4 (cirrhosis), following Refs. [9–11, 17]. The performance of the classifiers was evaluated by using the AUC, the accuracy

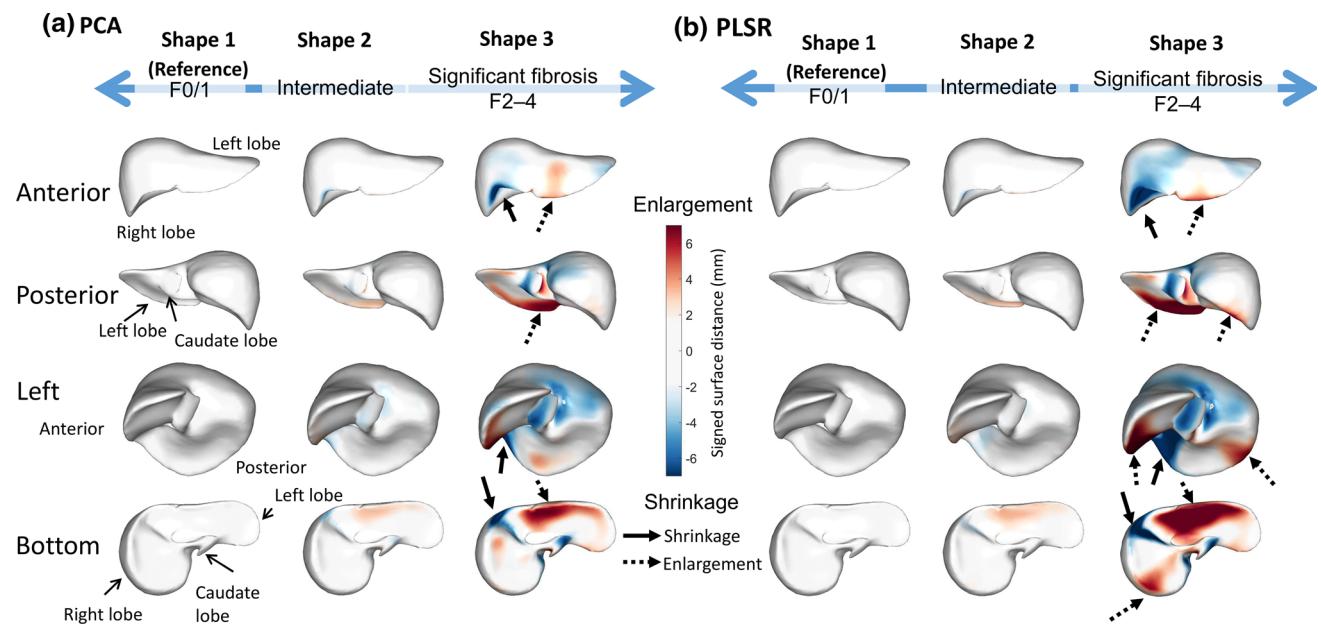


Fig. 3 Liver shape variations derived in the direction from F0/1 to F2–4 in the classification F0/1 versus F2–4 using PCA and PLSR scores. Shape 1 is the mean shape of the F0/1 shapes. Shape 2 corresponds with the scores of F0/1 projected onto the discrimination hyperplane,

and Shape 3 corresponds with the scores of Mean F2–4 projected onto the line passing through Shape 1 and Shape 2. Animated version of the variations can be found at the Supplementary Materials (ModesProjections.mp4)

(ACC) and the F1-score (F1). In order to further reduce the influence of the random selection of training/testing datasets in deriving the SVM models, the classification by SVM was repeated for 50 iterations with a fivefold cross-validation at each iteration. The statistical significance between the medians of the AUCs of the 50 iterations between PCA versus PLSR and manual versus automated segmentations was assessed by using a two-sided Wilcoxon rank sum test with $\alpha = 0.05$. To account for the multiple comparisons, the p values were adjusted with Bonferroni correction ($n = 6$) using R 3.4.3 (R Foundation for Statistical Computing, Vienna, Austria) [42]. Pearson ($\rho_P \in [-1, 1]$) and Spearman ($\rho_S \in [-1, 1]$) correlation coefficients were used to assess the linearity and monotonicity of the relationship, respectively. In order to assess the effect size of the differences between the mean AUCs, Cohen's d [43] was computed using R 3.4.3 [42]. d -values of 0.2, 0.5 and 0.8 were the cut-offs for small, medium and large effects, respectively [43].

Experimental results

Figure 3 shows the shape variations represented by the SSMs. The shape variations were computed in the direction perpendicular to the decision hyperplane at the classification F0/1 versus F2–4. Shape 1 corresponds with the scores of the mean of F0/1 shapes, whereas Shape 2 was calculated by projecting the score vector of Shape 1 onto the SVM hyperplane, and Shape 3 corresponds with the projection of the mean shape

F2–4 onto the line passing through Shape 1 and Shape 2. As the direction is perpendicular to the discrimination hyperplane, the shape variations would have the least correlation between the F0/1 and F2–4 shapes. Enlargements at the left and caudate lobes and posterior part of the right lobe with shrinkage in the anterior part of the right lobe were observed in the F2–4 direction. In order to better visualize the variations in the caudate lobe, Shape 3 was rigidly registered to Shape 1 based on ICP using 160 surface points sampled at the groove of the inferior vena cava and Spiegel's lobe. Figure 4 shows a magnified view for the variations in the caudate lobe represented by the displacement vectors after the rigid registration. Following the anatomical description of the sub-segments of the caudate lobe by Abdalla et al. [44], the enlargements were mainly observed at the caudate process, paracaval portion and papillary process.

