Expert Systems With Applications 89 (2017) 31–40

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Joint discriminative and collaborative representation for fatty liver disease diagnosis☆

Jinxing Li[a], Bob Zhang[b], David Zhang[a,c,*]

[a] Biometrics Research Center, Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[b] Department of Computer and Information Science, University of Macau, Macau, China
[c] Department of Computer Science, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

ARTICLE INFO

Article history:
Received 20 April 2017
Revised 26 June 2017
Accepted 15 July 2017
Available online 17 July 2017

Keywords:
Fatty liver
Multi-modal (task)
Tongue image
Facial image
Collaborative representation
Discriminative representation

ABSTRACT

Many people suffer from the Fatty Liver disease due to the changes in diet and lifestyle, and the convenient diagnosis of it has attracted many attentions in recent years. The computerized tongue or facial diagnosis as an important diagnostic tool provides a possible way to detect the disease in the daily life. Most of existing approaches only takes a single modality (e.g., tongue or face) into account, although various modalities would contribute complementary information which is beneficial for the improvement of the diagnosis accuracy. To circumvent this issue, a novel multi-modal fusion method is presented in this paper. Particularly, a noninvasive capture device is first used to captured the tongue and facial images, followed by the feature extraction. Our so-called joint discriminative and collaborative representation approach is then proposed to not only reveal the correlation between the tongue and facial images, but also keep the discriminative representation of each class simultaneously. To optimize the proposed method, an efficient algorithm is proposed, obtaining a closed-form solution and greatly reducing the computation. In identification of the Fatty Liver Disease for healthy controls, the proposed multi-modal fusion approach achieves 85.10% in average accuracy and 0.9363 in the area under ROC curve, which obviously outperform the case of using a single modality and some state-of-the-art methods.

1. Introduction

Due to the changes in diet and lifestyle in western countries and many Asian countries, the prevalence of obesity and metabolic syndrome have increased dramatically, increasing the number of people suffering from the Fatty Liver disease (Hashimoto, Tokushige, & Ludwig, 2015; Amarapurkar et al., 2007; Deng et al., 2014; Bucak & Baki, 2010; Gunasundari, Janakiraman, & Meenambal, 2016). The type-B ultrasonic and Computed Tomography (CT) as the general methods of detecting the Fatty Liver disease are widely applied in most hospitals. Although adopting these two methods has achieved a remarkable performance, it is inconvenient for many people to diagnose in the daily life. Thus a novel approach for the Fatty Liver disease diagnosis is necessary.

The tongue or facial diagnosis as a staple has been practiced in the traditional medicine system for many years. Particularly, in Traditional Chinese Medicine (TCM) (Wang, Zhang, Yang, Wang, & Zhang, 2013; Zhang, Kumar, & Zhang, 2014b; Kirschbaum, 2000; Wang & Zhang, 2013; Wang, Zhang, & Lu, 2016) for instance, the belief is that the color, texture and geometric changes of the patient's tongue and face are capable of reflecting symptoms of a certain disease. The traditional tongue or facial diagnostic is heavily based on the practitioner's years of experience, restricting its application in wider fields. Fortunately, some works quantitatively extracting and analyzing different features of tongue and facial images for diagnosis were established in recent years, and have shown their effectiveness in many disease detections. Therefore, in this paper, a novel multi-modal fusion method which considers the tongue and face simultaneously is proposed to detect the Fatty Liver disease.

Six centroids extracted from the facial color gamut were first presented by Zhang et al. (2014b). The facial color feature was then obtained according to the centroids, and achieved a superior performance in Diabetes Mellitus (DM) detection. In addition, the heart and hepatitis disease diagnostic systems based on facial color were proposed by Kim, Lee, Cho, and Oh (2008) and Liu and Guo (2008), respectively. Apart from the color feature, Zhang, Wang, Karray, Yang, and Zhang (2013) also took texture features

☆ Fully documented templates are available in the elsarticle package on CTAN.

* Corresponding author at : Biometrics Research Center, Department of Computing, The Hong Kong Polytechnic University,Hung Hom, Kowloon, Hong Kong, China.
E-mail addresses: csjxli@comp.polyu.edu.hk (J. Li), bobzhang@umac.mo (B. Zhang), csdzhang@comp.polyu.edu.hk (D. Zhang).

http://dx.doi.org/10.1016/j.eswa.2017.07.023
0957-4174/

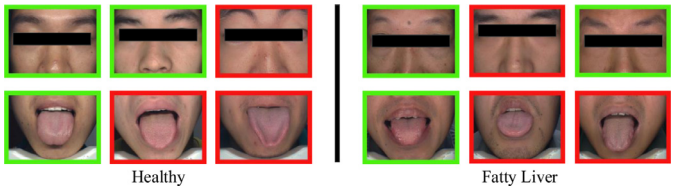

Fig. 1. Some typical instances classified by using SRC. The correct classification is described by the green border and the incorrect classification is denoted by the red border. The classified instances based on facial images are shown in the first row, and the classified instances based on tongue images are displayed in the second row. The left part denotes the 'healthy' samples and the right part denotes the 'Fatty Liver' samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

into account by Gabor filter and achieved a prominent performance. Wang, Zhang, Yang et al. (2013) first designed an image capture device which not only gets a high-resolution image but also reduces the influence of different lights. Similarly to Zhang et al. (2014b), 12 color centroids were statistically studied from a large number of tongue images in Wang's study (Wang, Zhang, Yang et al., 2013) to reflect the color changes in tongue images caused by diseases. Simultaneously, a novel DM and Nonproliferative Diabetic Retinopathy detection method was presented in Zhang's research (Zhang, Kumar, & Zhang, 2014a) by exploiting the color, texture and geometry features of the tongue.

Although a number of tongue or face based diagnostic systems have been established, most of them considered either the tongue or the face independently and ignored the relationship between them. It is reasonable to assume that there is some latent and correlated information between the tongue and face modalities, and these types of factors may have a contribution for the overall classification performance. Specially, a practitioner may be difficult to make a diagnostic decision only based on tongue or facial features, while an integration of these two modalities may contribute to obtaining more accurate detection results. As shown in Fig. 1, we can see that some Healthy or Fatty Liver samples can not be detected by exploiting facial features but tongue features does. Similarly, in some cases, using facial images may fail to get an accurate diagnosis, but using tongue images may succeed. Moreover, there is a case that we can not get an accurate diagnosis no matter using the tongue or the face, but a combination of these two images may be possible. Therefore, the fusion of the tongue and facial features is crucial for diagnosis.

A naive way of fusing the tongue and facial features is to concatenate them as a single one. However, it fails to discover the latent correlation between them. By utilizing both of the tongue features and the facial features for classification, an efficient multimodal approach is presented in this paper. Specially, our approach effectively exploits the similarity of two different modalities, extracting the relationship between tongue and facial features. Meanwhile, the discriminative information among each class is kept. In this case, the so-called joint discriminative and collaborative representation (JDCR) not only reveals the similar representation, but also encourages the representations of different categories to be more discriminative and de-correlate for the test sample (Xu, Zhong, Yang, You, & Zhang, 2016). An optimization algorithm is presented to calculate the closed-form solution in a computationally efficient way. Finally, we apply JDCR in discriminating between Fatty Liver control. To be honest, especially in the Guangdong Provincial TCM Hospital, there is no accurate way to detect the fatty liver disease using only type-B ultrasonic or CT. These two medical imaging examination methods can only give an uncertain result that the patient may suffer from the fatty liver disease, but not an accurate diagnosis. In order to get a reliable detection, further examinations such as the liver function test and the blood fat test are needed. In practice, there are two ways for the accurate fatty liver disease diagnosis. The first one is a combined judgment. For instance, in the Guangdong Provincial Traditional Chinese Medicine Hospital, a patient is detected with the disease if and only if his or her type-B ultrasonic (or CT), liver function, and blood fat tests are uncommon simultaneously. The second method is the needle biopsy of liver (Evans, 1952), which needs to locate certain cells in the liver. It can be easily seen that both strategies are inconvenient, invasive and painful. However, our method finds the valuable information in the tongue and facial images to do the fatty liver disease diagnosis, which is quite easy to achieve. Despite the fact that our detection accuracy may be a little lower than that of the two strategies mentioned above, we aim to provide a self reliant device for people to diagnose by themselves at any time. If a person is detected with the fatty liver disease by the proposed method, we would advise him or her to go to the hospital to have a further examination.

