Joint Similar and Specific Learning for Diabetes Mellitus and Impaired Glucose Regulation Detection

Jinxing Li[a], David Zhang[a,b], Yongcheng Li[b], Jian Wu[b], Bob Zhang[c]

[a] *Biometrics Research Center, Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China*
[b] *Department of Computer ScienceHarbin Institute of Technology Shenzhen graduate school, Shenzhen, China*
[c] *Department of Computer and Information Science, University of Macau, Avenida da Universidade, Taipa, Macau, China*

Abstract

Effective and accurate diagnosis of Diabetes Mellitus (DM), as well as its early stage Impaired Glucose Regulation (IGR), has attracted much attention recently. Traditional Chinese Medicine (TCM) [1], [2] etc. has proved that tongue, face and sublingual diagnosis as a noninvasive method is a reasonable way for disease detection. However, most previous works only focus on a single task (tongue, face or sublingual) for diagnosis, although different tasks may provide complementary information for the diagnosis of DM and IGR. In this paper, we propose a novel fusion method to jointly represent the tongue, face and sublingual information and discriminate between DM (or IGR) and healthy controls. Specially, the tongue, facial and sublingual images are first collected by using a non-invasive capture device. The color, texture and geometry features of these three types of images are then extracted, respectively. Finally, our so-called joint similar and specific learning (JSSL) approach is proposed to combine features of tongue, face and sublingual vein, which not only exploits the correlation but also extracts individual components among them. Experimental results on a dataset consisting of 192 Healthy, 198 DM and 114 IGR samples (all samples were obtained from Guangdong Provincial Hospital of Traditional Chinese

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

* Corresponding author
Email address: csdzhang@comp.polyu.edu.hk (David Zhang)

Medicine) substantiate the effectiveness and superiority of our proposed method for the diagnosis of DM and IGR, achieving 86.07% and 76.68% in average accuracy and 0.8842 and 0.8278 in area under the ROC curves, respectively. The source code can be found in https://github.com/sasky1/JSSLreleased.

Keywords: Diabetes Mellitus (DM), Impaired Glucose Regulation (IGR), joint representation, tongue image, facial image, sublingual image.

1. Introduction

The number of people suffering from diabetes mellitus (DM) [3], [4], [5] is increasing each year, which is predicted to reach 366 million by 2030 [6], causing disabilities, economic hardship and even death. An accurate diagnosis of DM, especially for its early stage also known as Impaired Glucose Regulation (IGR), is becoming more and more important. Until now, the fasting plasma glucose (FPG) test is a standard method to diagnose DM in many hospitals. FPG test is performed by analyzing the patient's blood glucose level after the patient has gone at least 12 hours without taking any food. This method is accurate, but inconvenient. The blood required detecting method can be consider invasive and slightly painful, and even has a risk of infection (piercing process).

In recent years, some works have been done on non-invasive methods to diagnose specific diseases by using body surface features (the tongue, face, and sublingual vein). In medical systems, the microcirculation ability of a patient suffering from DM or IGR would be influenced depending on his or her blood glucose level. This kind of change would result in clustering of red corpuscles and platelets, and even the endothelial injury and the basement membrane thickening. The human face is the best way to show these changes since it contains the most number of microvascular vessels. This is why the face based experiments often achieve a better performance. Simultaneously, as the tongue consists of a large number of microvascular vessels, it also has the valuable information for DM or IGR diagnosis. Besides this, due to the fact a reduction of the microcirculation ability is long term, peripheral vessel vascular blockage

may occur, which is indicated in the sublingual vein. As a result, these changes may have an influence on the color, texture and geometry of a patient's face, tongue and sublingual vein. Therefore, [7], [8], [9], [10], [11], [12], [13] have done many works on these three features for disease diagnosis.

The experimental results of disease diagnosis based on tongue, facial and sublingual images have proved the effectiveness and reasonability of aforementioned non-invasive methods. [14] first captured a precise facial image for diagnosis using a chamber with the LED light and digital camera. They established a five color scale for facial image to measure the changes caused by internal organs. Wang et al. [14] proposed a mathematically described tongue color space. They statistically studied main 12 types of color distribution in tongue with over 9000 tongue images, and their corresponding experiments illustrated that these colors contributed to the disease classification. In [2], Zhang et al. used facial block color with sparse representation classifier (SRC) [15], [16], [17], [18], [19], [20] for DM detection. Moreover, extracted color, texture and geometry features of tongue feature were exploited to detect DM and nonproliferative diabetic retinopathy (DR) [1]. A heart disease diagnostic system based on facial color [21] was proposed. Five facial blocks were extracted from the facial image to detect hepatitis in [22], and the average accuracy achieved 73.6% using the average RGB pixel intensities features. Zhang and Wang etc. [23] took both color and texture features into account for computerized facial diagnosis.

