A study of hepatic fibrosis staging methods using diffraction enhanced imaging

Check for
updates

Jing Wang[1*†], Hui Li[2†], Xiuling Zhou[3], Xiao-Zhi Gao[4] and Ming Wang[1]

* Correspondence: wang_jing@
bjmu.edu.cn
†Hui Li and Jing Wang contributed
equally to this work.
[1]Department of Health Information
Management, Peking University
Health Science Center, Beijing
100191, China
Full list of author information is
available at the end of the article

Abstract

The early hepatic fibrosis staging is very important for timely diagnosis, prognosis, and treatment of all chronic liver diseases. Diffraction-enhanced imaging, which can provide much more information on soft tissue morphology than conventional absorption radiography, might be a potential noninvasive technique to diagnose and stage hepatic fibrosis. This paper presents different feature extraction strategies and classification methods to automatically classify hepatic fibrosis using diffraction-enhanced imaging images. Texture features are obtained using a total of three methods including first order feature, gray level co-occurrence matrix, and grayscale gradient co-occurrence matrix. The fusion of these texture features is also studied. The principal component analysis is used to reduce the dimension of the features and redundant information among data. The features are classified using two popular classification techniques, namely, K-nearest neighbors and support vector machines. On the basis of the comparison of different feature strategies and classification methods, we can identify the suitable methods for grading hepatic fibrosis. The proposed approach efficiently classifies the hepatic fibrosis DEI images into four classes with the highest classification accuracy of 99.99%. We further demonstrate the potential of the DEI images in staging hepatic fibrosis.

Keywords: Hepatic fibrosis staging, Diffraction-enhanced imaging, Feature fusion, Machine learning

1 Introduction

Hepatic fibrosis is generally a pathophysiological process referring to the abnormal proliferation of connective tissues in the liver. There are various pathogenic factors of fibrosis, such as alcohol, viruses, and drugs. In the early stage, hepatic fibrosis is reversible under timely treatment. However, if the appropriate treatment opportunity is missing, fibrosis process will develop into cirrhosis, hepatic necrosis, and ultimately even liver cancer, which may seriously threaten the patient's health. Assessment of hepatic fibrosis is helpful to clinical decision-making, prognosis prediction, and curative effect evaluation. Therefore, accurate diagnosis and evaluation of the degree of hepatic fibrosis is indeed very important in practice.

Currently, liver biopsy is the gold standard for diagnosis and staging of fibrosis. However, it is invasive with many contraindications, e.g., high risk, low repeatability,

and poor compliance [1]. Hemorrhage, infection, and biliary peritonitis may occur. As a matter of fact, there are 3% morbidity and 0.03% mortality rates in liver biopsy [2]. In addition, the sampling volume of liver biopsy is only about 1/50,000 of the liver parenchymal tissue, leading to inevitable sampling errors [3, 4]. Imaging techniques have been widely used to evaluate the progress of liver fibrosis during the past decade, because of their advantages of noninvasive and high repeatability [5]. Those conventional imaging modalities including but are not only limited to ultrasonography (US), computed tomography (CT), and magnetic resonance imaging (MRI) provide morphological features of liver to evaluate liver fibrosis and could differentiate between cirrhosis or severe fibrosis and normal liver. However, accurate staging of fibrosis or diagnosis of mild fibrosis was often not achievable [6, 7]. X-ray Diffraction-enhanced imaging (DEI) is a phase contrast technology [8, 9]. For biological soft tissues, the phase shift is more than 1000 times larger than the absorption term. Therefore, DEI can generate excellent contrast of biological soft tissues compared to the conventional absorption radiography. Li et al. have demonstrated that the DEI images show great visibility of fibrous liver architectures, and DEI may be a potential noninvasive technique to diagnose and stage hepatic fibrosis [10].

In clinical practice, the interpretation of stages of hepatic fibrosis from medical images largely depends on the experiences of radiologists. Machine learning methods may assist the radiologists to analyze the medical images, to ease their workload, and to improve the accuracy of the diagnosis results [11–21]. Lee et al. [11] assessed gradations of hepatic fibrosis using a deep convolutional neural network (DCNN) with US. The DCNN showed high accuracy and achieved better performance than that of radiologists in the diagnosis of cirrhosis. Choi et al. [13] developed a deep learning system (DLS) for staging liver fibrosis by using CT images in the liver. Results showed that the DLS had a staging accuracy of 79.4% for diagnosing significant fibrosis, advanced fibrosis, and cirrhosis. K. Yasaka al [14]. staged liver fibrosis by using a deep learning model based on CT images, with moderate performance. Kayaalti et al. [15] obtained a comprehensive feature set from the CT images. The SVM and KNN were used to classify those features. The method proposed in this paper was successfully applied to pairwise fibrosis staging. However, its classification results for multi-class fibrosis still need to be further improved. Yasaka et al. [16] investigated the performance of a DCNN model in the staging of liver fibrosis using gadoxetic acid-enhanced hepatobiliary phase magnetic resonance (MR) imaging. The DCNN model exhibited a high diagnostic performance in the staging of liver fibrosis. House et al. [18] extracted texture features from combined MRI of T2-weighted, high-resolution, spin echo sequence. The receiver operating characteristics curve was employed to assess the diagnostic performance of texture. However, the performance of texture in MRI was modest when classifying patients with mild and intermediate fibrosis stages. Zhang et al. [21] used texture features calculated from GLCM as the input SVM to differentiate fibrosis using MRI and CT. The experimental results showed that the MRI images provided a higher classification accuracy than the CT images. Li et al. [10] applied DEI method to the imaging of hepatic fibrosis in rat liver sections and showed that DEI image can discriminate the degree of fibrosis. As a follow-up work of [10, 22] applied BP network to automatically discriminate liver fibrosis using texture feature and achieved the accuracy rate of 95.1%.

As aforementioned, the current work on staging of the liver fibrosis using machine learning mainly focuses on ultrasound images [11, 12], CT images [13–15, 21], and MRI images [16–21]. Most imaging techniques available at present can accurately detect cirrhosis or severe fibrosis, but usually fail to detect stage of fibrosis due to the limited sensitivity and specificity. In [10, 22], the potential of DEI image in the classification of hepatic fibrosis was explored. In [10, 22], only the texture features that extracted from GLCM were used to describe the characteristics of liver fibers. The extracted texture features were only preliminarily analyzed by human vision [10] or BP [22]. This paper is the continuation of [22]. In this paper, more kinds of texture features, fusion features, and features after dimensionality reduction are developed to describe the hepatic fibers of DEI image, and more machine learning algorithms are applied to automatically classify the extracted features into different fibrosis stages for the automatic staging of hepatic fibrosis. Compared with [22], the performance of liver fiber grading has been significantly improved. We also identify the potentials of DEI in liver fibrosis classification, providing a reference for the selection of optimal features, classification methods, and parameters in automatic staging of liver fibrosis

Our paper is structured as follows. Section 2 introduces the research background and relevant methods. Section 3 demonstrates the experimental results and discussions. Finally, Section 4 draws some remarks and conclusions.