We firstly describe the tongue and facial image capture device, as well as the feature extraction in Section 2. The proposed joint discriminative and collaborative representation algorithm is then analyzed in Section 3. The experimental results and the conclusion are illustrated in Section 4 and Section 5, respectively.

2. Image capture device and feature extraction

In this section, the tongue and facial image capture device is introduced, followed by the feature extraction of images in the two modalities.

2.1. Tongue and facial capture device

The primary issue for computerized tongue and facial diagnosis is the reliable and robust image capture device. Since the light in different environment have an influence on the representation of images, we have designed a semi-closed device, as shown in Fig. 2. As we can see that two components including two D65 fluorescent tubes and a SONY video camera construct the device. Concretely, these two fluorescent tubes are placed on either sides symmetrically and the camera is placed on tubes' center. In this way, a uniform illumination is provided which ensures the accurate and consistent representation of images. As shown in Fig. 2(b), we further design a chin rest which can be fixed in different height. Thus the tongue and face can be captured by changing the height of the chin rest. In this system, the captured image is saved in JPEG format with 640×480 size. Additionally, in order to reduce the variability in color images caused by the difference of the device dependence, a color correction algorithm (Wang & Zhang, 2010) is applied. Using this correction model, the pixels in the original tongue and facial images are normalized in a standard RGB color space.

2.2. Feature extraction

Three types of features including color, texture and geometry of captured images were analyzed in our previous work (Zhang et al., 2014a; 2014b). Since some parts (e.g., eyes, mouths and noses etc.) have no contribution to the diagnosis, four blocks with 64×64 size in the facial image are extracted around the face, which can reflect the health status of a person (Maciocia, 1989). For the facial image, both the color feature and texture feature extractions are based on these four blocks. Similarly, eight blocks with 64×64 size in the tongue image are also located, and the texture feature extraction for the tongue is based on these blocks. Details of the location of blocks are shown in Fig. 3 (Zhang et al., 2014a; Wang, Zhang, Guo, & Zhang, 2013).

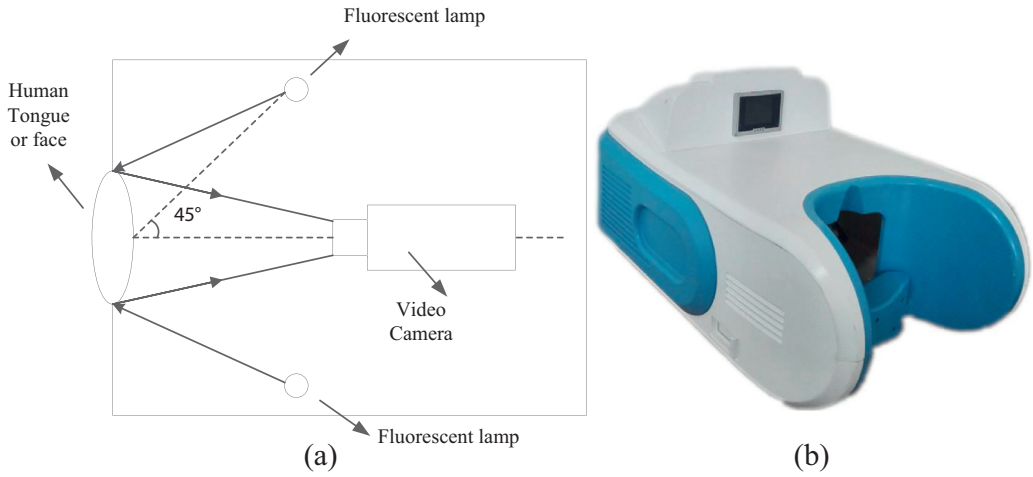

Fig. 2. Image capture device. (a) Viewing geometry and imaging path of the imaging device. (b) Appearance of the device and the system.

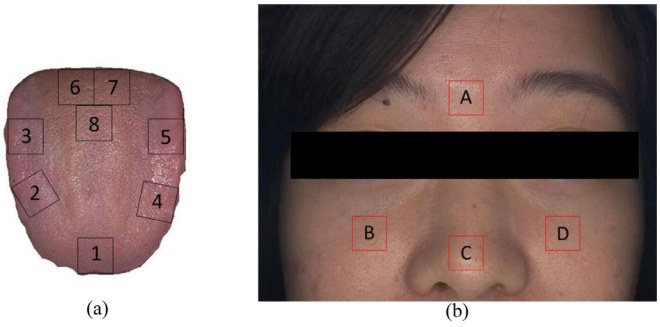

Fig. 3. The located blocks in tongue and facial images.

2.2.1. Color feature

The RGB values of captured tongue or facial images are first calculated and then converted to CIEXYZ

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

followed by CIEXYZ to CIELAB (Zhang, Wang, Jin, & Zhang, 2005)

$$\begin{aligned} L &= 166 f(Y/Y_0) - 16 \\ a &= 500[f(X/X_0) - f(Y/Y_0)] \\ b &= 200[f(Y/Y_0) - f(Z/Z_0)] \end{aligned} \tag{2}$$

where X_0, Y_0 and Z_0 are the CIEXYZ tristimulus values of the reference white point; $f(x) = x^{1/3}$ if $x > 0.008856$ or $f(x) = 7.787x + 16/116$ if $x \le 0.008856$.

Refer to our previous work (Zhang et al., 2014a; 2014b), 12 colors centroids for the tongue and 6 colors centroids for the face are extracted, respectively. When an image is captured, its corresponding LAB values are calculated following Eqs. (1) and (2). Each pixel is compared with the predefined colors to assign the closest one based on the l_2 norm or Euclidean distance. Then the ratio of each color with respect to the total number of image pixels is computed, which is regarded as the color feature. Therefore, a 12 dimensional tongue color feature is acquired. Similarly, a 6 dimensional face color feature for each block is obtained.

2.2.2. Texture feature

In Zhang's (Zhang et al., 2013) and Shu's (Shu & Zhang, 2015) study, the texture feature has been proved to be effective in disease diagnosis. In this subsection, we apply the 2-D Gabor filter to compute the corresponding value for each block.

Firstly, the 2-D Gabor filter is defined as:

$$G_k(x, y) = \exp\left(\frac{x'^2 + \gamma^2 y'^2}{-2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda}\right) \tag{3}$$

where $x' = x\cos\theta + y\sin\theta$, $y' = -x\sin\theta + y$, θ is the orientation, γ is the aspect ratio of the sinusoidal function, σ is the variance, and λ is the wavelength. In our system, four θ (0, 45, 90 and 135) and three σ (1, 2 and 3) choices are selected to get the best result. A response $R_k(x, y)$ is generated through convolving each filter with a texture block.

$$R_k(x, y) = G_k(x, y) * im(x, y) \tag{4}$$

where $*$ means 2-D convolution operation and the function $im(x, y)$ denotes the texture block. The maximum pixel intensity is then obtained according to $FR(x, y) = \max(R_1(x, y), \cdots, R_n(x, y))$, which is regarded as the texture feature. Meanwhile, the mean of all blocks for the tongue or facial image is also appended to be the last value of the texture feature. Totally, a 9 dimensional tongue texture feature and a 5 dimensional face texture feature are acquired, respectively.

2.2.3. Geometry feature

Statistically, the changes in the geometric shape of the tongue can also reveal some valuable information for diagnosis. In our system, 13 geometry features containing smaller half-distance, length, width, length-width ratio, area, center distance, circle area, triangle area, square area, center distance ratio, circle area ratio, triangle area ratio, and square area ratio (Zhang et al., 2014a) are proposed.

3. Multi-modal joint discriminative and collaborative representation

In this section, we introduce a novel multi-model fusion approach called joint discriminative and collaborative representation (JDCR). As the proposed method is based on the collaborative representation, we first briefly review the collaborative representation based classification (CRC).