However, despite of various tongue, face or sublingual diagnosis methods proposed for disease detection, most of them regarded either tongue, face or sublingual vein as an independent one and ignored the relationship among them which may have an effect on the overall classification performance. As shown in Fig. 1, it is easy to see that some healthy or DM samples can not be classified with tongue features but facial features or sublingual features do. Similarly, some samples can be detected by the tongue task or the sublingual task but the facial task is unable. In particular, some patients are difficult to be diagnosed with tongue, face and sublingual vein (e.g., the last column in 'Healthy' part of Fig. 1), while a combination of them may have a possibility for accurate

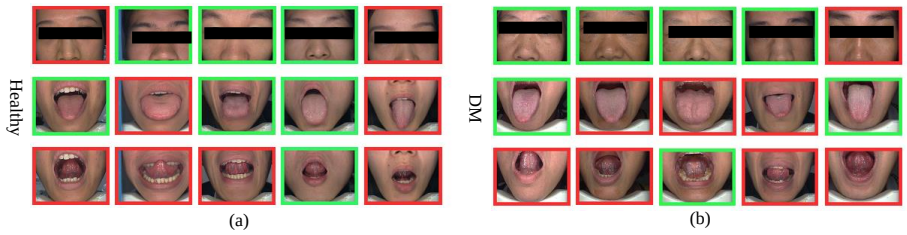

Figure 1: Example classifications using Sparse Representation Classification. A green border indicates correct classification and a red border represents incorrect classification. The first row represents experimental results using facial images, followed by the tongue images and sublingual images in the second row and third row of the 'Healthy' part. So does the 'DM' part.

diagnosis. Thus, an effective exploitation of the complementary information is beneficial for the diagnosis of diseases. A naive way of taking the tongue, face and sublingual vein into account is to concatenate these three vectors as a single one. However, it is not an efficient way since these the tongue, face and sublingual vein are different. Furthermore, the concatenated feature dose not exploit the cross correlated information among the original data. Therefore, it is necessary to have a research in combination with the tongue, face and sublingual vein.

In this paper, we propose a novel fusion classifier to discriminate between DM (or IGR) and healthy controls. In particular, our proposed method jointly represents three kinds of features obtained from tongue, facial and sublingual images and shares a similarity between them. In addition, consider differences between those tasks which contain useful information for classification, we also extract individual components that keeps the diversity between them. In this case, both similarity and distinctiveness of tongue, facial and sublingual features are exploited, being beneficial for the disease detection. An optimal algorithm based on Linearized Alternating Directions Method (LADM) [24], [25] and Augmented Lagrangian Multiplier (ALM) method [26], [27] are applied to solve the presented strategy.

The rest of this paper is organized as follows. In Section 2, we briefly describe

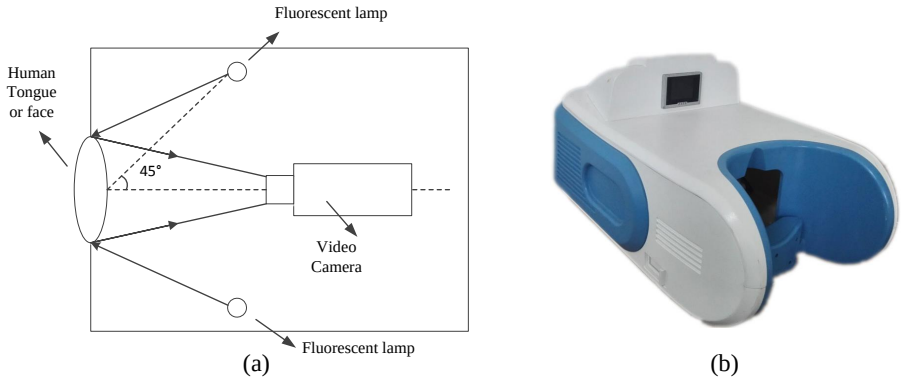

Figure 2: Image capture device. (a) Viewing geometry and imaging path of the imaging device. (b) Appearance of the device and the system.

our previous work including the image capture device and the corresponding feature extraction. In Section 3, we analyze the proposed fusion classifier. Section 4 illustrates the experimental results, followed by concluding remarks in Section 5.

2. Image Capture Device and Feature Extraction

In this section, we will first briefly introduce the image capture device of the tongue, face and sublingual vein, and then describe the feature extraction of these three types of images.

2.1. Tongue and Facial Capture Device

The image capture (tongue, face and sublingual) device is shown in Fig. 2, consisting of a SONY 3-CCD video camera placed in the center and two D65 fluorescent tubes situated symmetrically on either sides of the camera in order to produce a uniform illumination. Particularly, the angle between the incident light and emergent light is 45 (shown in Fig. 2(a)), recommended by Commission Internationale de l'Eclairage (CIE). A patient places his or her chin on a chin rest and show the tongue, face and sublingual vein to the camera (shown in Fig. 2(b)) to obtain corresponding images by changing the height

or position of the chin rest. Each image saved in JPEG format with 640×480 size is color corrected [28] to eliminate any variability in color images caused by changes of the illumination and the device dependence. Using this correction model, original tongue, facial and sublingual images are corrected, and pixels are transformed in the standard RGB (sRGB) color space.