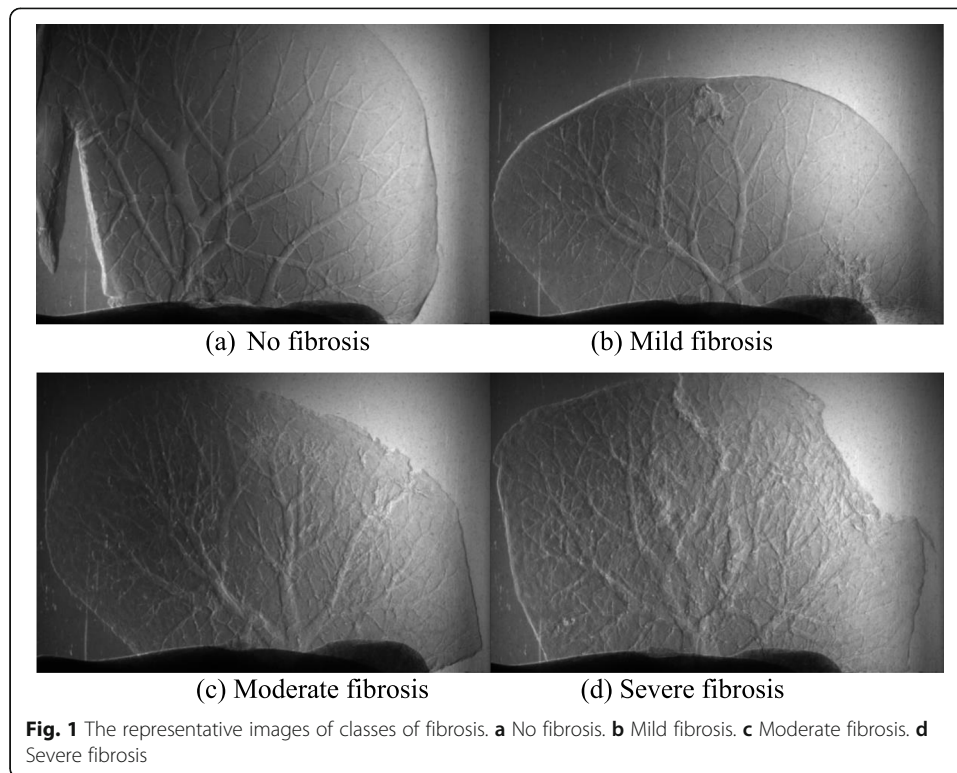## 2 Materials and methods

2.1 Materials

The hepatic fibrous samples of rats are prepared in the Peking University Health Science Center. Hepatic fibrosis in rats is induced by human albumin. The liver samples are cut into about 2 mm thick sections for DEI imaging. Procedures involving animals and their care are carried out in accordance with the NIH guidelines (NIH Pub. No. 85–23, revised 1996). The DEI images of rats are collected at the 4W1A topography and imaging station of Beijing Synchrotron Radiation Facility. There are four different grades of liver fibrosis: no fibrosis (normal), mild fibrosis, moderate fibrosis, and severe fibrosis. Figure 1 shows the representative DEI images of each class. After the imaging examination, the specimens are made into histology pictures, which are reviewed by three pathologists from the Third Hospital of Peking University. Fibrosis stages are evaluated according to the METAVIR scorning system. The diagnostic results obtained by the three pathologists are used as the gold standard for assessment.

2.2 Methods

The staging of hepatic fibrosis can be considered as a classification problem. The classification framework is shown in Fig. 2. All the algorithms are implemented using the Matlab R2016b software. More details are given as follows.

2.2.1 Image pre-processing stage

Pre-processing is performed on the DEI images to achieve reliable and acceptable classification accuracy. The main objective of this stage is to eliminate the background noise, reduce the difference of illumination, and normalize the DEI images. The background images with the same experiment conditions and parameters are collected. The

Fig. 1 The representative images of classes of fibrosis. **a** No fibrosis. **b** Mild fibrosis. **c** Moderate fibrosis. **d** Severe fibrosis

DEI images are divided by the background image with the same experiment condition to remove the background noise. After preprocessing, the light difference is removed, and the background noise is reduced. After the removal of background noise, the image gray level is standardized to be between 0 and 1.

The liver region is segmented manually. The regions of interest (ROIs) are obtained from liver regions, which do not contain big vessels in the normalized images. The size of ROIs is 100 × 100 pixels. A total of 11,169 ROIs are selected in the rat data set, including 2090 normal samples, 2246 mild samples, 3306 moderate samples, and 3527 severe samples. These ROI samples will be used for feature extraction.

2.2.2 Feature extraction stage

Feature extraction is a critical step in fibrosis staging. Texture pattern of hepatic fibrosis is one of the most important biomarkers in the diagnosis of hepatic fibrosis grade. There are different texture structures in different stages of liver fibrosis. Therefore, these texture features are extracted to describe the hepatic fibrosis. The first order statistic features (FO), texture feature-based gray level co-occurrence matrix (GLCM) [23], texture feature-based grayscale gradient co-occurrence matrix (GGCM) [24] are employed separately.

The FO features describes the distribution of image pixel intensities, which can be obtained based on ROI images. A total of 14 first order texture features including energy, entropy, kurtosis, maximum, mean, mean absolute deviation, median, minimum, range, root mean square, skewness, standard deviation, uniformity, and variance, are calculated.
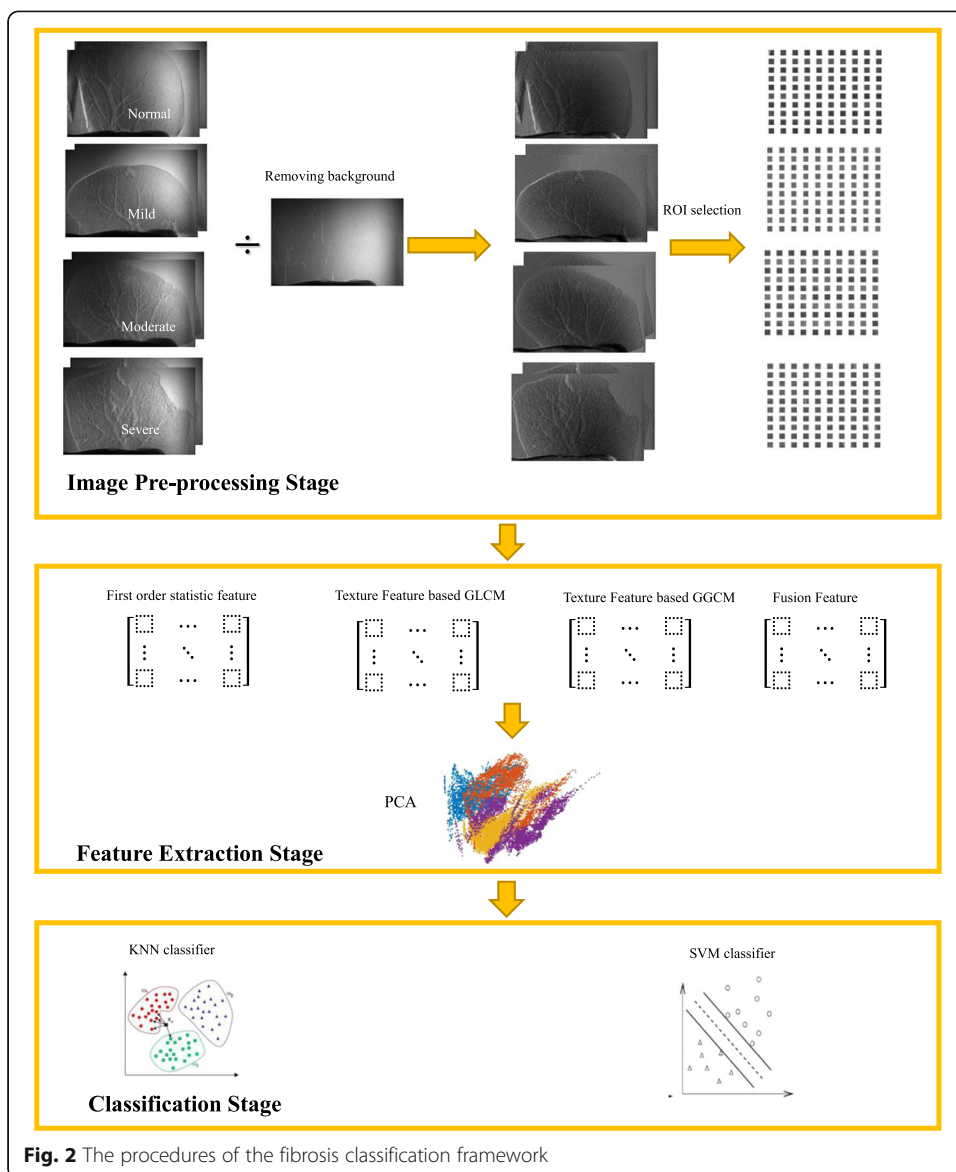
Fig. 2 The procedures of the fibrosis classification framework