3.1. Collaborative representation based classification

Recently, sparse representation (Wright, Yang, Ganesh, Sastry, & Ma, 2009; Zhang et al., 2016; Zhang et al., 2017; Shi, Wan, Wu, & Chen, 2017) has achieved a great success in many applications, such as face recognition (Wright et al., 2009), object decomposition (Peng, Ganesh, Wright, Xu, & Ma, 2012) and image restoration (Dong, Zhang, Shi, & Li, 2013), partially due to the progress

of l_0-norm and l_1-norm minimization techniques (Tibshirani, 1996; Donoho, 2006; Tropp & Wright, 2010). However, the minimization of l_1-norm in the sparse representation based classification would be very time consuming which greatly limits its application in practice. Fortunately, Zhang, Yang, and Feng (2011) proposed a simple collaborative representation based classification (CRC) scheme which not only reduces the computational complexity but also guarantees the effectiveness in object recognition. Given a test sample $\mathbf{y}$, it can be represented as follows by using CRC:

$$\hat{\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2 \tag{5}$$

where the matrix $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_J] \in \mathbb{R}^{m\times n}$ is constructed by training samples, J denotes the total number of categories, n is number of training samples, $\mathbf{D}_i \in \mathbb{R}^{m\times n_i}$ means the n_i number of training samples belonging to the ith class with m dimension of the input feature, and the λ is the regularization parameter to avoid over-fitting. Different from sparse representation using l_1-norm, collaborative representation utilizes the regularized least square: l_2-norm, which can be easily and analytically derived as

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{D}^T\mathbf{D} + \lambda\cdot\mathbf{I}\right)^{-1}\mathbf{D}^T\mathbf{y} \tag{6}$$

w.r.t. to the representation $\boldsymbol{\alpha}$. Note that $\mathbf{P} = \left(\mathbf{D}^T\mathbf{D} + \lambda\cdot\mathbf{I}\right)^{-1}\mathbf{D}^T$ is independent to the test sample $\mathbf{y}$, thus it can be obtained previously to be a projection matrix. In other words, the process of collaborative representation is very fast. More information about the collaborative representation can be found in Zhang's study (Zhang et al., 2011).

3.2. Joint discriminative and collaborative representation model

As mentioned in Section 1, different features from different modalities may have a certain correlation. With the multi-modal representation, however, how to exploit the rich structural information of each modality is a challenging task. Motivated by Yang, Zhang, Zhang, and Wang (2012), the following term is utilized to measure the similarity of various modalities.

$$\min_{\boldsymbol{\alpha}} \sum_{k=1}^{K} \left\|\mathbf{y}^k - \mathbf{D}^k\boldsymbol{\alpha}\right\|_2^2 \tag{7}$$

where $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1; \boldsymbol{\alpha}_2; \cdots; \boldsymbol{\alpha}_J]$ is the shared representation coefficient and $\boldsymbol{\alpha}_i$ is the partial vector associated with the ith category; $\mathbf{D}^k \in \mathbb{R}^{m_k\times n}$ is the training set (dictionary) corresponding to kth modality and m_k is the dimension of the instance belonging to kth modality. The features from different modalities are different both in scale and types. Thus directly analyzing the similarity among them is unreasonable. However, we can assume that all the modalities are a projection from a latent subspace. In this latent subspace, the distance of each modality is close to each other. Particularly, we can assume that a modality $\mathbf{y}^k$ of a test sample $\mathbf{y}$ is protected by its corresponding dictionary $\mathbf{D}^k$: $\mathbf{y}^k = \mathbf{D}^k\boldsymbol{\alpha}$. In this case, the dictionary $\mathbf{D}^k$ can be regarded as the mapping matrix which projects the latent and shared variable into each modality. In practice, a test sample consisting of the tongue and facial features can be represented linearly as a combination of the training samples: since all tasks are belonging to a same sample, they have a great probability to be represented by the training samples of each task on the same position and hence values of significant coefficients $\boldsymbol{\alpha}$ are similar or same.

Although the correlation among different tasks are extracted, it is still not preferable for the classification since the aforementioned method is unsupervised and does not use the label information for training. In order to enlarge the variance of different classes, we embed a discriminative regularization (Xu et al., 2016) into our model. The main formula of the discriminative regularization is shown as follows,

$$\left\|\mathbf{D}_i^k\boldsymbol{\alpha}_i + \mathbf{D}_j^k\boldsymbol{\alpha}_j\right\|_2^2 = \left\|\mathbf{D}_i^k\boldsymbol{\alpha}_i\right\|_2^2 + \left\|\mathbf{D}_j^k\boldsymbol{\alpha}_j\right\|_2^2 + 2tr((\mathbf{D}_i^k\boldsymbol{\alpha}_i)^T\mathbf{D}_j^k\boldsymbol{\alpha}_j) \tag{8}$$

where $\mathbf{D}_i^k$ denotes the training samples belonging to the ith category in the kth task. It is easy to see that in Eq. (8) the ith and jth category have the minimal correlation (maximize difference) by minimizing the term $(\mathbf{D}_i^k\boldsymbol{\alpha}_i)^T\mathbf{D}_j^k\boldsymbol{\alpha}_j$. Additionally, the function (8) is convex and differentiable, and closed-form solution w.r.t. the coefficient $\boldsymbol{\alpha}$ can be computed which greatly reduces the computational complexity.

Therefore, by combining the correlation formula Eq. (7) and the discriminative regularization Eq. (8), the objective function of our model can be obtained as follows:

$$\begin{aligned}\min_{\boldsymbol{\alpha}} &\sum_{k=1}^{K}\left\{\left\|\mathbf{y}^k - \mathbf{D}^k\boldsymbol{\alpha}\right\|_2^2\right\} + \lambda\|\boldsymbol{\alpha}\|_2^2 \\ &+ \sum_{k=1}^{K}\left\{\gamma\sum_{\substack{i=1\\ i\neq j}}^{J}\sum_{\substack{j=1\\ j\neq i}}^{J}\left\|\mathbf{D}_i^k\boldsymbol{\alpha}_i + \mathbf{D}_j^k\boldsymbol{\alpha}_j\right\|_2^2\right\}\end{aligned} \tag{9}$$

where $\mathbf{D}^k = \left[\mathbf{D}_1^k, \mathbf{D}_2^k, \cdot, \cdot, \cdot, \mathbf{D}_J^k\right]$, and λ and γ are the positive parameters that are used to trade-off the influence of three terms in Eq. (9). As we can see, our proposed model not only extracts the shared representation coefficient $\boldsymbol{\alpha}$ to exploit the correlation among different modalities, but also maximizes the difference between different classes by introducing the discriminative regularization.

The framework of JDCR is displayed in Fig. 4. At the second step, the shared representation coefficient is extracted. Meanwhile, since the discriminative information of different classes is utilized, we can see that most elements of the coefficients with large value would concentrate on the first class (assume that the test sample belongs to the first class). Thus its reconstruction error corresponding to the first class would be lower than that in other categories, which means the test sample has high probability of belonging to the first category.

3.3. Optimization of JDCR

We denote the objective function as

$$L = \sum_{k=1}^{K}\left\|\mathbf{y}^k - \mathbf{D}^k\boldsymbol{\alpha}\right\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2 + \sum_{k=1}^{K}\gamma\sum_{i=1,i\neq j}^{J}\sum_{j=1,j\neq i}^{J}\left\|\mathbf{D}_i^k\boldsymbol{\alpha}_i + \mathbf{D}_j^k\boldsymbol{\alpha}_j\right\|_2^2 \tag{10}$$

We then divide it into two parts:

$$\begin{aligned}L_1 &= \sum_{k=1}^{K}\left\|\mathbf{y}^k - \mathbf{D}^k\boldsymbol{\alpha}\right\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2 \\ L_2 &= \sum_{k=1}^{K}\gamma\sum_{i=1,i\neq j}^{J}\sum_{j=1,j\neq i}^{J}\left\|\mathbf{D}_i^k\boldsymbol{\alpha}_i + \mathbf{D}_j^k\boldsymbol{\alpha}_j\right\|_2^2\end{aligned} \tag{11}$$

Derivative of L_1 w.r.t $\boldsymbol{\alpha}$: It is easy to get that the derivative of L_1 w.r.t $\boldsymbol{\alpha}$ is

$$\frac{\partial L_1}{\partial\boldsymbol{\alpha}} = -2\sum_{k=1}^{K}(\mathbf{D}^k)^T(\mathbf{y}^k - \mathbf{D}^k\boldsymbol{\alpha}) + 2\lambda\boldsymbol{\alpha} \tag{12}$$

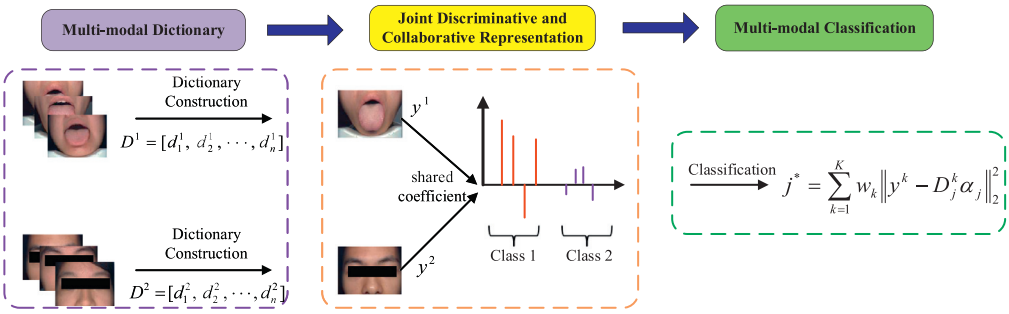

Fig. 4. The framework of the Joint Discriminative and Collaborative Representation model (JDCR). JDRC consists of three steps which are multi-modal dictionary construction, joint discriminative and collaborative representation, and the reconstruction error based classification. In the first step, training samples belonging to different classes construct the dictionary; multi-modal features extracted from the given test instance are then represented by the dictionary, and the shared and discriminative representation coefficient is obtained; thirdly, the label is output by comparing reconstruction errors of each class.