2.2. Feature Extraction

In our previous work [1], [2], we have proposed a method to extract color, texture and geometry features of the three types of images. In this subsection, we will briefly introduce them which would make readers understand following sections more easily. It should be noted that refer to [29], four blocks with 64×64 size strategically located around the face are extracted, which contain the information of the health status of the human. Details of the location of blocks can be found in [30]. Both color and texture extraction for facial images are based on these blocks. Similarly, we also have defined eight blocks with 64×64 size for tongue texture feature extraction. More details can be found in [1]. For the sublingual images, the main domain about the vein is first decomposed and its color and geometry features are then extracted.

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 [31]

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

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 $x \leq 0.008856$.

We then compare the obtained LAB values with 12 predefined colors for the tongue [1], 6 predefined colors for the face [2] and 6 predefined colors (the method of defining these colors is same in the three types of images) for the sublingual vein to assign the color value which is closest to it using Euclidean distance. After evaluating all tongue, facial or sublingual pixels, the total of each color is summed and divided by the total number of pixels. We regard these ratios as the color feature. For tongue image, we statistically extract 12 different colors as the feature. Thus, a 12 dimensional vector for the tongue is obtained. Similarly, we select 6 different colors and 6 different colors for face and sublingual, respectively.

2.2.2. Texture Feature

The 2-D Gabor filter is applied to calculate the texture of each block.

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

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. A response $R_k(x, y)$ is produced by convolving each filter with a texture block.

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

where the symbol $*$ denotes 2-D convolution and the function $im(x, y)$ represents the texture block. Then the maximum pixel intensity is selected following $FR(x, y) = \max(R_1(x, y), \cdots, R_n(x, y))$. Apart from the texture value of each block, we also add the mean of these values as an additional texture value. Thus, we can get a 9 dimensional vector for the tongue and 5 dimensional vector for each block of the face.

2.2.3. Geometry Feature

Statistically, a person who suffers from DM or IGR would affect the geometry of the tongue [1] and the sublingual vein [32]. In our previous work, we

have introduced the details of the geometry feature extraction [1] and these features are based on measurements, distances, areas, and their ratios. For tongue images, 13 geometry features are selected (width, length, length-width ratio, smaller half-distance, center distance, center distance ratio, area, circle area, circle area ratio, square area, square area ratio, triangle area, and triangle area ratio). For sublingual images, 6 geometry features are selected (length, width, and length ratios of each side vein). Note that, before extracting the geometry feature of sublingual images, we first decompose the image to get the sublingual vein.

3. Joint Similar and Specific Learning

A general framework for fusion is proposed in this section. Before introducing our proposed method, we first briefly review the sparse representation classifier (SRC).

3.1. Sparse Representation Classifier

Given a set of training samples and a test sample, the main idea of SRC is that the test sample is represented as a linear combination of the training samples, and the representation coefficients are required to be as sparse as possible. In practice, the l_1-norm minimization is applied to ensure the sparsest linear representation of the test sample over the training samples. Suppose that matrix $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \cdot, \cdot, \cdot, \mathbf{D}_J]$ called dictionary is the set of training samples, where J is the number of classes, $\mathbf{D}_i \in \mathbb{R}^{m \times n_i}$ is the training set of the *i-th* class with m dimension and n_i samples. Each column in $\mathbf{D}$ is called atom. A test sample $\mathbf{y} \in \mathbb{R}^{m \times 1}$ can be denoted by

$$\hat{\boldsymbol{\alpha}} = \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \tag{5}$$

where λ is the penalty parameter, $\|\cdot\|_2^2$ is the l_2 norm and $\|\cdot\|_1$ is l_1 norm. $\hat{\boldsymbol{\alpha}} = [\hat{\boldsymbol{\alpha}}_1; \hat{\boldsymbol{\alpha}}_2; \cdots; \hat{\boldsymbol{\alpha}}_J]$ is the sparse coefficient, and $\hat{\boldsymbol{\alpha}}_i$ is the sparse coefficient corresponding to $\mathbf{D}_i$.

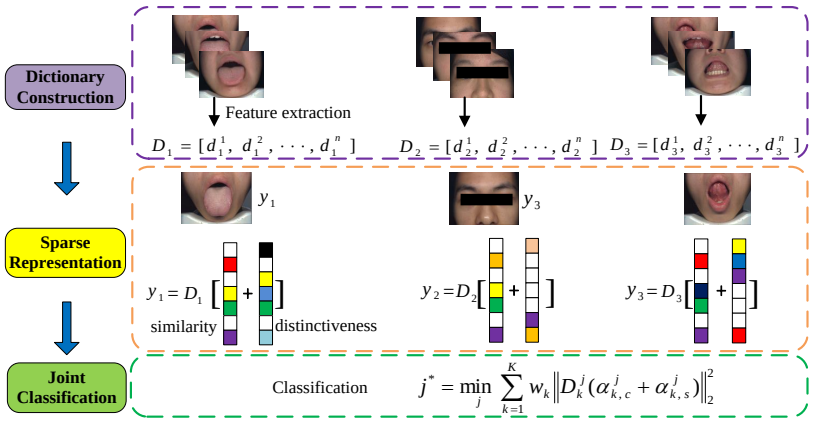

Figure 3: The framework of our proposed method JSSL. JSSL contains three parts: dictionary construction, sparse representation and joint classification. Firstly, the dictionary consists of training samples; secondly, tongue, facial and sublingual features of a given test sample are represented sparsely with the dictionary, and the representation coefficients are divided into two parts which are similar components and individual components; thirdly, the label of the test sample is decided according to the total reconstruction error.