At the automatic segmentation using 3D U-Net, the training hyper-parameters were empirically set as follows: the number of epochs was set to 32 epochs, the batch size to 8 images and the learning rate to 0.001, and the weights were optimized using an Adam optimizer based on a binary cross-entropy loss function. The ground-truth labels in the training of the 3D U-Net were the same as those obtained by the manual segmentation. On average, the DCs obtained using the automatic segmentation were $94.4 \pm 3.25\%$.

The PLSR scores obtained using the manual segmentations produced larger AUCs, ACCs and F1-scores than PCA. At F0/1 versus F2–4 and F0–2 versus F3–4, the

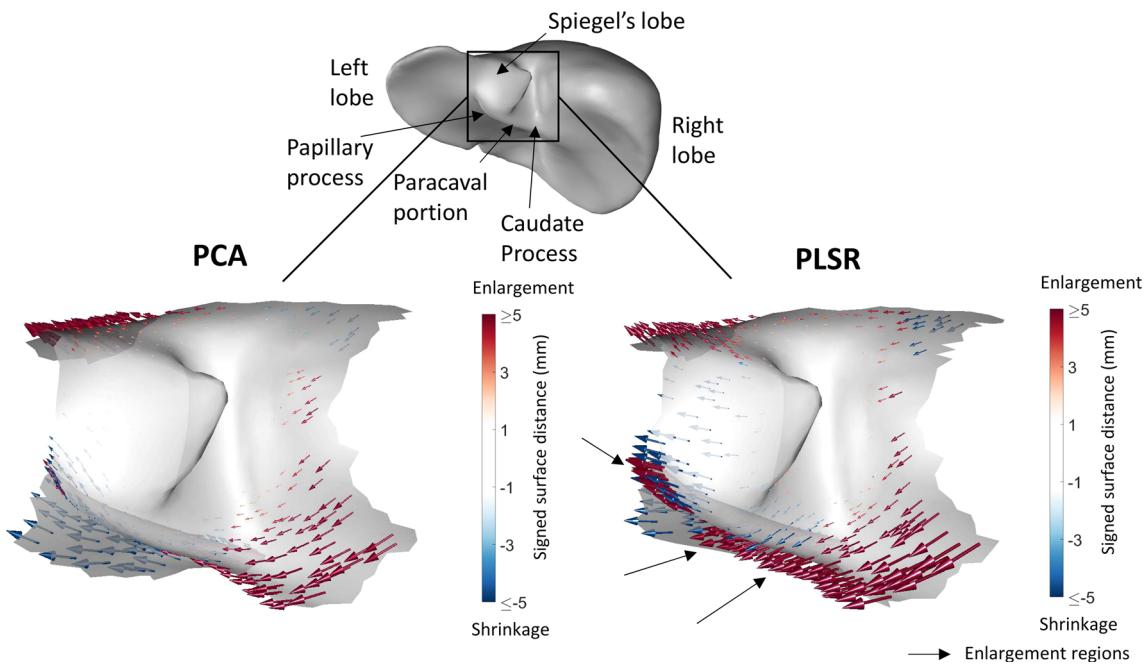


Fig. 4 Representation of the enlargement in the caudate lobe in Shape 3 (F2–4) compared with Shape 1 (F0/1) using PCA and PLSR. The deformations were estimated using an iterative closest point (ICP) registration to account for the translation under the overall enlargement of the left lobe and shrinkage of the right lobe. Blue and red arrows indicate the

deformation from Shape 1 to Shape 3, whereas the color represents the signed surface distance (blue: shrinkage; red: enlargement). Enlargements by PLSR were mainly observed at the caudate process, paracaval portion and papillary process, as indicated by the black arrows. PCA represented smaller enlargements at the caudate process