Derivative of L_2 w.r.t $\boldsymbol{\alpha}$: Factorize L_2, we can get

$$\begin{aligned} L_2 &= \gamma \sum_{k=1}^{K} \left(\sum_{j \neq t} \left\| \mathbf{D}_t^k \boldsymbol{\alpha}_t + \mathbf{D}_j^k \boldsymbol{\alpha}_j \right\|_2^2 + \sum_{i \neq t} \left\| \mathbf{D}_i^k \boldsymbol{\alpha}_i + \mathbf{D}_t^k \boldsymbol{\alpha}_t \right\|_2^2 + F \right) \\ &= \gamma \sum_{k=1}^{K} \left(2 \sum_{i \neq t} \left\| \mathbf{D}_t^k \boldsymbol{\alpha}_t + \mathbf{D}_i^k \boldsymbol{\alpha}_i \right\|_2^2 + F \right) \end{aligned} \tag{13}$$

where F is unrelated to $\boldsymbol{\alpha}_t$. Thus, the derivative of L_2 w.r.t. $\boldsymbol{\alpha}_t$ is

$$\begin{aligned} \frac{\partial L_2}{\partial \boldsymbol{\alpha}_t} &= \gamma \sum_{k=1}^{K} \left(4 (\mathbf{D}_t^k)^T \sum_{i \neq t} (\mathbf{D}_t^k \boldsymbol{\alpha}_t + \mathbf{D}_i^k \boldsymbol{\alpha}_i) \right) \\ &= 4\gamma \sum_{k=1}^{K} \left((\mathbf{D}_t^k)^T \left((J-1) \mathbf{D}_t^k \boldsymbol{\alpha}_t + \sum_{i \neq t} \mathbf{D}_i^k \boldsymbol{\alpha}_i \right) \right) \\ &= 4\gamma \sum_{k=1}^{K} (\mathbf{D}_t^k)^T \left((J-2) \mathbf{D}_t^k \boldsymbol{\alpha}_t + \sum_{i=1}^{J} \mathbf{D}_i^k \boldsymbol{\alpha}_i \right) \\ &= 4\gamma \sum_{k=1}^{K} (\mathbf{D}_t^k)^T \left((J-2) \mathbf{D}_t^k \boldsymbol{\alpha}_t + \mathbf{D}^k \boldsymbol{\alpha} \right) \end{aligned} \tag{14}$$

Therefore, the derivative of L_2 w.r.t $\boldsymbol{\alpha}$ is obtained as follows.

$$\begin{aligned} \frac{\partial L_2}{\partial \boldsymbol{\alpha}} &= \begin{bmatrix} 4\gamma \sum_{k=1}^{K} (\mathbf{D}_1^k)^T ((J-2) \mathbf{D}_1^k \boldsymbol{\alpha}_1 + \mathbf{D}^k \boldsymbol{\alpha}) \\ 4\gamma \sum_{k=1}^{K} (\mathbf{D}_2^k)^T ((J-2) \mathbf{D}_2^k \boldsymbol{\alpha}_2 + \mathbf{D}^k \boldsymbol{\alpha}) \\ \vdots \\ 4\gamma \sum_{k=1}^{K} (\mathbf{D}_J^k)^T ((J-2) \mathbf{D}_J^k \boldsymbol{\alpha}_J + \mathbf{D}^k \boldsymbol{\alpha}) \end{bmatrix} \\ &= 4\gamma (J-2) \begin{bmatrix} \sum_{k=1}^{K} (\mathbf{D}_1^k)^T \mathbf{D}_1^k \boldsymbol{\alpha}_1 \\ \sum_{k=1}^{K} (\mathbf{D}_2^k)^T \mathbf{D}_2^k \boldsymbol{\alpha}_2 \\ \vdots \\ \sum_{k=1}^{K} (\mathbf{D}_J^k)^T \mathbf{D}_J^k \boldsymbol{\alpha}_J \end{bmatrix} + 4\gamma \begin{bmatrix} \sum_{k=1}^{K} (\mathbf{D}_1^k)^T \mathbf{D}^k \\ \sum_{k=1}^{K} (\mathbf{D}_2^k)^T \mathbf{D}^k \\ \vdots \\ \sum_{k=1}^{K} (\mathbf{D}_J^k)^T \mathbf{D}^k \end{bmatrix} \boldsymbol{\alpha} \\ &= 4\gamma (J-2) \begin{bmatrix} \sum_{k=1}^{K} (\mathbf{D}_1^k)^T \mathbf{D}_1^k \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots \sum_{k=1}^{K} (\mathbf{D}_J^k)^T \mathbf{D}_J^k \end{bmatrix} \boldsymbol{\alpha} + 4\gamma \mathbf{M}_2 \boldsymbol{\alpha} \\ &= 4\gamma ((J-2) \mathbf{M}_1 + \mathbf{M}_2) \boldsymbol{\alpha} \end{aligned} \tag{15}$$

where

$$\mathbf{M}_1 = \begin{bmatrix} \sum_{k=1}^{K} (\mathbf{D}_1^k)^T \mathbf{D}_1^k \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots \sum_{k=1}^{K} (\mathbf{D}_J^k)^T \mathbf{D}_J^k \end{bmatrix}$$

and

$$\mathbf{M}_2 = \begin{bmatrix} \sum_{k=1}^{K} (\mathbf{D}_1^k)^T \mathbf{D}^k \\ \sum_{k=1}^{K} (\mathbf{D}_2^k)^T \mathbf{D}^k \\ \vdots \\ \sum_{k=1}^{K} (\mathbf{D}_J^k)^T \mathbf{D}^k \end{bmatrix}.$$

In conclusion, the total derivative of the objective function L w.r.t. $\boldsymbol{\alpha}$ is

$$\frac{\partial L}{\partial \boldsymbol{\alpha}} = -2 \sum_{k=1}^{K} (\mathbf{D}^k)^T (\mathbf{y}^k - \mathbf{D}^k \boldsymbol{\alpha}) + 2\lambda \boldsymbol{\alpha} + 4\gamma ((J-2) \mathbf{M}_1 + \mathbf{M}_2) \boldsymbol{\alpha} \tag{16}$$

Finally the optimal solution of Eq. (9) is

$$\boldsymbol{\alpha} = (2\gamma ((J-2) \mathbf{M}_1 + \mathbf{M}_2) + \lambda \mathbf{I} + \sum_{k=1}^{K} (\mathbf{D}^k)^T \mathbf{D}^k)^{-1} \sum_{k=1}^{K} (\mathbf{D}^k)^T \mathbf{y}^k \tag{17}$$

where $\mathbf{I}$ is the identity matrix. Let $\mathbf{P} = (2\gamma ((J-2) \mathbf{M}_1 + \mathbf{M}_2) + \lambda \mathbf{I} + \sum_{k=1}^{K} (\mathbf{D}^k)^T \mathbf{D}^k)^{-1}$. Note that $\mathbf{P}$ is independent of the testing sample $\mathbf{y} = [\mathbf{y}^1, \cdots, \mathbf{y}^K]$, thus we can pre-calculate it as a projection matrix. Given a sample including K tasks, we only need to calculate $\sum_{k=1}^{K} (\mathbf{D}^k)^T \mathbf{y}^k$ and then simply project it onto $\mathbf{P}$. In practice, this makes our proposed method JDCR very fast.