If the test sample $\mathbf{y}$ is from class i, then it can be well represented by the training samples from i-th class. In other words, among its representation coefficients $\hat{\boldsymbol{\alpha}}$ over all the training samples, only coefficients in class i will be significant while others will be insignificant. Then the class label of the test sample is determined by the following Eq. (6)

$$i^* = \min_i \|\mathbf{y} - \mathbf{D}_i \hat{\boldsymbol{\alpha}}_i\|_2^2 \tag{6}$$

More information about the SRC can be found in [15].

3.2. Joint Similar and Specific Learning Model

As mentioned above, different tasks or feature vectors from a same sample may share some similarity. It is reasonable to assume that representation coefficients coded on their associated dictionaries of different tasks should be similar which would make the representation stable. For example, a test sample, containing tongue vector, facial vector and sublingual vector, is represented as a

linear combination of the training samples; since all vectors are belonging to a same sample, they will be well represented by the training samples of their corresponding class, and hence the position and values of significant coefficients are similar. To achieve the above goal, we use [33] the following term to achieve the similarity of the tongue, facial and sublingual feature vectors.

$$\min_{\boldsymbol{\alpha}_k} \sum_{k=1}^{K} \|\boldsymbol{\alpha}_k - \bar{\boldsymbol{\alpha}}\|_2^2 \tag{7}$$

where $\boldsymbol{\alpha}_k$ is the representation coefficient of the k-th vector and $\bar{\boldsymbol{\alpha}} = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\alpha}_k$ is the mean vector of all $\boldsymbol{\alpha}_k$ (K is the number of different types of vectors). It is easy to see that Eq. (7) aims to reduce the variance of different representation coefficients $\boldsymbol{\alpha}_k$, making them similar to each one. However, this assumption is too restrictive since there is also distinctiveness among them. Therefore, it is necessary to not only exploit the similarity among all tasks but also keep the flexibility of each task. In this case, the balance between similarity and distinctiveness will represent the original sample more stable and accurate which is beneficial for classification.

To address aforementioned problem, we divide the representation coefficients $\boldsymbol{\alpha}_k$ into two parts: the similar part and the specific part. Specifically, $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s$, where $\boldsymbol{\alpha}_k^c$ denotes the similarity, while $\boldsymbol{\alpha}_k^s$ denotes the distinctiveness. The framework of our proposed method is shown in Fig. 3. The formulation of our model is

$$\min \sum_{k=1}^{K} \left\{ \|\mathbf{y}_k - \mathbf{D}_k\left(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s\right)\|_2^2 + \tau \left\|\boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c\right\|_2^2 \right\} + \sum_{k=1}^{K} \lambda \left(\|\boldsymbol{\alpha}_k^c\|_1 + \|\boldsymbol{\alpha}_k^s\|_1\right) \tag{8}$$

where $\mathbf{y}_k$ is the test sample, $\mathbf{D}_k = [\mathbf{D}_k^1, \mathbf{D}_k^2, \cdots, \mathbf{D}_k^J]$ is training samples of the k-th task, e.g. tongue, facial or sublingual vector, and $\mathbf{D}_k^i \in \mathbb{R}^{m_k \times n_k^i}$ is the training set of the k-th task of the i-th class with m_k dimension and n_k^i samples; $\bar{\boldsymbol{\alpha}}^c = \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{\alpha}_k^c$ is the mean value of similar sparse representation coefficients of all tasks, and τ and λ are the non-negative penalty constants. From Eq. (8), we can see that our model aims to extract the similar components of each task

through $\mathbf{x}_k^c$, while also keeps individual components of each task through $\mathbf{x}_k^s$. Note that at the step of exploiting correlation, we do not directly set each $\boldsymbol{\alpha}_k^c$ to be equal, but instead of minimizing the distance between them. This way also makes our approach more flexible. In addition, consider the test sample would be linearly represented by atoms of the dictionary belonging to its own class, we apply l_1 norm on both $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^s$ to keep the sparsity.

3.3. Optimization of JSSL

We alternatively update the similar coefficients $\boldsymbol{\alpha}_k^c$ and special coefficient $\boldsymbol{\alpha}_k^s$. For example, we update $\boldsymbol{\alpha}_k^c$ by fixing $\boldsymbol{\alpha}_k^s$, and vice versa.

Update $\boldsymbol{\alpha}_k^c$: By fixing $\boldsymbol{\alpha}_k^s$, the optimization solution of Eq. (8) with respect to $\boldsymbol{\alpha}_k^c$ equals to the following problem

$$\boldsymbol{\alpha}_k^c = \arg\min \|\mathbf{y}_k - \mathbf{D}_k(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s)\|_2^2 + \tau\|\boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c\|_2^2 + \lambda\|\boldsymbol{\alpha}_k^c\|_1 \tag{9}$$

we apply Augmented Lagrangian method (ALM) algorithm to modify Eq. (9).