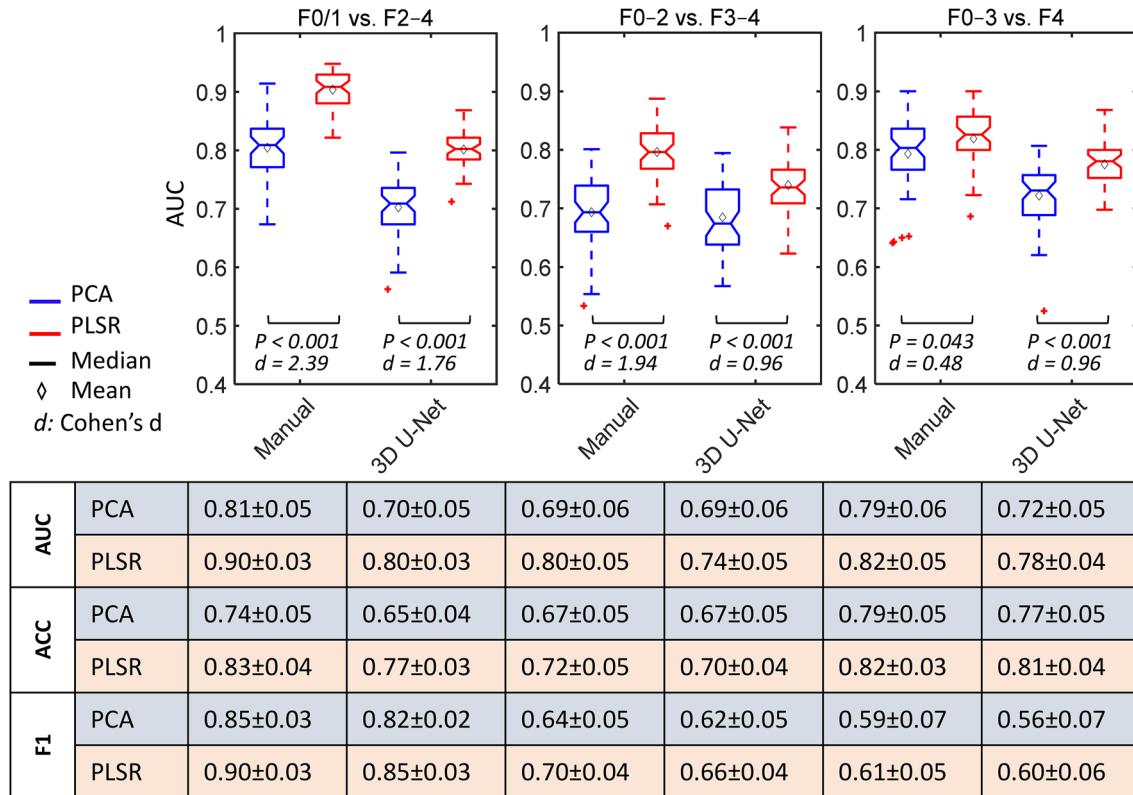


Fig. 5 Accuracy of liver fibrosis staging represented by the area under the receiver operating characteristic curve (AUC), accuracy (ACC) and F1-score (F1) using PCA (blue) and PLSR (red) applied to manual and

automatic (3D U-Net) segmentations. (P : adjusted p value, Wilcoxon rank sum test, Bonferroni correction $n = 6$; d : Cohen's d)

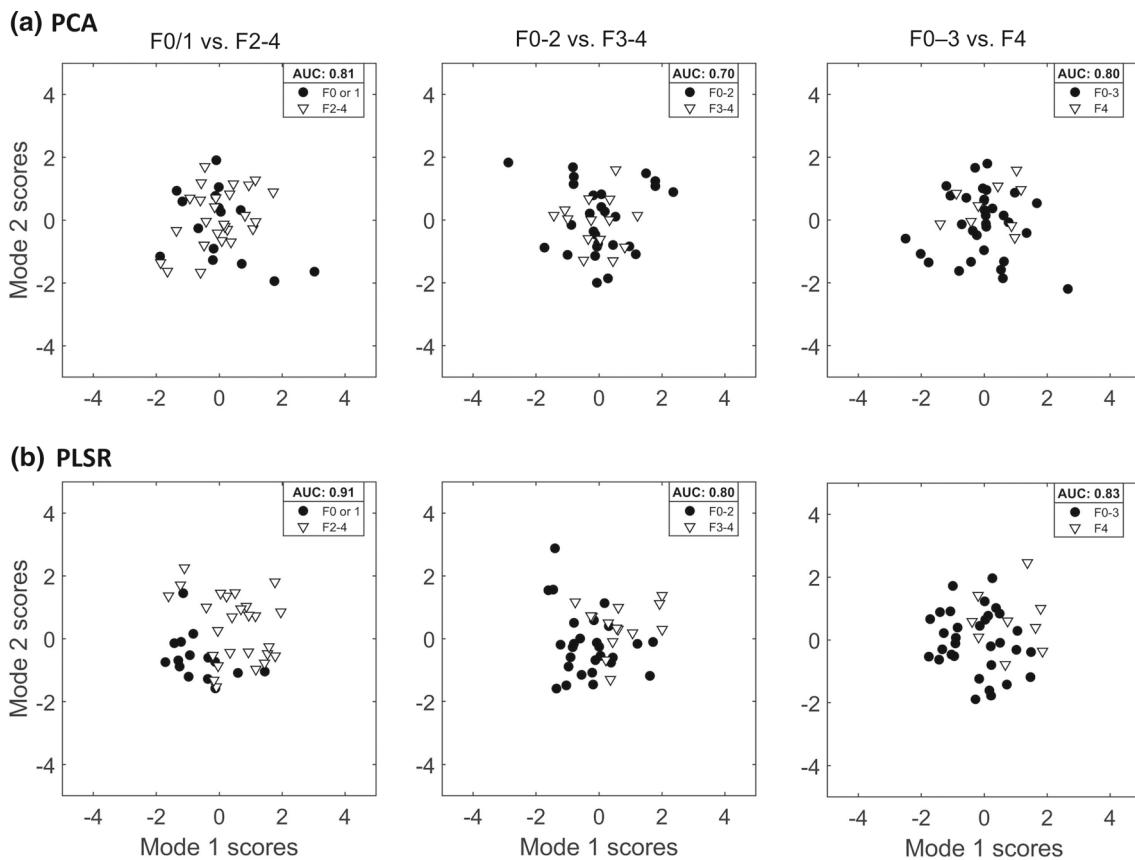


Fig. 6 Distributions of the scores corresponding with the first and second modes obtained by using PCA (a) and PLSR (b) with respect to the fibrosis stages. The black circles and white triangles indicate the lower and higher fibrosis classes, respectively