GLCM describes the distribution of co-occurring pixel values at a given offset and reflects the relationship between the neighbor pixels. The gray level of ROI images is reduced to 32 bits to save the computation time. The matrices of GLCM are calculated symmetrically at 0, 45, 90, and 135° with 1-pixel distance and normalized thereafter. For one ROI image, four GLCMs are obtained. Nine features including angular second moment (energy), inertia, inverse, difference moment, entropy, sum average, difference average, sum entropy, and difference average are extracted from each GLCM matrix. Thus, a total of 36 features can be extracted from each ROI image.

GGCM can discover the joint distribution of gray and gradient of pixel in the image and reflects the spatial relationship between each pixel and its neighboring pixels, which depicts the image texture in a clear way. The texture feature-based GGCM is calculated based on ROI images. Gradient image is obtained by using the gradient of each pixel in the ROI image. The gradient of each pixel is calculated by 3 × 3 Sobel operator. The gradient images are normalized to 32 gray levels to reduce the computation

time. GGCM is achieved with the ROI image and gradient image. A total of 15 statistical characteristics including small gradient advantage, big gradient advantage, inhomogeneity of grayscale distribution, non-uniformity of gradient distribution, energy, gray mean, gradient mean, grayscale standard deviation, gradient standard deviation, correlation, gray entropy, gradient entropy, entropy of mixing, differential moment, and inverse moment can be acquired by using GGCM.

Each kind of features represents the characteristics of one aspect of the image. Therefore, it is necessary to study the fusion features. To simplify our presentation, we introduce the following abbreviations: the fused feature-based FO and GLCM are abbreviated as FO_GLCM, the fused feature-based FO and GGCM FO_GGCM, the fused feature-based GLCM and GGCM GLCM_GGCM, the fused feature-based FO, GLCM, and GGCM are FO_GLCM_GGCM.

The next step is dimension reduction for improving the classification efficiency. Principal component analysis (PCA) [25] is adopted here to reduce feature dimension and minimize the feature space.

2.2.3 Classification stage

The K-nearest neighbors (KNN) [26] and support vector machines (SVM) [27, 28] are used as classification approaches in our work.

In the K-NN classification, samples are classified based on their distances to other samples in training data set. In this paper, different neighbor number k and different distance functions are studied to obtain the best classification result.

The SVM classifier aims at finding the super hyperplane with the largest margin between two classes of samples in the feature space. Kernel functions project the data from a low dimensional space into a high-dimensional feature space to improve linear separability as much as possible. Different kernel functions and kernel scales are used. Liver fibrosis staging is a multi-class problem, and we need to construct a suitable multi-class classifier. In this paper, one-versus-one method is applied to combine multiple binary classifiers to build up multi-class classifier.

The performance of classification depends on both the training and test samples. Ten-fold cross validation is used to train and evaluate classification model in our work. The total number of training samples is 11,169. These samples are divided into 10 groups, of which nine groups have 1118 samples and one group has 1107 samples. Cross validation is repeated for 10 times with each sub-sample validated once. The experimental results demonstrated in our paper are the average of 10 trials.

Accuracy, specificity, and sensitivity are deployed to evaluate the performances of the classifiers, such as training time and prediction accuracy. Parallel computing can be utilized to improve the efficiency [29, 30].

3 3 results and discussions

3.1 Results and analysis

In our experiments, FO, GLCM, GGCM, FO_GLCM, FO_GGCM, GLCM_GGCM, and FO_GLCM_GGCM are separately calculated. Each type of features is classified using the KNN and SVM classifier separately.

3.1.1 Performances of different features using KNN classifier

The performance of KNN classifier is affected by the number of neighbors (k) and distance metric. The KNN classifiers are tested with different numbers of neighbors and different distance metrics so as to find the best parameters for staging hepatic fibrosis. The number of neighbors is set to be between 1 and 50, and distance metrics considered are Euclidean distance, cosine angle distance, Murkowski distance, Mahalanobis distance, city distance, Chebychev distance, correlation distance, Hamming distance, Jaccard distance, and Seuclidean distance. Experimental results show that the KNN classification accuracy is the highest when the number of neighbors is set 1. Therefore, in the following experiments, the number of neighbors is set to 1.

Figure 3 shows accuracy of KNN classifiers with different distance metrics. Apparently, Minkowski distance, Euclidean distance, and Cosine distance have achieved the best classification accuracy among the ten distance metrics. Table 1 shows the performance of the KNN classifiers with Euclidean distance, Cosine distance, and Minkowski distance in more detail. We can observe that when KNN classifier is used in staging fibrosis, Minkowski distance gives the highest accuracy and sensitivity, Cosine distance metric the highest specificity, and Euclidean distance the highest efficiency. Although the accuracy of the KNN classification obtained by Euclidean distance is slightly lower than that of Minkowski distance, the training time is shorter than that of the latter, and the prediction speed of the former is much faster than that of the latter. Thus, when KNN is used to staging fibrosis, to pursue algorithm accuracy and sensitivity, we can choose Minkowski distance; to pursue efficiency, we can choose Euclidean distance; and to pursue specificity, we can choose Cosine distance metric.