Applying the ALM, the problem of (9) can be modified as follows.

$$\boldsymbol{\alpha}_k^c = \arg\min \|\mathbf{y}_k - \mathbf{D}_k(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s)\|_2^2 + \tau\|\boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c\|_2^2 + \lambda\left\|\boldsymbol{\alpha}_k^{c\prime}\right\|_1 + \frac{\mu}{2}\left\|\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c\prime} + \frac{\mathbf{z}_k}{\mu}\right\|_2^2 \tag{10}$$

where $\boldsymbol{\alpha}_k^{c\prime}$ is the relaxed variable, $\mathbf{z}_k$ is the *k-th* lagrangian multiplier, and μ is the step value. Then we can optimize $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^{c\prime}$ alternatively.

(a) Firstly, we fix $\boldsymbol{\alpha}_k^{c\prime}$ to get $\boldsymbol{\alpha}_k^c$

$$\boldsymbol{\alpha}_k^c = \arg\min \|\mathbf{y}_k - \mathbf{D}_k(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s)\|_2^2 + \tau\|\boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c\|_2^2 + \frac{\mu}{2}\left\|\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c\prime} + \frac{\mathbf{z}_k}{\mu}\right\|_2^2 \tag{11}$$

Follow the Ref. [33], a closed-form solution of $\boldsymbol{\alpha}_k^c$ can be derived:

$$\boldsymbol{\alpha}_k^c = \boldsymbol{\alpha}_{0,k}^c + \frac{\tau}{K}\mathbf{P}_k\mathbf{Q}\sum_{\eta=1}^{K}\boldsymbol{\alpha}_{0,\eta}^c \tag{12}$$

where $\mathbf{P}_k = (\mathbf{D}_k^T\mathbf{D}_k + (\tau + \frac{\mu}{2})\mathbf{I})^{-1}$, $\boldsymbol{\alpha}_{0,k}^c = \mathbf{P}_k(\mathbf{D}_k^T(\mathbf{y}_k - \mathbf{D}_k\boldsymbol{\alpha}_k^s) + \frac{\mu}{2}\boldsymbol{\alpha}_k^{c'} - \frac{\mathbf{z}_k}{2})$, and $\mathbf{Q} = (\mathbf{I} - \frac{\tau}{K}\sum_{\eta=1}^{K}\mathbf{P}_\eta)^{-1}$.

(b) Secondly, after fixing $\boldsymbol{\alpha}_k^c$, the optimization solution of Eq. (10) can be reduced to Eq. (13) at the step of updating $\boldsymbol{\alpha}_k^{c\prime}$.

$$\boldsymbol{\alpha}_k^{c\prime} = \arg\min \lambda\left\|\boldsymbol{\alpha}_k^{c\prime}\right\|_1 + \frac{\mu}{2}\left\|\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c\prime} + \frac{\mathbf{z}_k}{\mu}\right\|_2^2 \tag{13}$$

Algorithm 1 Algorithm of updating $\boldsymbol{\alpha}_k^s$ in JSSL

Input: σ, $\gamma = \lambda/2$, $\mathbf{y}_k$, $\mathbf{D}_k$, and $\boldsymbol{\alpha}_k^c$, $k = 1, \cdots, K$

Initialization: $\tilde{\boldsymbol{\alpha}}_k^{s(1)} = \mathbf{0}$ and $h = 1$,

1: **for** $k = 1, ..., K$ **do**

2: **while** not converged **do**

3: h=h+1

4: $\tilde{\boldsymbol{\alpha}}_k^{s(1)} = \mathbf{S}_{\gamma/\sigma}\left(\tilde{\boldsymbol{\alpha}}_k^{s(h-1)} - \frac{1}{\sigma}\nabla\mathbf{F}(\tilde{\boldsymbol{\alpha}}_k^{s(h-1)})\right)$ where $\nabla\mathbf{F}(\tilde{\boldsymbol{\alpha}}_k^{s(h-1)})$ is the derivative of the left of Eq. (15) $\|\mathbf{y}_k - \mathbf{D}_k(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s)\|_2^2$, and $\mathbf{S}_{\gamma/\sigma}$ is a soft threshold operator that defined in Eq. (14);

5: **end while**

6: **end for**

Output: $\boldsymbol{\alpha}_k^s = \tilde{\boldsymbol{\alpha}}_k^{s(h)}$, $k = 1, \cdots, K$

Then $\boldsymbol{\alpha}_k^{c\prime}$ could be derived by operating Threshold$(\boldsymbol{\alpha}_k^c + \frac{\mathbf{z}_k}{\mu}, \frac{\lambda}{\mu})$. The operation of soft threshold is shown as follows.

$$\left[\mathbf{S}_{\lambda/\mu}(\boldsymbol{\beta})\right]_i = \begin{cases} 0 & |\boldsymbol{\beta}_j| \leq \lambda/\mu \\ \boldsymbol{\beta}_i - \text{sign}(\boldsymbol{\beta}_i)\lambda/\mu & \text{otherwise} \end{cases} \tag{14}$$

where $\boldsymbol{\beta}_i$ means the value of the *i-th* component of $\boldsymbol{\beta}$. After getting $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^{c\prime}$, $\mathbf{z}_k$ and μ can be updated following $\mathbf{z}_k = \mathbf{z}_k + \mu(\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c\prime})$ and $\mu = 1.2\mu$. In order to avoid μ being to large, we pre-fix the maximum of μ, following $\mu = \min(1.2\mu, 1000)$ and we set the initial value of μ as 0.01.