AUC increased from 0.81 ± 0.05 to 0.90 ± 0.03 ($p < 0.001$; $d = 2.39$; 95%CI [1.78, 3.00]) and from 0.69 ± 0.06 to 0.80 ± 0.05 ($p < 0.001$; $d = 1.94$; 95%CI [1.39, 2.48]). The largest mean AUC, ACC and F1-score were 0.90 ± 0.03 , 0.83 ± 0.04 and 0.90 ± 0.03 , respectively, obtained in the classification F0/1 versus F2-4 (significant fibrosis detection). Low F1-scores were obtained for F0-2 versus F3-4 and F0-3 versus F4, which indicate low sensitivity (detection of true positives) in these classifications. The largest AUC at the automated segmentation was 0.80 ± 0.03 (F0/1 vs. F2-4).

Figure 6 shows a comparison between PCA and PLSR scores of the first and second modes. The PLSR scores clearly show, especially in F0/1 versus F2-4, a better class separation in the latent space compared to PCA.

Figure 7 shows the correlations between the fibrosis stage and the scores at the first mode derived from PCA and PLSR. From Fig. 7a, the Pearson correlation coefficient shows that the correlation is not statistically significant. From Fig. 7b, the Pearson correlation coefficient at F0/1 versus F2-4 ($\rho_P = 0.461$, $p = 0.004$) showed that PLSR scores have a moderate correlation with the fibrosis stage.

Figure 8 shows the ROC curves corresponding with the median AUCs (upper middlemost values) obtained from the

manual segmentations compared with those of the MRE and serum/blood measurements. MRE had the largest AUC (0.965) obtained in F0-2 versus F3-4. Serum/blood measurements outperformed the SSMs at F0-1 versus F2-4 and F0-2 versus F3-4.

Discussion

Effectiveness of the proposed method

In this study, an application of SSM using PLSR for understanding and staging of liver fibrosis based on MR images was proposed. In addition to commonly known variations associated with liver fibrosis, such as the shrinkage of the entire right lobe or enlargement of the entire left lobe, even though a small dataset was used, our SSM based on PLSR represented locally detailed variations observed in HCV-related cirrhotic livers [15]. These variations include shrinkage in the anterior part of the right lobe, and enlargement of the caudate lobe and posterior part of the right lobe, as shown in Figs. 3 and 4. In addition, the SSM could capture enlargement at the caudate and papillary processes of