The performance of each feature set-based KNN is provided in Table 2. In the KNN, the number of the nearest neighbors is set to be 1, and Euclidean distance is used as the distance metric. Accuracy of all KNN classifiers based on different types of features is all higher than 98.5%. Among the classification results of single type features, GGCM achieves the highest classification accuracy and specificity by using only 15 features, and GLCM achieves the highest accuracy and sensitivity by using 36 features. Compared with GLCM, the training time of GGCM is shorter, and the prediction speed is faster. Thus, although the accuracy obtained by GLCM and GGCM are comparable, it is more suitable to use GGCM. GGCM features are the best choices for grading liver fibrosis with single type features and KNN.
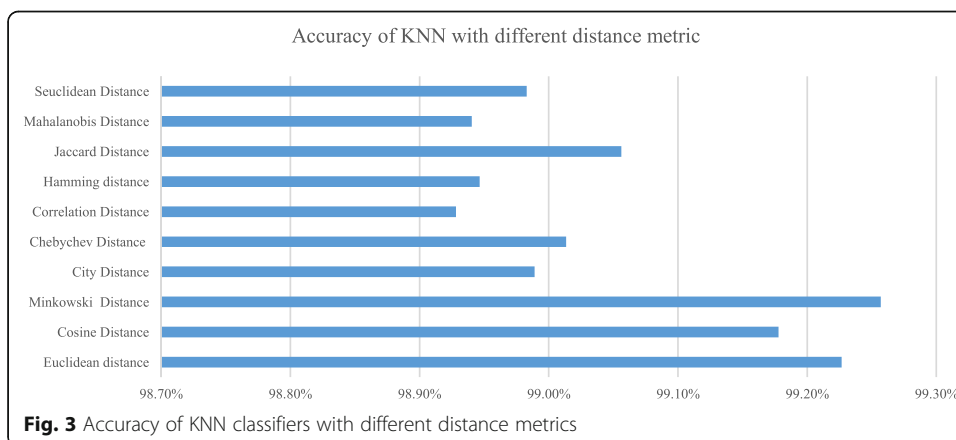

Fig. 3 Accuracy of KNN classifiers with different distance metrics

Table 1 Performances of KNN with Euclidean, Cosine, and Minkowski distance metrics

Distance metric	Accuracy	Sensitivity	Specificity	Training time(s)	Prediction speed (obs/s)
Euclidean	99.21%	99.38%	98.54%	2.4916	18,000
Cosine	99.13%	99.22%	98.80%	2.503	18,000
Minkowski	99.23%	99.39%	98.54%	58.548	570

The accuracy, sensitivity, and specificity of KNN classifiers are improved by the fusion of different types of features. Among the classification results of the fusion features with two types of features, their accuracies are the same. However, FO_GLCM obtains the highest sensitivity, FO_GGCM obtains the highest specificity, and FO_GGCM obtains the highest efficiency. With regard to the fusion of FO, GLCM, and GGCM, the accuracy is improved to 99.8%, and the specificity is improved to 99.92%. FO_GLCM_GGCM obtains the highest accuracy and specificity among all the features using KNN classifier. To conclude, if KNN classifier is used for staging hepatic fibrosis, FO_GLCM_GGCM feature is the best choice for pursuing high accuracy and low misdiagnosis rate, FO_GLCM feature is the best choice for pursuing diagnosis accuracy, and GGCM is the best choice for pursuing efficiency.
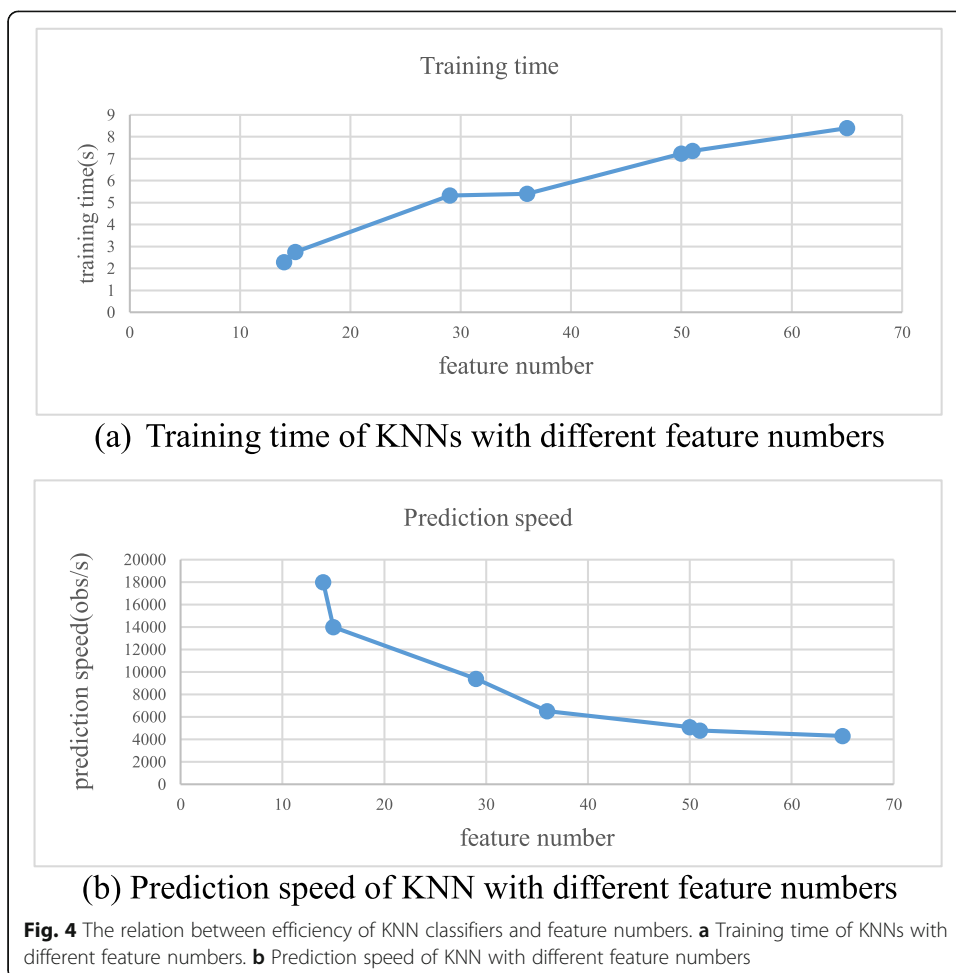
Figure 4 a, b shows the training time and testing speed of KNN classifiers with different numbers of features, respectively. The training time increases, and the testing speed decreases with the growth of the number of features. We can conclude that the training time of KNN classifier with fusion feature is longer, and the prediction speed is slower, because the number of fusion features is more than that of single type of features.