3.4. The classification rule of JDCR

After linearly representing the test sample and acquiring the corresponding coefficients, the label is decided according to the lowest total reconstruction error over K modalities.

$$j^* = \min_j \sum_{k=1}^{K} w_k \left\| \mathbf{y}^k - \mathbf{D}_j^k \boldsymbol{\alpha}_j \right\|_2^2 \tag{18}$$

where $\mathbf{D}_j^k$ and $\boldsymbol{\alpha}_j$ are the parts of the dictionary $\mathbf{D}^k$ and the shared coefficient corresponding to the jth class, respectively; w_k denotes the weight value associated with the kth modality. In this paper, we calculate it by using the strategy mentioned in Yuan's study (Yuan, Liu, & Yan, 2012). The detail steps of JDCR algorithm are illustrated in Algorithm 1. Note that, values of two parameters λ and γ are set through the cross validation.

4. Experimental results

4.1. Image dataset

The tongue and facial dataset consists of 961 samples containing 500 Healthy samples and 461 Fatty Liver samples. In this

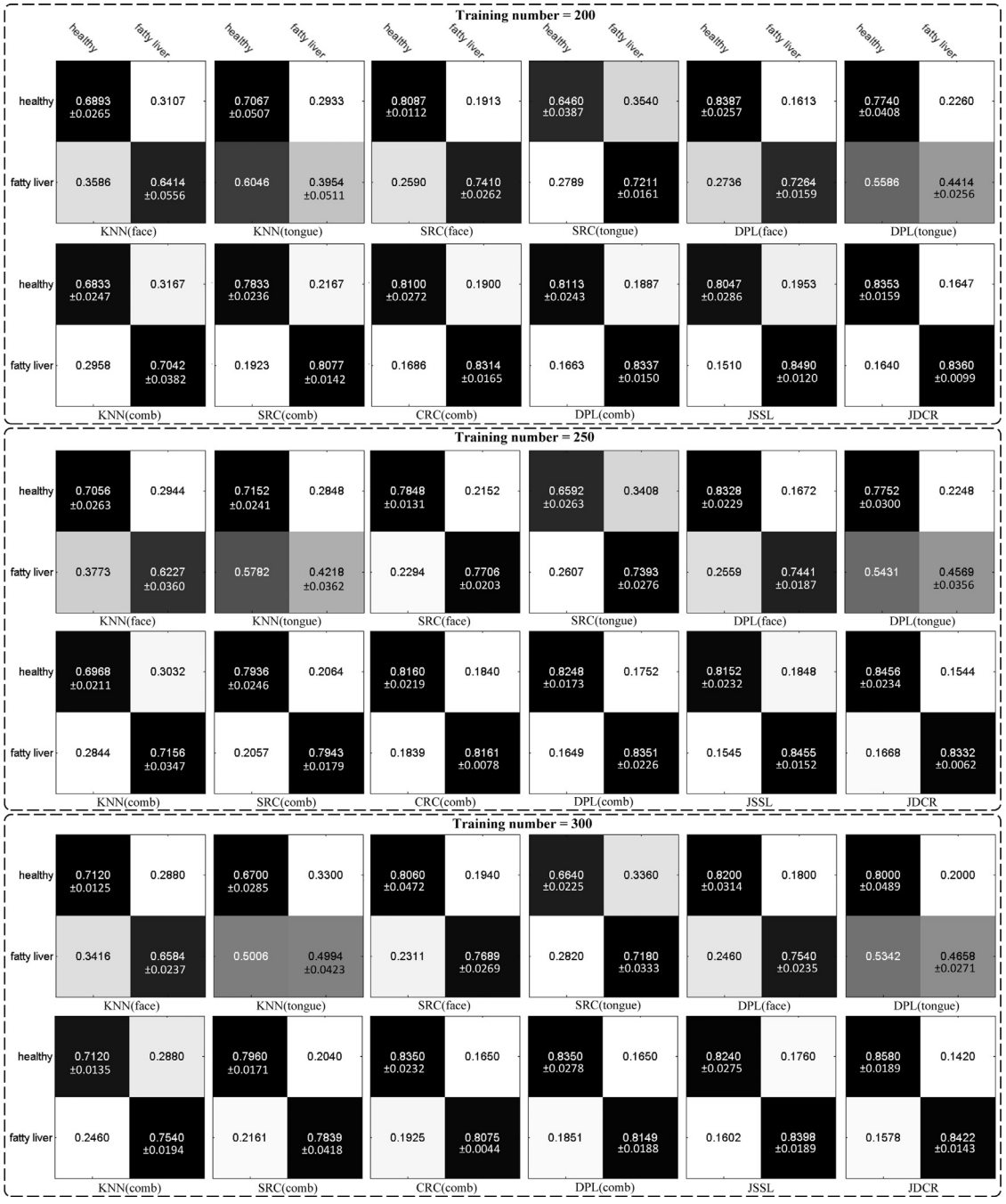

Fig. 5. The confusion matrices in 5 independent experiments for Fatty Liver detection. The top left, top right, bottom left and bottom right corners are sensitivity, false negative rate, false positive rate and specificity, respectively, followed by their corresponding error bars.

dataset, each sample has two different types of images: tongue image and facial image, respectively. These images were collected at the Guangdong Provincial TCM Hospital, Guangdong, China, from the early 2014 to the late 2015. In the sampling process, people who were healthy were verified through a blood test and other examinations. According to the standardized rule designed by the Guangdong Provincial TCM Hospital, people were verified to be healthy if indicators from the tests fall within a certain range. The type-B ultrasonic (CT), liver function, and blood fat tests was exploited to decide whether the person was suffering from the Fatty Liver. If and only if the results of these three examinations are uncommon, the person is detected with the Fatty Liver.

4.2. Experiments set

An experiment for Healthy versus Fatty Liver Disease classification is conducted in this section. In the experiment, KNN (Cover & Hart, 1967), sparse representation based classification (SRC) (Wright et al., 2009), CRC (Zhang et al., 2011), dictionary pair learning (DPL) (Gu, Zhang, Zuo, & Feng, 2014), as well as the multi-view method-joint similar and specific learning (JSSL) (Li, Zhang, Li, Wu, & Zhang, 2017) are applied for the single-modal or multi-modal classification. Without loss of generality, different features of the tongue or facial modalities are concatenated as a single one, and we denote it as tongue feature or facial feature. Since KNN, SRC, CRC and DPL are designed for individual task,

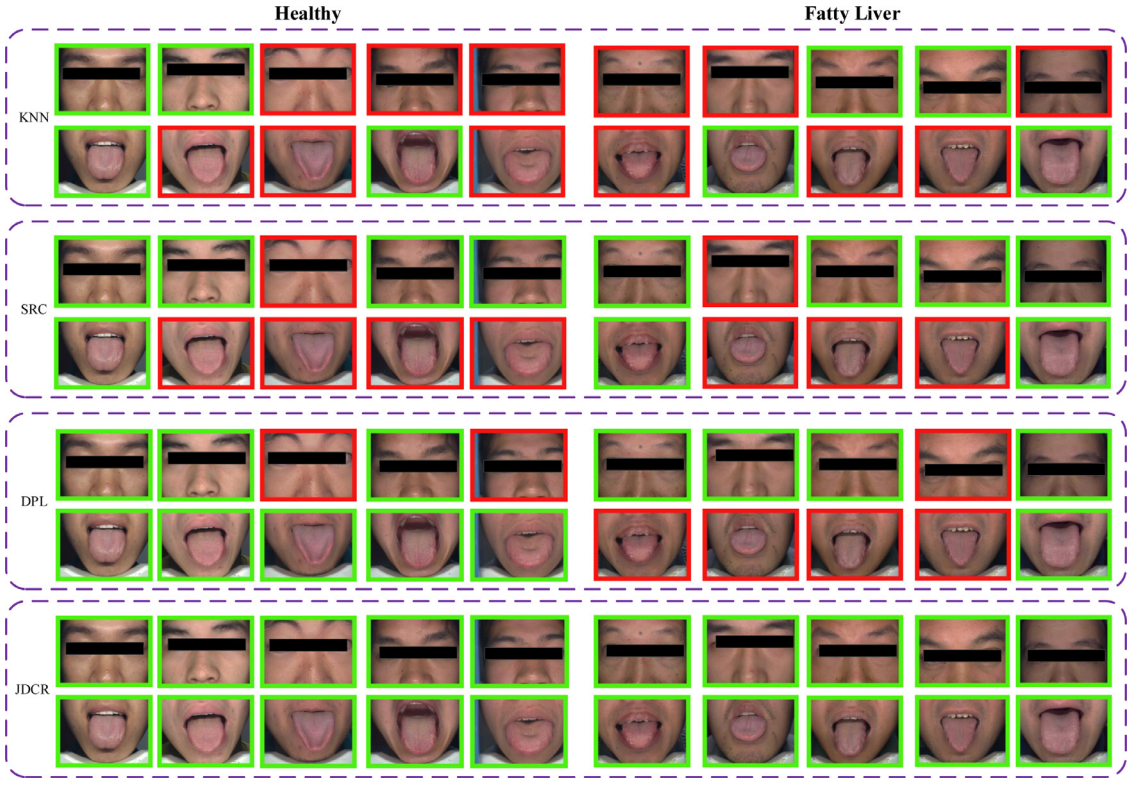

Fig. 6. Some typical instances classified by using different classifiers. The correct classification is described by the green border and the incorrect classification is denoted by the red border. The classified instances based on facial images are shown in the first row, and the classified instances based on tongue images is displayed in the second row. The left part denotes the 'healthy' samples and the right part denotes the 'Fatty Liver' samples. KNN, SRC and DPL fail in some cases. In contrast, JDCR can obtain an obvious improvement. Particularly, even though some instances cannot be detected by the strategies with the individual modality, JDCR is able to diagnose. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