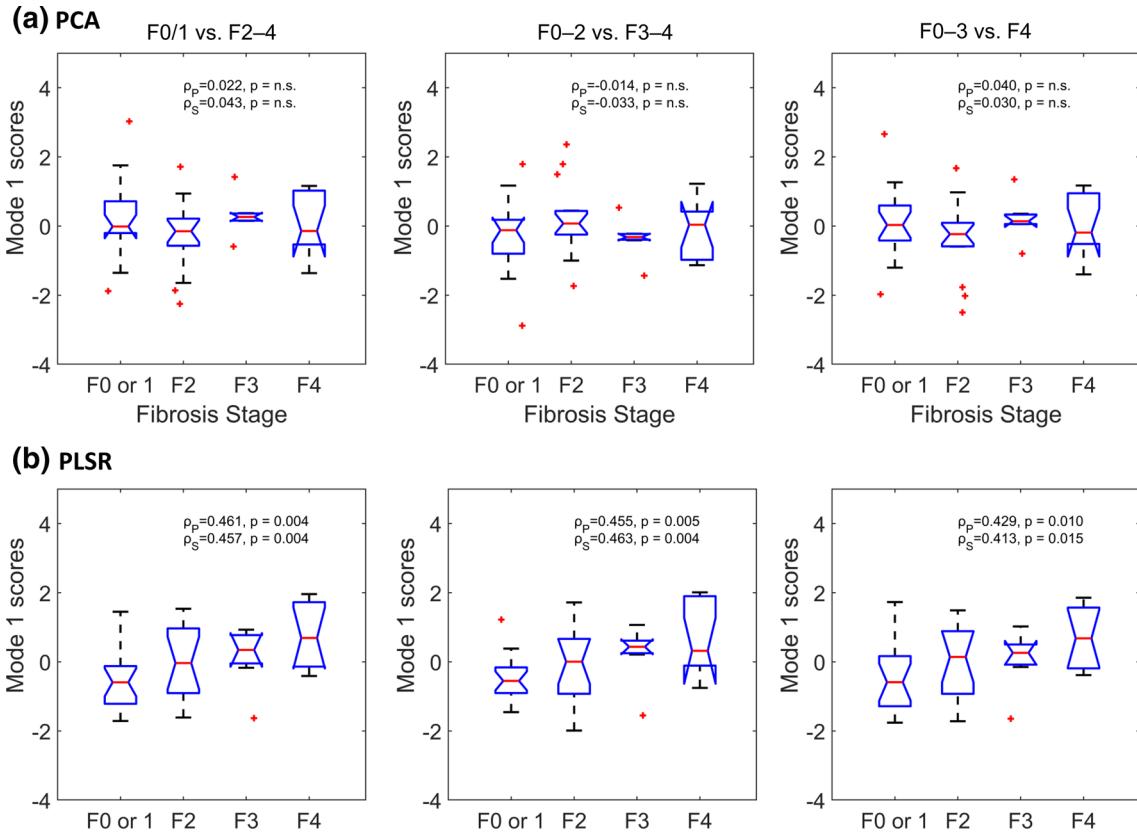


Fig. 7 Correlations between the fibrosis stage (*x*-axis) and SSM scores (*y*-axis) obtained by PCA (a) and PLSR (b). ρ_P : Pearson correlation coefficient; ρ_S : Spearman correlation coefficient; *p*: adjusted *p* value (Bonferroni correction, $n = 6$); n.s.: not statistically significant

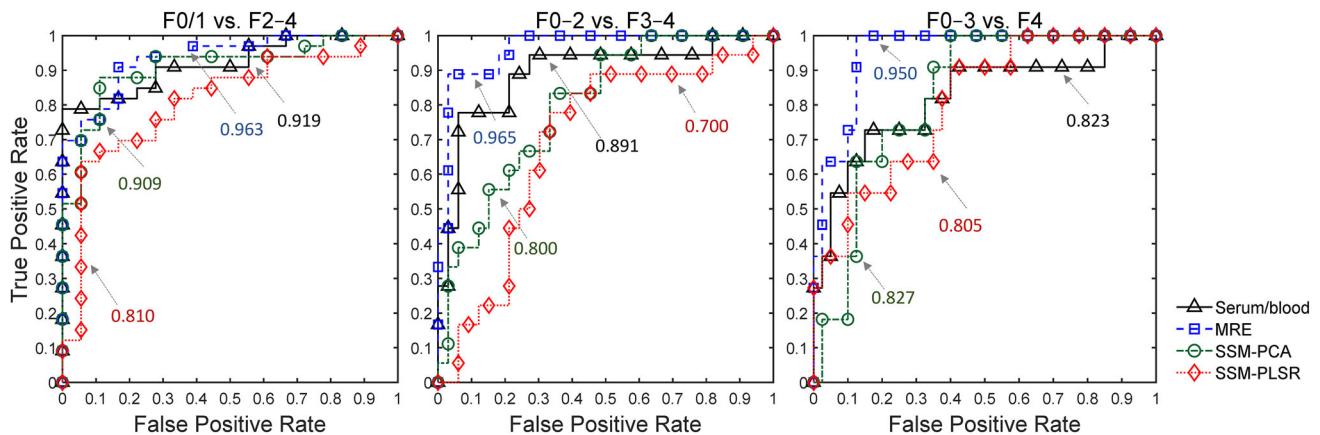


Fig. 8 ROC curves of the binary classifications using PCA and PLSR scores corresponding with median AUCs (upper middlemost values) under manual segmentation, with the ROCs of the MRE and serum/blood measurements. Annotations with gray arrows indicate the AUC

the caudate lobe, as shown in Fig. 4. These modelled variations qualitatively agree with localized variations (at the segment and sub-segment levels) reported in [15, 16]. It is also noteworthy that compared to PCA, PLSR could better represent the variations, thereby yielding better classification accuracy, as shown in Figs. 5, 6, 7, and 8. Therefore, we presume that the derived direction using PLSR scores could be potentially representing the direction of the fibrosis dis-

ease progression. However, even though not a few number of cases in our dataset had HCV ($n = 39$; 76.5%), an analysis at the segment level on a larger dataset is needed.

Comparison with existing methods

Table 1 shows comparisons between the proposed and latest approaches [9–11, 17]. Choi et al. [11] demonstrated the

Table 1 Comparison with previous studies using image features

Study	Technique	ROI	No. of cases	Distribution (%)			Modality	Segmentation F0/1 versus F2–4		F0–2 versus F3–4		F0–3 versus F4	
				F0/1	F2	F3	F4	AUC	ACC	AUC	ACC	AUC	ACC
Proposed	SSM (PLSR)	Whole liver	51	18 (35)	15 (29)	7 (14)	11 (22)	Gd-EOB MR	Manual	0.90±0.03	0.83±0.04	0.80±0.05	0.72±0.05
Hori et al. [17]	SSM (PCA)	Whole liver	91	61 (67)	3 (3)	1 (1)	26 (29)	CT	Automated	0.80±0.03	0.77±0.03	0.74±0.05	0.70±0.04
House et al. [9]	Texture (Haralick) features	Three transverse slices from the whole liver	49	27 (55)	5 (10)	9 (18)	8 (16)	T2-weighted MR	Manual	0.95	—	0.96	—
									Manual	0.81	—	0.81	—
									Automated	0.81	—	0.87	—
Yasaka et al. [10]	CNN	Patches from axial slices	634	127 (20)	96 (15)	133 (21)	278 (44)	Hepatobiliary phase MR	Manual	0.85	0.80	0.84	0.77
Choi et al. [11]	CNN	Whole liver	8352	3697 (44)	445 (5)	633 (8)	3577 (43)	Contrast-enhanced CT	Automated	0.96	0.94	0.97	0.95
									Manual	0.92±0.03	0.90±0.03	0.95	0.92

largest AUCs and ACCs using CNNs applied to CT images. Yasaka et al. [10] demonstrated a high performance for the classifications F0–2 versus F3–4 (AUC: 0.84) and F0–3 versus F4 (AUC: 0.84) using a CNN applied to patches extracted from axial MR images. Given the large datasets used in the studies [10, 11], it is evident that CNN-based approaches demonstrate a remarkably high performance with respect to fibrosis stage prediction. Nevertheless, CNNs have a limited interpretability with respect to the fibrosis-relevant variations in the classified images.