3.1.2 Performances of different types of features using SVM classifier

The kernel function and kernel scale have a significant influence on the performance of SVM classifiers. The SVM classifiers with different kernel functions and kernel scales are trained to find the best parameters, which are suitable for classifying hepatic fibrosis. Four kernel functions are adopted, i.e., linear kernel, quadratic polynomial kernel, cubic polynomial kernel, and Gaussian kennel. Kernel scales are set to be between 0.01 and 2 at an interval of 0.01. The performances of the SVM classifiers with the best kernel scale and different kernel functions are given in Table 3. Apparently, when SVM classifier is used in staging fibrosis, Cubic kernel function yields the highest accuracy and specificity among the four kernel function, but its training time is longer. Gaussian kernel obtains the highest sensitivity, and it is the most efficient in training phase as its

Table 2 The performances of KNN classifiers based on different kind of features

Feature set	Feature number	Accuracy	Sensitivity	Specificity	Training time (s)	Prediction speed (obs/s)
FO	14	98.60%	98.85%	97.63%	2.2781	18,000
GLCM	36	99.20%	99.48%	97.96%	5.3947	6500
GGCM	15	99.20%	99.42%	98.47%	2.7419	14,000
FO_GLCM	50	99.60%	99.90%	98.65%	7.2376	5100
FO_GGCM	29	99.60%	99.66%	99.25%	5.3189	9400
GLCM_GGCM	51	99.60%	99.76%	99.10%	7.354	4800
FO_GLCM_GGCM	65	99.80%	99.19%	99.92%	8.3886	4300

Fig. 4 The relation between efficiency of KNN classifiers and feature numbers. **a** Training time of KNNs with different feature numbers. **b** Prediction speed of KNN with different feature numbers

training time is the shortest. However, its accuracy, specificity, and prediction speed are slightly lower than those of Cubic kernel function. Therefore, when SVM classifier is applied to staging fibrosis, to pursue accuracy and specificity, Cubic kernel can be chosen. On the other hand, if the sensitivity and training speed of the algorithm are considered, Gaussian kernel function is a competitive candidate.

The performances of different types of features based SVM are given in Table 4. In the SVM classifiers, Gaussian kernel is adopted.

Accuracies of all SVM classifiers based on different type of features are higher than 98.4%. Among the SVM classification results of single type features, GLCM achieves the highest accuracy and highest sensitivity; GGCM can obtain the highest specificity with relatively shorter training time and faster prediction speed. Although the classification accuracy and sensitivity obtained by using GGCM features are slightly lower than

Table 3 Performances of SVM classifiers

Kernel function	Kernel scale	Accuracy	Sensitivity	specificity	Training time (s)	Prediction speed (obs/s)
Linear	0.07	93.83%	94.32%	91.90%	2367.9	170,000
Quadratic	0.37	99.31%	99.66%	97.93%	1655.7	72,000
Cubic	0.78	99.57%	99.69%	99.10%	699.22	76,000
Gaussian	0.55	99.49%	99.73%	98.14%	93.538	6900

Table 4 The performances of SVM classifiers based on different kinds of features

Feature Set	Accuracy	Sensitivity	Specificity	Training time (s)	Prediction speed (obj/s)
FO	98.41%	98.68%	97.39%	58.413	12,000
GLCM	99.49%	99.73%	98.56%	93.538	6900
GGCM	99.31%	99.35%	99.15%	61.643	11,000
FO_GLCM	99.73%	98.85%	99.25%	40.354	13,000
FO_ GGCM	99.68%	99.75%	99.40%	45.398	13,000
GLCM_GGCM	99.73%	99.82%	99.37%	41.978	12,000
FO_GLCM_GGCM	99.81%	99.84%	99.69%	48.903	10,000

those of GLCM features, the number of features is far less, the training time is shorter, and the prediction speed is faster. Hence, using GGCM features is more efficient for feature extraction and classification than GLCM features.

The accuracy, sensitivity, and specificity of SVM classifiers are improved by the fusion of two types of features. Among the classification results of the fusion features with the two types of features, FO_GLCM and GLCM_GGCM can obtain the highest accuracy, FO_GLCM the highest sensitivity, and FO_GGCM the highest specificity. Moreover, with the fusion of all the types of features, the accuracy is improved to 99.81%, sensitivity 99.69%, and specificity 99.69%. The training time is 48.903 s, and prediction speed is 10,000 objects each second. The conclusion is that when SVM classifier is used for staging the liver fibrosis, the highest accuracy, sensitivity, and specificity can be obtained with FO_GLCM_GGCM. In other words, FO_GLCM_GGCM feature is the best choice when pursuing accuracy, low misdiagnosis rate, and high diagnosis accuracy.

3.1.3 3Feature dimension reduction

The above experimental results show that the accuracy, sensitivity, and specificity of KNN classifier are improved by the fusion of different types of features compared with single type features. However, with the growth of fusion feature dimension, the efficiency of classification may decrease, which means that the training time will increase, and prediction speed will decrease. To cope with this issue, PCA method is used to reduce dimensions as well as improve the efficiency of classifier. The relationship between the accuracy of KNN classifier and PCA dimensionality reduction is shown in Fig. 5. It is clearly visible that the classification accuracy increases with the growth of feature dimension. However, when the feature dimension increases to a certain amount, accuracy becomes stable.

Figure 6 shows the performance of the classifiers before and after dimensionality reduction using KNN classifiers with different types of features. From Fig. 6, it can be seen that PCA dimension reduction improves classification accuracy, sensitivity, and specificity, reduces training time, and speeds up prediction. With the PCA dimensionality reduction, we can use fewer features to achieve higher accuracy, sensitivity, and specificity. This indicates that the feature has redundant information that can be removed by PCA. After dimension reduction of FO_GLCM_GGCM feature by PCA, the highest accuracy reaches 99.99%, the highest sensitivity 100%, and the highest specificity 99.97%.