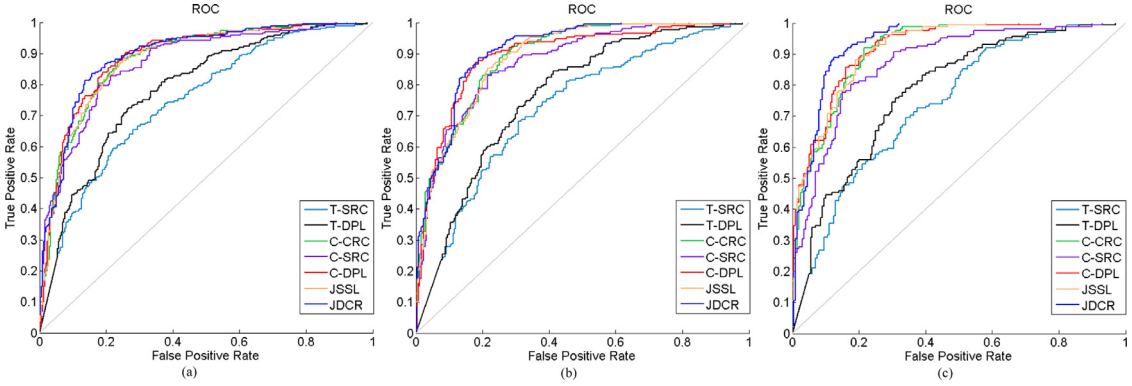

Fig. 7. The ROC curves of various strategies and different features for the Fatty Liver disease detection when the number of training data is 200, 250, and 300 in each category. T-SRC and C-SRC mean the curves are obtained by SRC(tongue) and SRC(Comb), respectively, which are similar to T-DPL and C-DPL.

Algorithm 1 Joint discriminative and collaborative representation (JDCR).

Input: λ, γ, $\mathbf{y} = \{\mathbf{y}^k\}$, $\mathbf{D} = \{\mathbf{D}^k\}$, $k = 1, \cdots, K$
Output: d_j, t, and $\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_J]$
Initialization: $\mathbf{P}$, $\sum_{k=1}^{K} (\mathbf{D}^k)^T \mathbf{y}^k$,
1: **$\boldsymbol{\alpha}$ is calculated using:**
$\boldsymbol{\alpha} = \mathbf{P} \sum_{k=1}^{K} (\mathbf{D}^k)^T \mathbf{y}^k$
2: **The distance between the test sample and the jth class isobtained using class-specific residual:**
$d_j = \sum_{k=1}^{K} w_k \left\| \mathbf{y}^k - \mathbf{D}_j^k \boldsymbol{\alpha}_j \right\|_2^2$
3: **The test sample is classified to the tth category:**
if $t = \arg\min_j d_j$

we also concatenate all features from two tasks into a single one (denoted as 'comb') to have a comparison with our proposed method. For instance, there are color, texture and geometric features denoted as x_t^c, x_t^t, and x_t^g for the tongue image, respectively. The tongue feature is then denoted as $\mathbf{x}_t = [x_t^c; x_t^t; x_t^g] \in \mathbb{R}^{d_t \times 1}$, where d_t is its corresponding dimensionality. Similar, the facial feature can be defined as $\mathbf{x}_f = [\mathbf{x}_f^c; \mathbf{x}_f^t] \in \mathbb{R}^{d_f \times 1}$, where $\mathbf{x}_f^c$, $\mathbf{x}_f^t$ and d_f are its color feature, texture feature and dimensionality. The combined vector is then generated through $\mathbf{x}_{comb} = [\mathbf{x}_t; \mathbf{x}_f]$.

In order to quantitatively evaluate our presented approach, we apply the Sensitivity, Specificity, False Positive Rate (FPR), False Negative Rate (FNR) and Accuracy to measure the performance in the experiment. The Sensitivity is an index to evaluate the probability of positives are correctly classified; the Specificity indicates

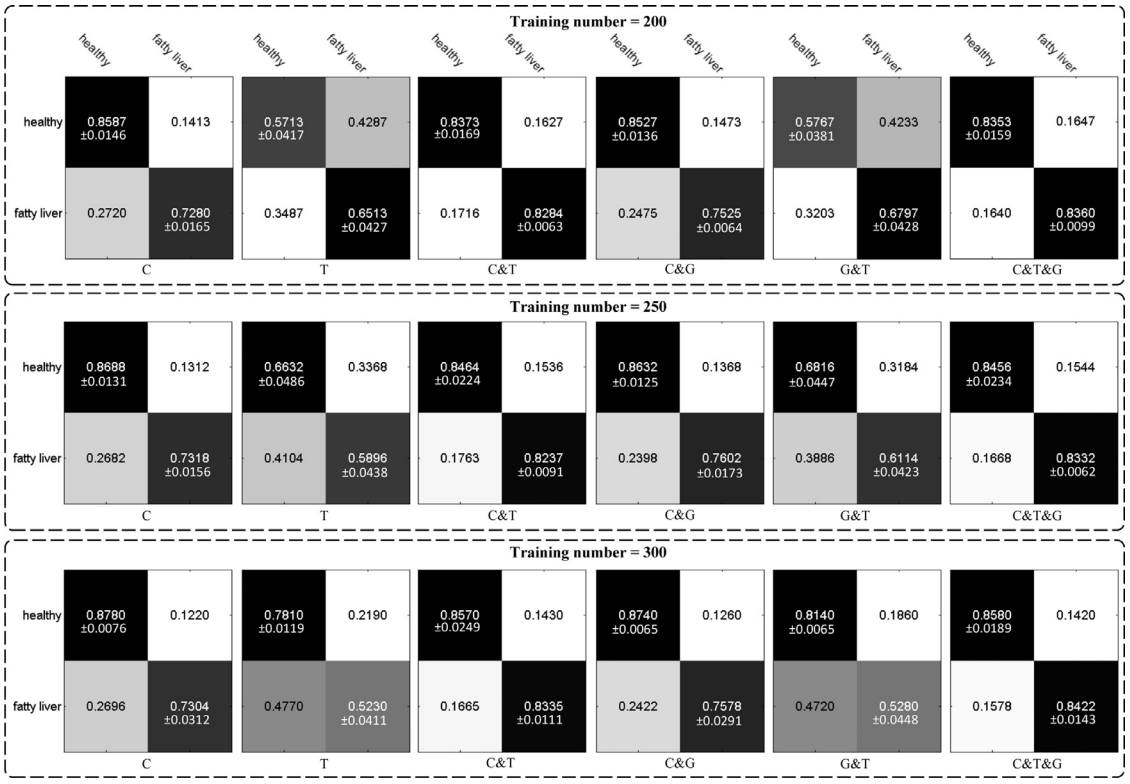

Fig. 8. The confusion matrices of the contributions of different features in the Fatty Liver detection experiments. The top left, top right, bottom left and bottom right corners are sensitivity, false negative rate, false positive rate and specificity, respectively, followed by their corresponding error bars.

the rate that negatives are correctly detected; the FPR refers to a positive result corresponds to rejecting the null hypothesis; FNR is a negative result corresponds to not rejecting the null hypothesis; the Accuracy measures the proportion of both positives and negatives that are correctly identified. They are mathematically defined as follows:

$$\begin{aligned} Sensitivity &= \frac{TruePos.}{TruePos. + FalseNeg.} \\ Specificity &= \frac{TrueNeg.}{TrueNeg. + FalsePos.} \\ FPR &= \frac{FalsePos.}{FalsePos. + TrueNeg.} \\ FNR &= \frac{FalseNeg.}{TruePos. + FalseNeg.} \\ Accuracy &= \frac{TruePos. + TrueNeg.}{TruePos. + FalsePos. + TrueNeg. + FalseNeg.} \end{aligned} \tag{19}$$

In addition, we conduct another experiment focusing on the contribution of different features and display time consumption in the subsection 4.4 and 4.5.