House et al. [9] also investigated the potential of Haralick texture analysis on whole liver slices sampled from T2-weighted MR images, which enhance the contrast between fibrotic and non-fibrotic livers. They could identify reduced sets of features that produced a high prediction accuracy, especially in F0–3 versus F4 (AUC: 0.87). However, the feature performance had high dependency on the acquisition parameters, such as the echo time, which affects the reproducibility under different acquisition parameters [9, 12]. Hori et al. [17] reported high prediction accuracy using PCA-based SSM features based on whole livers extracted from CT images. The essential difference in their SSM is using the liver shapes without normalization. By normalizing the shapes (see section “**Construction of statistical shape model based on PLSR**”), our analysis could represent locally detailed variations.

In summary, the proposed approach has two advantages compared with the previous studies [9–11]. First, even though a small dataset was used, it provided an explicit mathematical representation of the variations associated with fibrosis stage, which was not offered by the implicit representations from CNNs [10, 11] or the texture features [9]. The second is the modality-independence, where as long as the whole liver shape is provided, SSMs can be derived regardless of the imaging modality.

On the other side, the proposed approach has a disadvantage of the degraded performance when automatically segmented liver shapes were used. We analyzed the impact of the automated segmentation by investigating the shapes that had a notable decrease in their classification accuracy. For this, a shape-wise AUC (i.e., AUC of classifying the shape against those from the other fibrosis stages) was computed. At F0–1 versus F2–4, when excluding the shapes whose AUC decreased $\geq 25\%$ ($n = 5$), overall AUCs of 0.92 ± 0.03 and 0.90 ± 0.03 for the manual and automated segmentations were obtained, respectively. Compared with the difference between the AUCs of the manual (0.90 ± 0.03) and automated (0.80 ± 0.03) segmentations before the exclusion (as shown in Fig. 5), the difference after the exclusion has notably decreased. This indicates a large impact of the 5 shapes on the decreased classification accuracy. However, by analyzing the segmentation accuracy of those shapes, one had a low DC of 74.4%, while the remaining 4 shapes had an

average DC of $93.9 \pm 2.3\%$, which is relatively close to that of the non-excluded shapes ($95.1 \pm 1.8\%$). Thus, although these 4 shapes did not have a large segmentation failure, their local segmentation failures negatively affected the overall classification accuracy. In our future work, this impact will be further investigated.

Limitations and future work

This study has the following limitations. First, a relatively small database ($n = 51$) was used. Increasing the number of cases would help in further clarifying the shape variations at the intermediate and advanced fibrosis stages and improving their classification accuracies. Second, the MRE measurements still have a better performance, as shown in Fig. 8, which requires further investigations about correlations between the liver shape and MRE measurements to improve the staging accuracy.

Our future work will focus on the following directions. First, a larger database of MR images of approximately 250 cases is being constructed. For such a database, instead of constructing the SSM collectively from all stages, it will be possible to obtain SSMs for each stage independently while interpolating among the neighboring models to represent the variations within intermediate stages, as Kasahara et al. [45] proposed for fetal development modelling. Second, in order to investigate improvements at the classification accuracy, the PLSR scores will be combined with other image features representing tissue characteristics, such as texture features [9] or learned representations via CNNs [10, 11], as well as physiological parameters, such as the serum/blood parameters.

Conclusions

This study offered an explicit representation of commonly known as well as detailed shape variations associated with liver fibrosis in contrast-enhanced MR images. Therefore, the application of PLSR-based shape analysis is feasible for improving our understanding of the shape variations associated with fibrosis stage and predicting it.

Acknowledgements The authors are grateful for Fukuda Norio, Yuki Suzuki, Steven Lim, Yukio Oshiro and Toshiyuki Okada for their contributions to this study. This research was supported by Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) Number 26108004/26461789/19K20711 and 17K10403.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent The requirement for informed consent was waived for this study.