Update $\boldsymbol{\alpha}_k^s$: After acquiring $\boldsymbol{\alpha}_k^c$, the optimization of Eq. (9) can be reformulated to Eq. (15) at the step of updating $\boldsymbol{\alpha}_k^s$.

$$\boldsymbol{\alpha}_k^s = \arg\min \|\mathbf{y}_k - \mathbf{D}_k(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s)\|_2^2 + \lambda \|\boldsymbol{\alpha}_k^s\|_1 \tag{15}$$

In fact, there are many methods to tackle the problem (13). For example, both ALM and Iterative Projection Method (IPM) [34] could deal with it. In this paper, we use IPM to address Eq. (13), as described in Algorithm 1.

The complete algorithm of JSSL is summarized in Algorithm 2. The values of parameters λ and γ are decided through the cross validation. Note that, since we introduce a new variable $\boldsymbol{\alpha}_k^{c'}$ at the step of updating $\boldsymbol{\alpha}_k^c$, another reasonable

Algorithm 2 Joint Similar and Special Learning (JSSL)

Input: λ, τ, $\mathbf{y}_k$, $\mathbf{D}_k$, $k = 1, \cdots, K$

Initialization: $\boldsymbol{\alpha}_k^c = \mathbf{0}$, $\boldsymbol{\alpha}_k^s = \mathbf{0}$, $\boldsymbol{z}_k = \mathbf{0}$

1: **while** not converged **do**

2: **Update coefficients** $\boldsymbol{\alpha}_k^c$: fix $\boldsymbol{\alpha}_k^s$

(a) compute $\boldsymbol{\alpha}_k^c$ following Eq. (12)

(b) compute $\boldsymbol{\alpha}_k^{c\prime}$ following Eq. (13)

(c) $\mathbf{z}_k = \mathbf{z}_k + \mu(\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c\prime})$

3: **Update coefficients** $\boldsymbol{\alpha}_k^s$: fix $\boldsymbol{\alpha}_k^c$, and solve $\boldsymbol{\alpha}_k^s$ following **Algorithm 1**

4: **end while**

Output: $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^s$ $k = 1, \cdots, K$

way is that we could alternatively update $\boldsymbol{\alpha}_k^{c'}$ and $\boldsymbol{\alpha}_k^c$ until convergence, and then get the solution of $\boldsymbol{\alpha}_k^c$. At this time, the value of $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^{c'}$ is similar.

3.4. The Classification Rule of JSSL

After obtaining the representation coefficients, the decision is ruled in favor of the class with total lowest reconstruction residual over all K vectors.

$$j^* = \min_j \sum_{k=1}^{K} w_k \left\| \mathbf{y}_k - \mathbf{D}_{k,j}(\boldsymbol{\alpha}_{k,j}^c + \boldsymbol{\alpha}_{k,j}^s) \right\|_2^2 \tag{16}$$

where $\mathbf{D}_{k,j}$, $\boldsymbol{\alpha}_{k,j}^c$ and $\boldsymbol{\alpha}_{k,j}^s$ are the elements of the dictionary $\mathbf{D}_k$, the similar coefficient $\boldsymbol{\alpha}_k^c$ and the specific coefficient $\boldsymbol{\alpha}_k^s$ of *j-th* category, respectively; w_k is the weight value corresponding to the *k-th* vector, which could be obtained by exploiting the method mentioned in [35].

4. Experimental Results

In this section, we conduct two types of experiments. Healthy versus DM classification is first provided. Then we present the numerical results of Healthy versus IGR classification. In both experiments, KNN [36], SVM [37], [38], SRC [15], GSRC (group sprase) [39] which are general and effective classifiers are used for each individual task classification. Without loss of generality, we concatenate

different features of different tasks as a single vector and refer it as tongue feature, facial feature or sublingual feature, respectively. Thus, the K in Eq. (8) is equal to 3. Considering that KNN, SVM, SRC and GSRC can only handle a single-task classification, we also concatenate multiple tasks as a long feature vector, and use this long feature as the inputs for these four algorithms, symbolled as 'combination'. In order to qualitatively illustrate the effectiveness of JSSL, two feature fusion methods called multi-task joint sparse representation classifier (MTJSRC) [35] and relax collaborative representation (RCR) [33] are also taken into to account as the comparison methods.