Fig. 5 The relationship between the accuracy of KNN classifier and PCA dimensionality reduction. **a** GLCM features. **b** GGCM features. **c** FO_GLCM features. **d** GLCM_GGCM features. **e** FO_GGCM features. **f** FO_GLCM_GGCM features
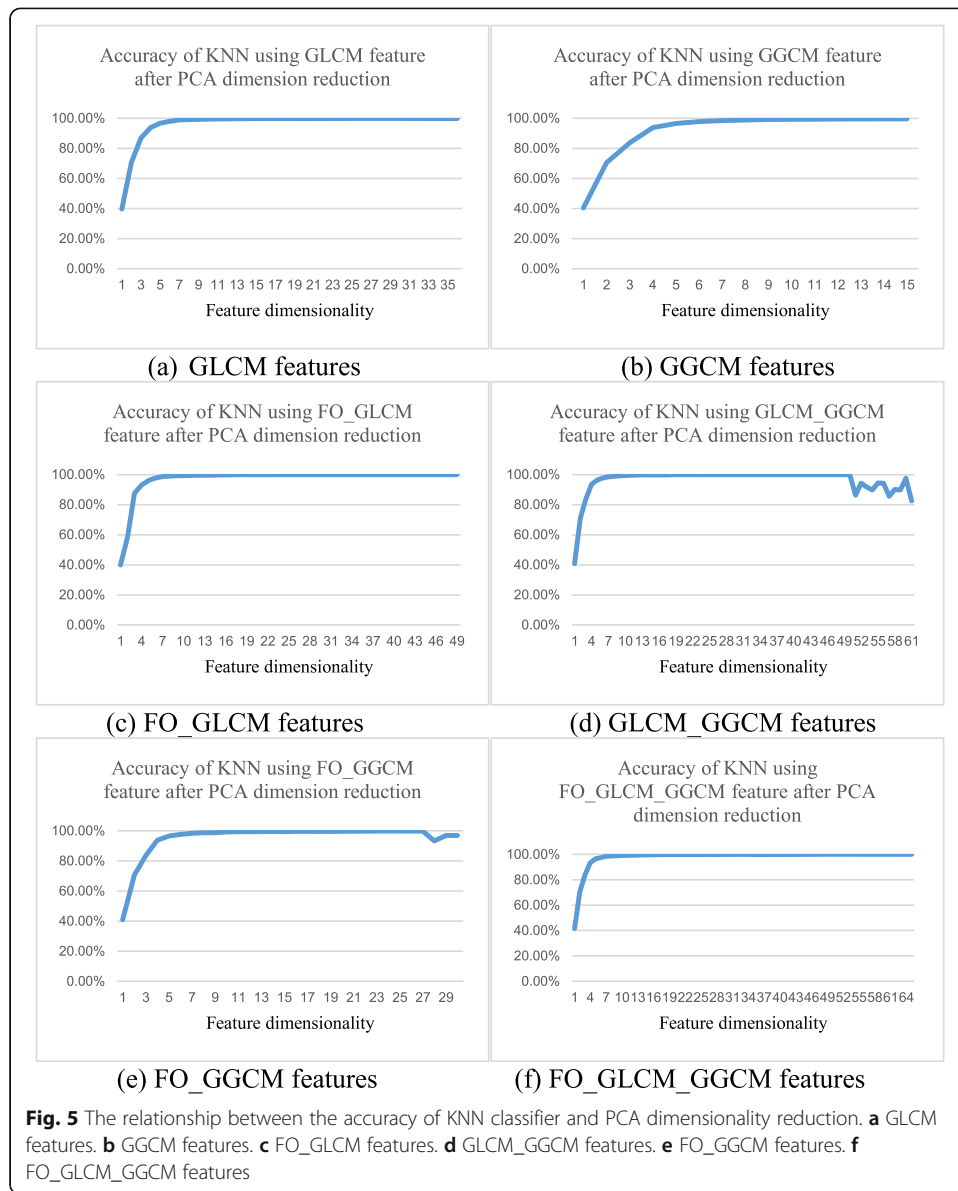
3.2 Discussions
The experimental results given in the above sections are discussed as follows:

1. Performance of different classifiers

Figure 7 shows the performance of KNN and SVM for different feature sets. It is clearly visible that (1) for 6/7 feature sets, the accuracy of SVM is better than that of KNN; (2) for 5/7 feature sets, the sensitivity of SVM is better than that of KNN; (3) for 6/7 feature sets, the specificity of SVM is better than that of KNN; (3) the training time of KNN is far lower than that of SVM, so the training speed of KNN is much faster; and (4) for FO and GGCM, the prediction speed of KNN is faster than that of SVM for the two feature sets with low feature dimension, while for the data sets with high feature dimension, the prediction speed of SVM
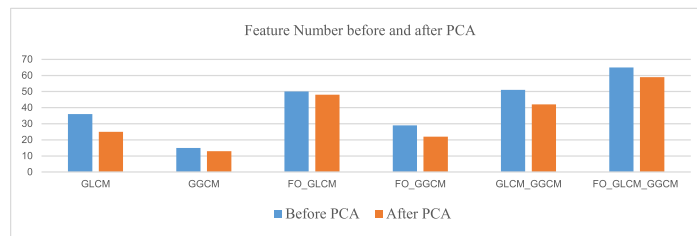
(a) Feature numbers before and after dimensionality reduction
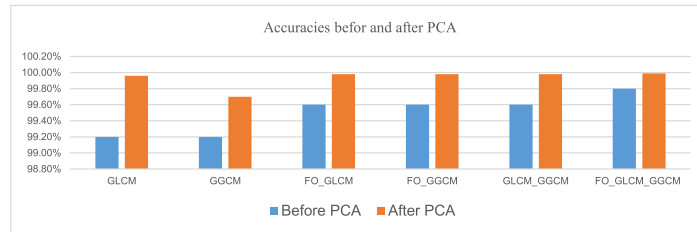
(b) Accuracy before and after dimensionality reduction
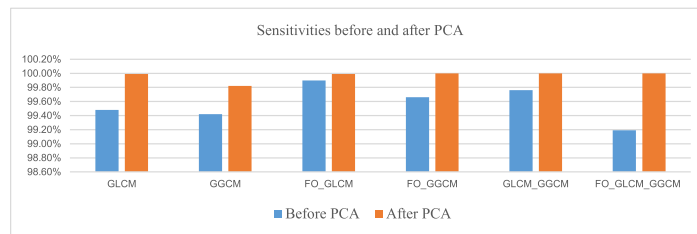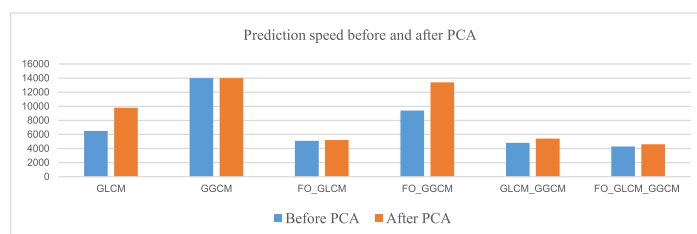

(c) Sensitivity before and after dimensionality reduction
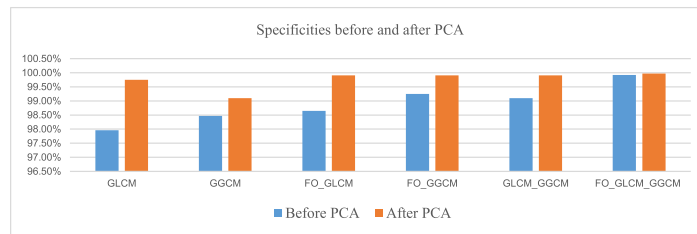
(d) Specificity before and after dimensionality reduction
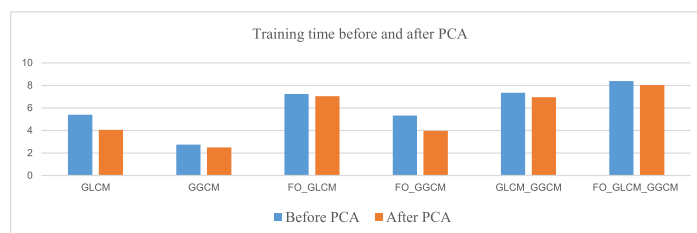
(e) Training time before and after dimensionality reduction

(f) Prediction speed before and after dimensionality reduction

Fig. 6 (See legend on next page.)