4.3. Healthy versus fatty liver classification

The evaluation of the proposed JDCR in detection of Fatty Liver compared with other methods is shown in this subsection. Some samples are randomly selected as training samples, and the remaining instances are applied for testing. To quantitatively and scientifically evaluate the strategies, we conduct 5 times and show the average accuracy and error bar.

Table 1 and Fig. 5 demonstrate the average accuracy, sensitivity, specificity, false positive and false negative classification rates on 5 independent experiments obtained by our presented approach JDCR and other comparison strategies. In each experiment, we randomly select 200, 250 and 300 samples of each categories from the original dataset as training samples, and the remaining samples are exploited for testing. As we can see that, compared with the single task based methods, our approach achieves an obvious improvement. In particular, when the training number is 200, JDCR gets 83.57% in accuracy while the best value obtained by face based method (DPL) only reaches 78.65% which is far below than ours'. In terms of sensitivity and specificity, JDCR also outperforms face or tongue based strategies. In contrast to the combination methods, our proposed algorithm also has a better results in average accuracy and sensitivity. Besides, with the increasing number of training samples, the average accuracy obtained by our proposed method has a remarkable improvement, which rises from 83.99% to 85.10%. Simultaneously, the values of sensitivity and specificity also increase gradually. Compared with the face or tongue task based methods, like KNN, SRC and DPL, JDCR reaches a higher average accuracy, when the training number is 250 and 300. In contrast to the combination based approaches, our method is also competitive. Although DPL (combination) gains higher values in specificity (when the training number is 250), its values of sensitivity are much lower than that acquired by JDCR. In comparison to JSSL, the presented method also achieves a better performance, although the specificity value obtained by JDCR is slightly inferior to that obtained by JSSL when the number of training samples is 200 and 250. Theoretically, JSSL aims to learn the common and distinctive components among various modalities by introducing the l_1 norm. However, similar to the traditional sparse representation methods, JSSL does not take the discriminant information across different categories into account, resulting in an inferior classification performance compared with JDCR in most cases. Additionally, there is no closed-form solution for this algorithm and a large number of iterations is required at the optimization step, which leads to be time consuming, as shown in Table 4. By contrast, the proposed method not only exploits the category-

Table 1
The classification rate and their corresponding error bar in 5 independent experiments for Fatty Liver detection. Best results are highlighted in bold.

	Number of training samples				Number of training samples		
Methods	200	250	300	Methods	200	250	300
KNN(face)	66.70% ± 2.38%	66.77% ± 2.29%	68.81% ± 1.01%	KNN(comb)	69.31% ± 2.15%	70.54% ± 1.95%	73.07% ± 1.31%
KNN(tongue)	56.19% ± 0.57%	58.09% ± 1.91%	59.39% ± 1.07%	SRC(comb)	79.47% ± 1.05%	79.39% ± 1.97%	79.06% ± 2.13%
SRC(face)	77.72% ± 1.15%	77.83% ± 1.36%	78.95% ± 1.96%	CRC(comb)	82.00% ± 1.53%	81.61% ± 1.18%	82.27% ± 1.34%
SRC(tongue)	68.09% ± 1.55%	69.59% ± 2.34%	68.81% ± 1.23%	DPL(comb)	82.17% ± 1.09%	82.95% ± 1.28%	82.60% ± 1.67%
DPL(face)	78.65% ± 1.35%	79.22% ± 1.20%	79.06% ± 1.51%	JSSL	82.53% ± 1.61%	82.91% ± 1.45%	83.10% ± 1.45%
DPL(tongue)	61.93% ± 1.65%	62.95% ± 1.85%	65.10% ± 2.42%	JDCR	**83.57% ± 0.75%**	**83.99% ± 1.25%**	**85.10% ± 1.32%**

Table 2
The AUC values on the Fatty Liver diagnosis when the number of training instances is 200, 250, and 300, respectively. Best results are highlighted in bold.

	Number of training samples		
Num	num=200	num=250	num=300
Methods	AUC	AUC	AUC
SRC(tongue)	0.7536	0.7345	0.7440
DPL(tongue)	0.7912	0.7740	0.7838
CRC(Comb)	0.8825	0.8923	0.9149
SRC(Comb)	0.8690	0.8740	0.8726
DPL(Comb)	0.8882	0.8947	0.9197
JSSL	0.8823	0.8900	0.9155
JDCR	**0.8940**	**0.9112**	**0.9363**

Table 3
The contributions of different features on the classification accuracy in Fatty Liver detection experiments.

	Number of training samples		
Features	num=200	num=250	num=300
C	79.79 ± 1.19%	80.61 ± 1.17%	81.22 ± 1.38%
T	60.86 ± 1.36%	62.95 ± 1.98%	66.59 ± 1.59%
C&T	83.32 ± 0.82%	83.60 ± 1.07%	84.65 ± 1.60%
C&G	80.61 ± 0.67%	81.61 ± 1.10%	82.22 ± 1.40%
G&T	62.46 ± 2.27%	64.59 ± 1.83%	68.64 ± 2.20%
C&T&G	83.57 ± 0.75%	83.99 ± 1.25	85.10 ± 1.32%

Table 4
The time consumption using the proposed method with different training data numbers.

Training data	200	250	300
JSSL	22.4168 (s)	32.5025 (s)	39.5441 (s)
JDCR	0.0346 (s)	0.0407 (s)	0.0600 (s)

based information, which is adaptive for the classification, but also has a closed-form solution to get the representation coefficients.

From Fig. 5 we also can see that the proposed method achieves lower FPR and FNR values compared with the single task based methods. In particular, when the training number is 300, JDCR obtains 15.78% and 14.20% in FPR and FNR, respectively, while the best results calculated by the face based method (DPL) reach 24.60% and 18.00%, which are much higher than ours'. In comparison to the combination methods, the proposed algorithm also gains an remarkable improvement. In contrast to the feature fusion approach JSSL, JDCR is superior in the two metrics. For instance, JDCR achieves at least a 3% decrease in FNR compared with that obtained by JSSL. Referring to the error bar, our method gets lower values at most time in comparison to other methods, which indicates that our method is more stable and suitable for the practical application.

Some selected examples of tongue and facial images of classification on Fatty Liver disease and Healthy are provided in Fig. 6, when the number of training samples is 200. It is easy to see that the individual task based approaches fail to classify some samples. And some samples cannot even be detected by the single modality, both tongue and face, while our proposed method JDCR does. The reason is that JDCR exploits the complementary information of tongue and facial images and minimizes the correlations among different classes, which are beneficial for diagnosis.

In addition, Fig. 7 further plots the ROC curves of various classification approaches for the Fatty Liver disease diagnosis when the number of training instances is 200, 250 and 300, respectively. Besides, their corresponding covered areas are listed in Table 2. Note that since our classification decision is based on the reconstruction error, which is similar with CRC, SRC, DPL and JSSL, the ROC and Area Under Curve (AUC) results obtained by our method are only compared with CRC, SRC, DPL and JSSL based experimental results. From the Fig. 7 we can see that the covered area obtained by the proposed method is larger than that computed by different approaches. Specially, our method JDCR outperforms other comparison methods remarkably when the training number is 300 in each class. Furthermore, from Table 2 it is also easy to observe that the JDCR is much superior to other existing methods. The area covered by JDCR is 0.8940, 0.9112 and 0.9363, while the best results among the comparison strategies are only 0.8882, 0.8947, and 0.9197, respectively.