4.1. Image Dataset

The tongue, facial and sublingual sample database comprises 504 samples split into 192 Healthy samples, 198 DM samples and 114 IGR samples. Each sample has three types of images including tongue image, facial image and sublingual image, respectively. All images were captured at the Guangdong Provincial TCM Hospital, Guangdong, China, from the early 2014 to the late 2015. Healthy samples were verified through a blood test and other examination. If indicators from these tests fall within a certain range (set by the Guangdong Provincial TCM Hospital), they were regarded as healthy. The FPG test was applied to diagnose whether a sample was suffering from the DM or IGR. When using the FPG test, all the samples had gone at least 12 hours without taking any food. For the DM patients, the blood glucose level was equal or larger than 7.11mmol/L, while the blood glucose level of IGR samples was between 6.1mmol/L and 7.11mmol/L. All these standard indictors are decided by the Guangdong Provincial TCM Hospital.

4.2. Healthy Versus DM Classification

We first test the performance of our proposed fusion method in identification of DM for healthy controls, with the tongue, facial and sublingual feature vectors. We randomly select the number of training samples from 30 to 100, and the rest samples are used for testing. Fig. 4 illustrates the experimental results

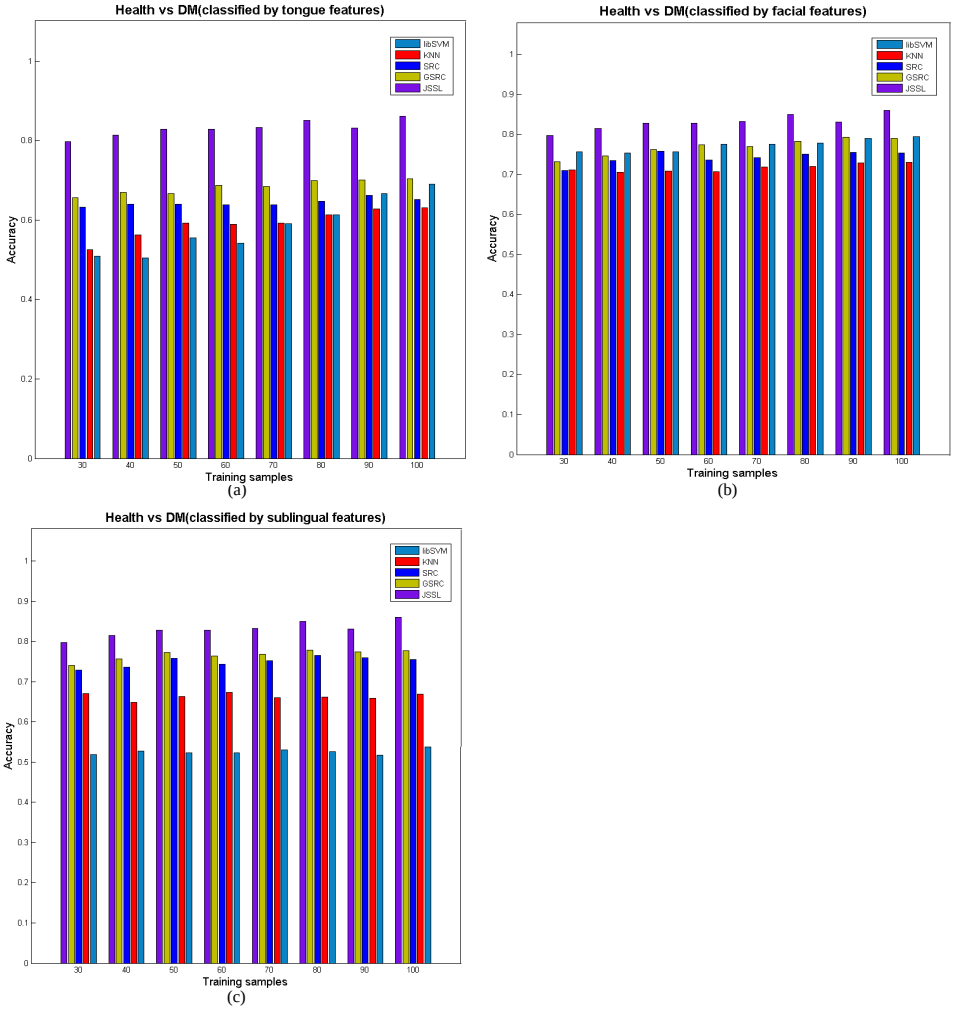

Figure 4: Comparison of Healthy and DM performance of the tongue, facial, sublingual feature and their fusion based classification methods. (a) Comparison of our method with the tongue image based feature. (b) Comparison of our method with the facial image based feature. (c) Comparison of our method with the sublingual image based feature.

of our JSSL approach, compared with the strategies using each individual task with different classifiers. Note that Fig. 4 only shows the averaged results of 5 independent experiments. It is easy to see that the combined measurement of tongue, facial and sublingual features consistently achieves more accurate discrimination between DM patients and healthy controls. Particularly, compared with single tongue feature, our method JSSL achieves about more than 15% accuracy. The classification accuracies obtained by JSSL are gradually rising with the increasing number of training samples, and they are almost all higher than 80%. In contrast, the best accuracy on facial feature based classification is only close to 80%. The accuracies obtained by our proposed algorithm along with the error bar and several state-of-the-art results directly from the single task are tabulated in Table 1. One can observe that JSSL based method performs much better than single task based approaches not only on the average accuracy, but also on the error bar.