(See figure on previous page.)
Fig. 6 The performances of KNN classifier with different types of features before and after dimensionality reduction. **a** Feature numbers before and after dimensionality reduction. **b** Accuracy before and after dimensionality reduction. **c** Sensitivity before and after dimensionality reduction. **d** Specificity before and after dimensionality reduction. **e** Training time before and after dimensionality reduction. f Prediction speed before and after dimensionality reduction

is faster than that of KNN. In general, SVM is better than KNN in performance. KNN is better than SVM in efficiency.

2. Performance of different features

It can be discovered from Tables 2 and 4 that GLCM and GGCM are capable of achieving better accuracy, sensitivity, and specificity. On the other hand, the performance of FO is worse. This indicates that texture features extracted from the GLCM and GGCM are more discriminative than the FO in staging of hepatic fibrosis. Therefore, GGCM and GLCM are more suitable than FO for characterizing different grades of liver fibrosis, when a single type of feature is chosen to represent the degree of fibrosis.

Fusion of different types of features can improve the classification performance. Especially, the best classification results are obtained by the fusion of three types of features. There is complementary information useful for grading liver fibers among different types of features. Although the classification performance of single FO feature is not ideal, the classification performance is improved, when combined with GLCM or GGCM. Thus, we can conclude that FO, GLCM, and GGCM all contain valuable information for fibrosis grading.

3. Best classifier and feature combination

From Tables 2 and 4, it can be found out that combination of SVM classifier and FO_GLCM_GGCM feature achieves the highest accuracy; combination of KNN classifier and FO_GLCM feature achieves the highest sensitivity; and combination of KNN classifier and FO_GLCM_GGCM achieves the highest specificity. The shortest training time is achieved by KNN classifier with FO features; with both the performance and efficiency of the classifier under consideration, hybridization of KNN classifier with GGCM or FO_GLCM_GGCM is a good choice.

4. The effect of dimension reduction

After dimensionality reduction by PCA, a higher accuracy can be achieved by using fewer numbers of features. Obviously, features have redundant information to be removed by feature selection, and dimension reduction by PCA can improve the efficiency of grading liver fibrosis. From the experimental results presented in this paper, we discover that using FO_GLCM_GGCM after dimensionality reduction by PCA and KNN classifier achieves the highest accuracy, sensitivity, and specificity.

5. Performance comparison with the previous study

(a) Accuracy of KNN and SVM

(b) sensitivity of KNN and SVM

(c) specificity of KNN and SVM

(d) Training time of KNN and SVM

(e) Prediction speed of KNN and SVM

Fig. 7 The performances of KNN and SVM classifier. **a** Accuracy of KNN and SVM. **b** Sensitivity of KNN and SVM. **c** specificity of KNN and SVM. **d** Training time of KNN and SVM. **e** Prediction speed of KNN and SVM

In our preliminary work [22], GLCM texture feature and BP classifier were used to classify liver fibrosis. Figure 8 shows the result when the GLCM texture feature is classified by BP classifier in [22] and KNN and SVM classifier in this paper, respectively. It can be seen from it the classification of accuracy, sensitivity, and

Fig. 8 The performances of KNN, SVM, and BP classifier
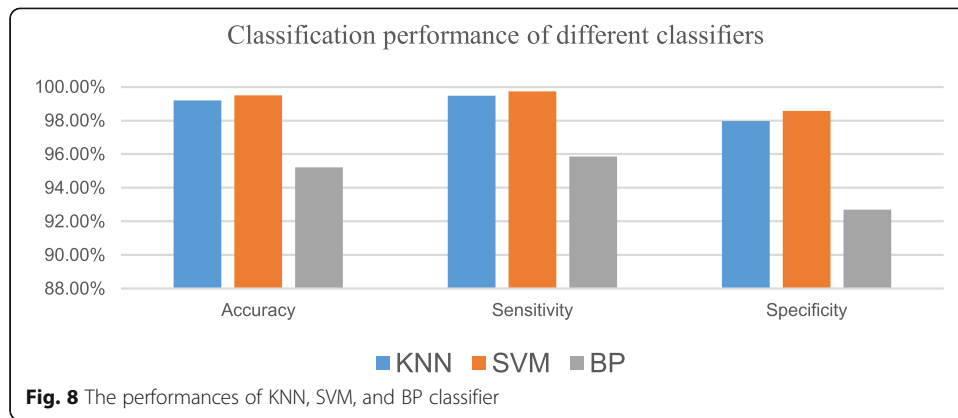
specificity of KNN and SVM in this paper are better than those of BP classifier in [22].

In addition to the further study of GLCM in [22], this paper also studies FO, GGCM, and fusion features. Figure 9 shows the performance of different features using the same classifier. As can be seen from it that except for FO feature, the accuracy and specificity of other features are better than GLCM. The sensitivity of FO_GLCM, FO_ GGCM, and GLCM_ GGCM is better than that of GGCM. Better feature extraction method to describe the hepatic fibrosis than GLCM is presented in this paper.

Table 5 summarizes similar study that has been carried out in the staging of liver fibrosis. Most of the results are based on US [11, 12], CT [13, 15, 21], and MRI [21] images. The work in [22] and this paper explore the DEI-based staging of hepatic fibrosis. According to the accuracies in Table 5, DEI images have certain advantages over US, CT, and MRI in hepatic fibrosis staging. In our preliminary work [22], BP network is applied to discriminate liver fibrosis in DEI image-based GLCM texture feature, which can achieve the accuracy rate of 95.1%. In the present paper, more types of texture features, fusion features, and features after dimensionality reduction are proposed to describe the hepatic fibers of DEI images, and a few more machine learning algorithms are applied to automatically classify different fibrosis stages. The classification performances have been significantly improved.

4 Conclusions

Due to the advantages of containing much more information on soft tissue morphology than the conventional absorption radiography, DEI images are used in this paper. In
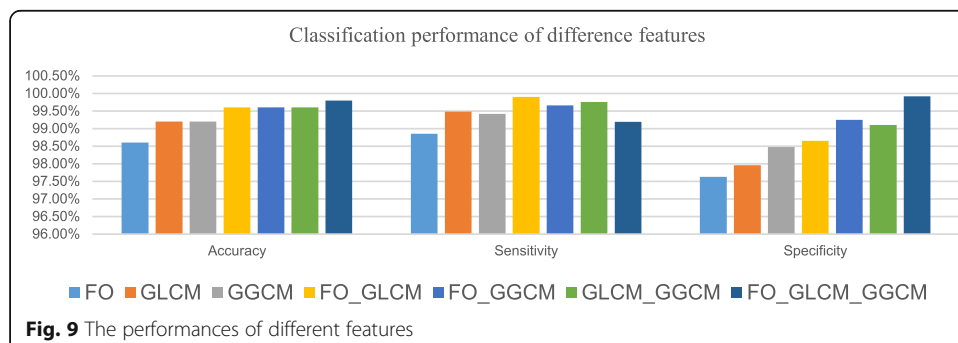

Fig. 9 The performances of different features

Table 5 Summary of similar study carried out on hepatic fibrosis staging

Author	Method	Image modality	Accuracy
J. H. Lee et al. [11]	DCNN	US	83.5%
Y. Chen et al. [12]	SVM, Naïve Bayes, random forest, KNN	US	82.87%
K.J. Choiet al [13].	DCNN	CT	79.4%
O. Kayaalti et al. [15]	GLCM, DWT, Gabor, KNN, SVM	CT	94%
X. Zhang et al. [21]	GLCM, SVM	CT/MRI	79%/82%
J. Wang et al. [22]	GLCM, BP	DEI	95.1%
Our methods	FO, GGAM, GLCM, Fusion feature, KNN, SVM,	DEI	**99.99%**

the proposed framework, different feature extraction strategies and classification methods are adopted and compared, which can provide a reference for the appropriate selection of models and parameter settings in hepatic fibrosis grading. Our approach can efficiently classify the hepatic fibrosis DEI images into four classes with a classification accuracy of 99.99%. The DEI has been proved to be a promising non-invasive technique in staging of hepatic fibrosis. The experimental results also demonstrate that grading the liver fibrosis using DEI images has interesting potentials in clinical applications. In our future work, new liver fibrous grading approaches based on deep learning techniques will be investigated.