References

- Ellis EL, Mann DA (2012) Clinical evidence of the regression of liver fibrosis. *J Hepatol* 56(5):1171–1180. <https://doi.org/10.1016/j.jhep.2011.09.024>
- Tsochatzis EA, Bosch J, Burroughs AK (2014) Liver cirrhosis. *Lancet* 383(9930):1749–1761. [https://doi.org/10.1016/S0140-6736\(14\)60121-5](https://doi.org/10.1016/S0140-6736(14)60121-5)
- Ichida F, Tsuji T, Omata M, Ichida T, Inoue K, Kamimura T, Yamada G, Hino K, Yokosuka O, Suzuki H (1996) New Inuyama classification; new criteria for histological assessment of chronic hepatitis. *Int Hepatol Commun* 6(2):112–119. [https://doi.org/10.1016/S0928-4346\(96\)00325-8](https://doi.org/10.1016/S0928-4346(96)00325-8)
- Liaw YF (2013) Reversal of cirrhosis: an achievable goal of hepatitis B antiviral therapy. *J Hepatol* 59(4):880–881. <https://doi.org/10.1016/j.jhep.2013.05.007>
- Cadranel JF, Rufat P, Degos F (2000) Practices of liver biopsy in France: results of a prospective nationwide survey. For the Group of Epidemiology of the French Association for the Study of the Liver (AEEF). *Hepatology* 32(3):477–481. <https://doi.org/10.1053/jhep.2000.16602>
- Rockey DC, Bissell DM (2006) Noninvasive measures of liver fibrosis. *Hepatology* 43(Suppl 1):S113–S120. <https://doi.org/10.1002/hep.21046>
- Wang QB, Zhu H, Liu HL, Zhang B (2012) Performance of magnetic resonance elastography and diffusion-weighted imaging for the staging of hepatic fibrosis: a meta-analysis. *Hepatology* 56(1):239–247. <https://doi.org/10.1002/hep.25610>
- Tang A, Cloutier G, Szeverenyi NM, Sirlin CB (2015) Ultrasound elastography and MR elastography for assessing liver fibrosis: part 2, diagnostic performance, confounders, and future directions. *Am J Roentgenol* 205:33–40. <https://doi.org/10.2214/AJR.15.14553>
- House MJ, Bangma SJ, Thomas M, Gan EK, Ayonrinde OT, Adams LA, Olynyk JK, St Pierre TG (2015) Texture-based classification of liver fibrosis using MRI. *J Magn Reson Imaging* 41(2):322–328. <https://doi.org/10.1002/jmri.24536>
- Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S (2018) Liver fibrosis: deep convolutional neural network for staging by using gadoteric acid-enhanced hepatobiliary phase MR images. *Radiology* 287(1):146–155. <https://doi.org/10.1148/radiol.2017171928>
- Choi KJ, Jang JK, Lee SS, Sung YS, Shim WH, Kim HS, Yun J, Choi JY, Lee Y, Kang BK, Kim JH, Kim SY, Yu ES (2018) Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT images in the liver. *Radiology*. <https://doi.org/10.1148/radiol.2018180763>
- Shafiq-Ul-Hassan M, Zhang GG, Latif K, Ullah G, Hunt DC, Balagurunathan Y, Abdallah MA, Schabath MB, Goldgof DG, Mackin D, Court LE, Gillies RJ, Moros EG (2017) Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys* 44(3):1050–1062. <https://doi.org/10.1002/mp.12123>
- Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JAWM, van Ginneken B, Cl Sánchez (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88. <https://doi.org/10.1016/j.media.2017.07.005>
- Kudo M, Zheng RQ, Kim SR, Okabe Y, Osaki Y, Iijima H, Itani T, Kasugai H, Kanematsu M, Ito K, Usuki N, Shimamatsu K, Kage