4.4. The contribution of different features

In order to find the contribution of different features (color, texture and geometry), we conduct an experiment using the proposed method in this subsection. The experimental results are displayed in Table 3 and Fig. 8. Note that the symbols 'C', 'T' and 'G' denote the color feature, texture feature and geometric feature, respectively; the symbol 'C&T' means that we only take the color and texture features into account but remove the geometric feature, and so do 'C&G', 'G&T' and 'C&T&G'. From the Table 3 and Fig. 8 we can see that the color feature provides more valuable information for classification than that of the texture feature. Among the combination of pair features, the 'C&T' has a remarkable improvement in Fatty Liver detection, followed by 'C&G' and 'G&T'. Particularly, by combining the 'Color' and 'Texture', the proposed method achieve about 16% in both FNR and FPR. Also note that although the combination of the color feature and the geometric feature achieves a high rate in sensitivity and a low rate in FNR, its specificity (FPR) values are much lower (higher), which

means this combination may be impractical in applications. Fortunately, a total integration of the color, texture and geometric features not only gains the highest values of accuracy, but also keeps a good performance in both sensitivity (FNR) and specificity (FPR).

4.5. Time consumption

The time consumption values using JSSL and the presented approach with different number of training data are shown in Table 4. In comparison to JSSL, our proposed method is much more efficient in terms of time consumption. As mentioned above, the proposed method generates the mapping matrix $\mathbf{P}$ previously. When a test sample come, we only need to calculate $\sum_{k=1}^{K}(\mathbf{D}^k)^T\mathbf{y}^k$ and then simply project it onto $\mathbf{P}$, which greatly reduces the time consumption and obtains the result in real-time.

5. Conclusion

In this paper, an efficient multi-modal fusion approach is presented for Fatty Liver disease diagnosis. Different from existing strategies for Fatty Liver diagnosis, we exploit the valuable information of tongue and facial images. Firstly, the tongue and facial images are captured by exploiting a non-invasive capture device. The color, texture and geometric features are then extracted. In order to exploit the correlation between them, we propose a novel fusion method called joint discriminative and collaborative representation to learn a shared representation coefficient between two modal samples. Simultaneously, a discriminative regularization is added into our model to maximize the difference among different classes. The extensive experiments prove the effectiveness and superiority of the proposed method in identification of Fatty Liver from healthy controls.

Acknowledgment

This work is supported by the GRF (15224015) fund from the HKSAR Government, the central fund from Hong Kong Polytechnic University, the NSFC fund (61332011, 61272292, 61271344), Shenzhen Fundamental Research fund (JCYJ20150403161923528, JCYJ20140508160910917), and the Science and Technology Development Fund (FDCT) of Macau 124/2014/A3.

References

Amarapurkar, D. N., Hashimoto, E., Lesmana, L. A., Sollano, J. D., Chen, P.-J., & Goh, K.-L. (2007). How common is non-alcoholic fatty liver disease in the Asia-Pacific region and are there local differences? *Journal of Gastroenterology and Hepatology, 22*(6), 788–793.

Bucak, İ. Ö., & Baki, S. (2010). Diagnosis of liver disease by using cmac neural network approach. *Expert Systems with Applications, 37*(9), 6157–6164.

Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27.

Deng, M., Dahmen, U., Sun, J., Huang, H., Sehestedt, C., Homeyer, A., … Dirsch, O. (2014). Limited correlation between conventional pathologist and automatic computer-assisted quantification of hepatic steatosis due to difference between event-based and surface-based analysis. *IEEE Journal of Biomedical and Health Informatics, 18*(4), 1473–1477.

Dong, W., Zhang, L., Shi, G., & Li, X. (2013). Nonlocally centralized sparse representation for image restoration. *IEEE Transactions on Image Processing, 22*(4), 1620–1630.

Donoho, D. L. (2006). For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics, 59*(6), 797–829.

Evans, F. (1952). Needle biopsy of liver. *British Medical Journal, 1*(4770), 1249.

Gu, S., Zhang, L., Zuo, W., & Feng, X. (2014). Projective dictionary pair learning for pattern classification. In *Advances in neural information processing systems* (pp. 793–801).

Gunasundari, S., Janakiraman, S., & Meenambal, S. (2016). Velocity bounded boolean particle swarm optimization for improved feature selection in liver and kidney disease diagnosis. *Expert Systems with Applications, 56*, 28–47.

Hashimoto, E., Tokushige, K., & Ludwig, J. (2015). Diagnosis and classification of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis: Current concepts and remaining challenges. *Hepatology Research, 45*(1), 20–28.

Kim, B.-h., Lee, S.-h., Cho, D.-u., & Oh, S.-Y. (2008). A proposal of heart diseases diagnosis method using analysis of face color. In *International conference on advanced language processing and web information technology, 2008. ALPIT'08* (pp. 220–225).

Kirschbaum, B. (2000). *Atlas of Chinese tongue diagnosis*: 1. Eastland Press.

Li, J., Zhang, D., Li, Y., Wu, J., & Zhang, B. (2017). Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection. *Information Sciences, 384*, 191–204.

Liu, M., & Guo, Z. (2008). Hepatitis diagnosis using facial color image. In *Medical biometrics* (pp. 160–167). Springer.

Maciocia, G. (1989). *The foundations of Chinese medicine*. Churchill Livingstone.

Peng, Y., Ganesh, A., Wright, J., Xu, W., & Ma, Y. (2012). Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(11), 2233–2246.

Shi, Y., Wan, Y., Wu, K., & Chen, X. (2017). Non-negativity and locality constrained laplacian sparse coding for image classification. *Expert Systems with Applications, 72*, 121–129.

Shu, T., & Zhang, B. (2015). Non-invasive health status detection system using gabor filters based on facial block texture features. *Journal of Medical Systems, 39*(4), 1–8.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 267–288.

Tropp, J. A., & Wright, S. J. (2010). Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE, 98*(6), 948–958.

Wang, D., Zhang, D., & Lu, G. (2016). An optimal pulse system design by multichannel sensors fusion. *IEEE Journal of Biomedical and Health Informatics, 20*(2), 450–459.

Wang, X., Zhang, B., Guo, Z., & Zhang, D. (2013). Facial image medical analysis system using quantitative chromatic feature. *Expert Systems with Applications, 40*(9), 3738–3746.

Wang, X., Zhang, B., Yang, Z., Wang, H., & Zhang, D. (2013). Statistical analysis of tongue images for feature extraction and diagnostics. *IEEE Transactions on Image Processing, 22*(12), 5336–5347.

Wang, X., & Zhang, D. (2010). An optimized tongue image color correction scheme. *IEEE Transactions on Information Technology in Biomedicine, 14*(6), 1355–1364.

Wang, X., & Zhang, D. (2013). A new tongue colorchecker design by space representation for precise correction. *IEEE Journal of Biomedical and Health Informatics, 17*(2), 381–391.

Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2009). Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(2), 210–227.

Xu, Y., Zhong, Z., Yang, J., You, J., & Zhang, D. (2016). A new discriminative sparse representation method for robust face recognition via l2 regularization. *IEEE Transactions on Neural Networks and Learning Systems*.

Yang, M., Zhang, L., Zhang, D., & Wang, S. (2012). Relaxed collaborative representation for pattern classification. In *Ieee conference on computer vision and pattern recognition (CVPR)* (pp. 2224–2231).

Yuan, X.-T., Liu, X., & Yan, S. (2012). Visual classification with multitask joint sparse representation. *IEEE Transactions on Image Processing, 21*(10), 4349–4360.

Zhang, B., Kumar, B., & Zhang, D. (2014a). Detecting diabetes mellitus and nonproliferative diabetic retinopathy using tongue color, texture, and geometry features. *IEEE Transactions on Biomedical Engineering, 61*(2), 491–501.

Zhang, B., Kumar, B., & Zhang, D. (2014b). Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier. *IEEE Transactions on Biomedical Engineering, 61*(4), 1027–1033.

Zhang, B., Wang, X., Karray, F., Yang, Z., & Zhang, D. (2013). Computerized facial diagnosis using both color and texture features. *Information Sciences, 221*, 49–59.

Zhang, H., Li, F., Liu, P., Chen, Y., Ren, D., & Wang, K. (2017). How can a sparse representation be made applicable for very low-dimensional data? *Expert Systems with Applications, 77*, 66–70.

Zhang, H., Wang, F., Chen, Y., Zhang, W., Wang, K., & Liu, J. (2016). Sample pair based sparse representation classification for face recognition. *Expert Systems with Applications, 45*, 352–358.

Zhang, H.-Z., Wang, K.-Q., Jin, X.-S., & Zhang, D. (2005). Svr based color calibration for tongue image. In *Proceedings of 2005 international conference on machine learning and cybernetics: 8* (pp. 5065–5070).

Zhang, L., Yang, M., & Feng, X. (2011). Sparse representation or collaborative representation: Which helps face recognition? In *International conference on computer vision* (pp. 471–478).