Author details
[1]Department of Health Information Management, Peking University Health Science Center, Beijing 100191, China. [2]Department of Medical Technology, Peking University Health Science Center, Beijing 100191, China. [3]Department of Technology and Industry Development, Beijing City University, Beijing 100083, China. [4]School of Computing, University of Eastern Finland, Kuopio, Finland.

References
1.   D.C. Rockey, S.H. Caldwell, Z.D. Goodman, et al., Liver biopsy. Hepatology. **49**, 1017–1044 (2009)

2. S. Saadeh, G. Cammell, W.D. Carey, et al., The role of liver biopsy in chronic hepatitis C. Hepatology. **33**, 196–200 (2001)
3. A. Regev, M. Berho, L.J. Jeffers, et al., Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am. J. Gastroenterol. **97**, 2614–2618 (2002)
4. V. Ratziu, F. Charlotte, A. Heurtier, et al., Sampling variability of liver biopsy in nonalcoholic fatty liver disease. Gastroenterology. **128**, 1898–1906 (2005)
5. C. Aubé, Imaging modalities for the diagnosis of hepatic fibrosis and cirrhosis. Clinics and Research in Hepatology and Gastroenterology **39**, 38–44 (2015)
6. R.P. Mathew, S.K. Venkatesh, et al., Imaging of hepatic fibrosis. Current Gastroenterology Reports **20**, 45 (2018)
7. S. Bonekamp, I. Kamel, S. Solga, et al., Can imaging modalities diagnose and stage hepatic fibrosis and cirrhosis accurately? J. Hepatol. **50**, 17–35 (2009)
8. D. Chapman, W. Thomlinson, R.E. Johnston, et al., Diffraction enhanced x-ray imaging. Phys. Med. Biol. **42**, 2015–2025 (1997)
9. R. Fitzgerald, Phase-sensitive x-ray imaging. Phys. Today **53**, 23–26 (2000)
10. H. Li, L. Zhang, X. Wang, et al., Investigation of hepatic fibrosis in rats with x-ray diffraction enhanced imaging. Appl. Phys. Lett. **94**, 124101 (2009)
11. J.H. Lee, L. Joo, T.W. Kang, et al., Deep learning with ultrasonography: Automated classification of liver fibrosis using a deep convolutional neural network. Eur. Radiol. **30**, 1264–1273 (2020)
12. Y. Chen, Y. Luo, W. Huang, et al., Machine-learning-based classification of real-time tissue elastography for hepatic fibrosis in patient with chronic hepatitis B. Comput. Biol. Med. **89**, 18–23 (2017)
13. K.J. Choi, J.K. Jang, S.S. Lee, et al., Development and validation of a deep learning system for staging liver fibrosis by using contrast agent-enhanced CT images in the liver. Radiology **289**, 688–697 (2018)
14. K. Yasaka, H. Akai, A. Kunimatsu, et al., Deep learning for staging liver fibrosis on CT: A pilot study. Eur. Radiol. **28**, 4578–4585 (2018)
15. O. Kayaalti, B.H. Aksebzeci, I.O. Karahan, et al., Liver fibrosis staging using CT image texture analysis and soft computing. Appl. Soft Comput. **25**, 399–413 (2014)
16. K. Yasaka, H. Akai, A. Kunimatsu, et al., Liver fibrosis: Deep convolutional neural network for staging by using gadoxetic acid-enhanced hepatobiliary phase MR images. Radiology **287**, 146–155 (2018)
17. N. Daginawala, B. Li, K. Buch, H. Yu, et al., Using texture analyses of contrast enhanced CT to assess hepatic fibrosis. Eur. J. Radiol. **85**, 511–517 (2016)
18. M.J. House, S.J. Bangma, M. Thomas, et al., Texture-based classification of liver fibrosis using MRI. J. Magn. Reson. Imaging **41**, 2322–2328 (2015)
19. D. Yoshimaru, T. Miyati, Y. Suzuki, et al., Diffusion kurtosis imaging with the breath-hold technique for staging hepatic fibrosis: A preliminary study. Magn. Reson. Imaging **47**, 33–38 (2018)
20. R. Cannella, A.A. Borhani, M. Tublin, et al., Diagnostic value of MR-based texture analysis for the assessment of hepatic fibrosis in patients with nonalcoholic fatty liver disease (NAFLD). Abdominal Radiol. **44**, 1816–1824 (2019)
21. X. Zhang, X. Gao, J. Brent, et al., Effective staging of fibrosis by the selected texture features of liver: Which one is better, CT or MR imaging? Comput. Med. Imaging Graph. **46**, 227–236 (2015)
22. J. Wang et al. Evaluation of texture features at staging liver fibrosis based on phase contrast X-ray imaging. BioMedical Engineering Online., 17(2018)
23. D.A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization. Can. J. Remote. Sens. **28**, 45–62 (2002)
24. J. Hong, Gray level-gradient co-occurrence matrix texture analysis method. Acta Automat. Sin. **10**, 22–25 (1984)
25. H. Abdi, L.J. Williams, Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics. **2**, 433–459 (2010)
26. W.C. Lin, S.W. Ke, C.F. Tsai, CANN: An intrusion detection system based on combining cluster centers and nearest neighbors. Knowl.-Based Syst. **78**, 13–21 (2015)
27. A. K. Sangaiah, D. V. Medhane, T. Han, et al., Enforcing position-based confidentiality with machine learning paradigm through mobile edge computing in real-time industrial informatics. IEEE Transactions on Industrial Informatics, 1-1 (2019)
28. M. Jordan, J. Kleinberg, B. Scholkopf, Support vector machine. Information Science and Statistics. 287-330 (2008)
29. A.K. Sangaiah, M.Y. Suraki, M. Sadeghilalimi, et al., A new meta-heuristic algorithm for solving the flexible dynamic job-shop problem with parallel machines. Symmetry **11**, 165 (2019)
30. A. K.Sangaiah, E. B. Tirkolaee, Goli, A. Dehnavi-Arani S, et al. Robust optimization and mixed-integer linear programming model for LNG supply chain planning problem. Soft Computing,1-21(2019)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.