- M, Kojiro M (2008) Diagnostic accuracy of imaging for liver cirrhosis compared to histologically proven liver cirrhosis. *Intervirol* 51(Suppl 1):17–26. <https://doi.org/10.1159/000122595>
15. Ozaki K, Matsui O, Kobayashi S, Sanada J, Koda W, Minami T, Kawai K, Gabata T (2010) Selective atrophy of the middle hepatic venous drainage area in hepatitis C-related cirrhotic liver: morphometric study by using multidetector CT. *Radiology* 257(3):705–714. <https://doi.org/10.1148/radiol.10100468>
 16. Hunt FOM, Lubner MG, Ziemlewicz TJ, Muñoz Del Rio A, Pickhardt PJ (2016) The liver segmental volume ratio for noninvasive detection of cirrhosis: comparison with established linear and volumetric measures. *J Comput Assist Tomogr* 40(3):478–484. <https://doi.org/10.1097/RCT.0000000000000389>
 17. Hori M, Okada T, Higashiiura K, Sato Y, Chen YW, Kim T, Onishi H, Eguchi H, Nagano H, Umeshta K, Wakasa K, Tomiyama N (2015) Quantitative imaging: quantification of liver shape on CT using the statistical shape model to evaluate hepatic fibrosis. *Acad Radiol* 22(3):303–309. <https://doi.org/10.1016/j.acra.2014.10.001>
 18. Chen YW, Luo J, Dong C, Han X, Tateyama T, Furukawa A, Kanasaki S (2013) Computer-aided diagnosis and quantification of cirrhotic livers based on morphological analysis and machine learning. *Comput Math Methods Med* 2013:264809. <https://doi.org/10.1155/2013/264809>
 19. Foruzan AH, Chen YW, Hori M, Sato Y, Tomiyama N (2014) Capturing large shape variations of liver using population-based statistical shape models. *Int J Comput Assist Radiol Surg* 9(6):967–977. <https://doi.org/10.1007/s11548-014-1000-5>
 20. Mwangi B, Tian TS, Soares JC (2014) A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12:229–244. <https://doi.org/10.1007/s12021-013-9204-3>
 21. Maitra S, Yan Y (2008) Principle component analysis and partial least squares: two dimension reduction techniques for regression: applying multivariate statistical models, vol 79. Casualty Actuarial Society, Quebec City
 22. Tibshirani R (2011) Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Ser B (Statistical Methodology)* 73:273–282. <https://doi.org/10.1111/j.1467-9868.2011.00771.x>
 23. Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Statistical Methodology)* 67(2):301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
 24. Geladi P, Kowalski B (1986) Partial least squares regression: a tutorial. *Anal Chim Acta* 185:1–17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9)
 25. Wold S, Ruhe A, Wold H, Dunn WJ III (1984) The collinearity problem in linear regression: the partial least squares (PLS) approach to generalized inverses. *SIAM J Sci Stat Comput* 5(3):735–743. <https://doi.org/10.1137/0905052>
 26. Lekadir K, Albi X, Pereañez M, Frangi AF (2015) Statistical shape modeling using partial least squares: application to the assessment of myocardial infarction. In: Proceedings: revised selected papers of the 6th international workshop on statistical atlases and computational models of the heart. Imaging and modelling challenges, vol 9534, pp 130–139. https://doi.org/10.1007/978-3-319-28712-6_14
 27. Lekadir K, Hoogendoorn C, Pereañez M, Albà X, Pashaei A, Frangi AF (2014) Statistical personalization of ventricular fiber orientation using shape predictors. *IEEE Trans Med Imaging* 33(4):882–889. <https://doi.org/10.1109/TMI.2013.2297333>
 28. Martinez AM, Kak AK (2001) PCA versus LDA. *IEEE Trans Pattern Anal Mach Intell* 23(2):228–233. <https://doi.org/10.1109/34.908974>
 29. Belhumeur P, Hespanha J, Kriegman D (1997) Eigenfaces versus fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720. <https://doi.org/10.1109/34.598228>
 30. Çiçek Ö, Abdulkadir A, Lienkamp S, Brox T, Ronneberger O (2016) 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Proceedings of MICCAI, Springer, LNCS 9901, pp 424–432. https://doi.org/10.1007/978-3-319-46723-8_49
 31. Okada T, Shimada R, Hori M, Nakamoto M, Chen YW, Nakamura H, Sato Y (2008) Automated segmentation of the liver from 3D CT images using probabilistic atlas and multilevel statistical shape model. *Acad Radiol* 15(11):1390–1403. <https://doi.org/10.1016/j.acra.2008.07.008>
 32. Okada T, Linguraru MG, Hori M, Summers RM, Tomiyama N, Sato Y (2015) Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors. *Med Image Anal* 26(1):1–18. <https://doi.org/10.1016/j.media.2015.06.009>
 33. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ (1999) Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 18(8):712–721. <https://doi.org/10.1109/42.796284>
 34. IRTK <https://biomedia.doc.ic.ac.uk/software/irtk/>. Last accessed 28 Sept 2018
 35. Heimann T, Meinzer HP (2009) Statistical shape models for 3D medical image segmentation: a review. *Med Image Anal* 13(4):543–563. <https://doi.org/10.1016/j.media.2009.05.004>
 36. de Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. *Chemomet Intell Lab Sys* 18(3):251–263. [https://doi.org/10.1016/0169-7439\(93\)85002-X](https://doi.org/10.1016/0169-7439(93)85002-X)
 37. Cawley G, Talbot N (2010) On over-fitting in model selection and subsequent selection bias in performance evaluation. *J Mach Learn Res* 11:2079–2107
 38. Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, Sulkowski M, Torriani FJ, Dieterich DT, Thomas DL, Messinger D, Nelson M, APRICOT Clinical Investigators (2006) Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 43(6):1317–1325. <https://doi.org/10.1002/hep.21178>
 39. Dice LR (1945) Measures of the amount of ecological association between species. *Ecology* 26(3):297–302
 40. Aspert N, Santa-Cruz D, Ebrahimi T (2002) MESH: measuring errors between surfaces using the Hausdorff distance. In: Proceedings: IEEE international conference on multimedia and expo, Switzerland, Lusanne, pp 705–708. <https://doi.org/10.1109/ICME.2002.1035879>
 41. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin J-C, Pujol S, Bauer C, Jennings D, Fennessy FM, Sonka M, Buatti J, Aylward SR, Miller JV, Pieper S, Kikinis R (2012) 3D slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging* 30(9):1323–1341. <https://doi.org/10.1016/j.mri.2012.05.001>
 42. R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
 43. Cohen J (1988) Statistical power analysis for the behavioral sciences. Routledge Academic, New York. <https://doi.org/10.1016/C2013-0-10517-X>
 44. Abdalla EK, Vauthey JN, Couinaud C (2002) The caudate lobe of the liver: implications of embryology and anatomy for surgery. *Surg Oncol Clin N Am* 11(4):835–848. [https://doi.org/10.1016/S1055-3207\(02\)00035-2](https://doi.org/10.1016/S1055-3207(02)00035-2)
 45. Kasahara K, Saito A, Takakuwa T, Yamada S, Matsuzoe H, Honnati H, Shimizu A (2018) A spatiotemporal statistical shape model of the brain surface during human embryonic development. *Adv Biomed Eng* 7:146–155