Besides, the results by feature concatenation methods (concatenate the tongue, facial and sublingual features as a single vector) and another two fusion methods (MTJSRC and RCR) are listed in Table 1. As we can observe from Table 1, the accuracies obtained by our proposed algorithm are higher at most time. The independent K-NN, SVM, SRC and GSRC combination based approaches are further inferior to our method since JSSL takes the correlation of tongue, facial and sublingual feature vectors into account. Note that, the results obtained by simply concatenating different vectors may even suffer a large performance drop, such as K-NN and SRC. The reason of this phenomenon is that the concatenating methods roughly combine different tasks together. However, different tasks have different contributions to the classification accuracy, and a simple combination is possible to take a drop of the final classification rate. For example, in this DM detection experiment, because of the inferior results based on the sublingual vector, a simple concatenation reduces the contribution of face vector, which consequently influences the final accuracy. Simultaneously, compared with another two multi-task fusion approaches: MTJSRC and RCR, JSSL is also competitive. Although the MTJSRC achieves the best result when

the number of training samples is 90, JSSL gains about more than 2% improvement than MTJSRC and RCR at most time. Moreover, values of the error bar by our method are also smaller which means JSSL is more stable.

In addition, Fig. 5 further plots the ROC curves of different classification methods for DM detection when the number of training samples is 100. And their corresponding covered areas are listed in Table 2. Note that since our classification decision is based on the reconstruction error, we only show the ROC curves of SRC, GSRC (single and combination), MTJSRC and RCR, whose classification decisions are similar with ours, to compare with our approach. From Fig. 5 and Table 2, we can see that the area covered by JSSL based ROC curve is obviously larger than tongue, facial or sublingual vector based curves. In contrast to the independent combination based strategies as long as the MTJSRC and RCR, JSSL also has an improvement. It is also interesting to find that, in DM detection, the facial feature based methods gain a both better average accuracy and AUC value, which means the face may be larger informative in practice. In fact, as mentioned in Section 1, the human face contains the most number of microvascular which is a best way to reveal the changes caused by the DM disease. This is why the face based methods often get a better performance than tongue or sublingual feature based approaches.

Some exemplar results of classification on DM and Healthy are shown in Fig. 6 (the corresponding exemplar results of SRC are shown in Fig. 1). As expected, SVM, SRC and GSRC are capable of discriminating healthy samples using tongue images, but fail to detect DM patients at many times. In contrast, SRC and GSRC have a prominent result on classifying DM using facial images, but fail to detect healthy samples. Different from aforementioned methods using an individual task, our presented approach get an accurate result of each sample by jointly taking tongue, face and sublingual into account.

4.3. Healthy Versus IGR Classification

In this subsection, we then apply our proposed fusion method in identification of IGR for healthy controls, with the tongue, facial and sublingual tasks,

Table 1: The average accuracy and error bar (percentage) in 5 independent experiments for DM detection.

	Training samples							
Methods	30	40	50	60	70	80	90	100
JSSL	**79.82**±1.92	**81.45**±1.50	**82.82**±1.46	**82.88**±2.15	**83.27**±1.55	**85.06**±0.87	83.17±2.15	**86.07**±1.07
K-NN (face)	71.15±1.89	70.55±3.35	70.86±2.94	70.74±2.18	71.83±1.89	71.99±3.26	72.89±2.21	72.98±1.8
libSVM (face)	75.62±2.03	75.43±2.24	75.67±2.51	77.53±1.47	77.53±2.49	77.92±3.05	79.10±1.70	79.48±3.14
SRC (face)	71.00±3.22	73.51±3.35	75.84±1.67	73.62±3.79	74.14±2.90	75.11±3.84	75.59±2.60	75.39±2.37
GSRC (face)	73.14±1.51	74.59±2.74	76.32±2.26	77.45±2.19	76.93±1.82	78.31±3.85	79.34±2.11	79.00±2.39
K-NN (combine)	68.58 ±2.48	70.35±2.36	68.18 ±1.43	71.96 ±2.24	72.03 ±2.24	71.86 ±2.69	72.99 ± 1.21	74.56 ±1.84
libSVM (combine)	77.34 ±2.91	76.08 ±2.21	76.84 ±7.20	78.60 ±1.43	76.33 ±3.67	79.05 ±1.58	79.91 ±2.11	81.15 ±1.78
SRC (combine)	61.33 ± 8.96	65.40 ±6.43	66.32±7.59	68.63 ±9.35	75.14 ±4.72	74.03 ±2.31	77.54±4.17	74.35 ±7.70
GSRC (combine)	79.03 ± 1.72	80.51 ±2.80	79.11 ±4.05	77.05 ±6.53	80.39 ± 1.83	79.83 ±2.46	82.09 ±2.36	83.35 ±2.55
MTJSRC	77.76 ± 2.93	79.04±2.56	80.48 ±1.18	80.96±2.00	80.08 ±2.16	82.51 ±1.49	81.04 ±2.17	83.14 ±1.19
RCR	77.81 ± 2.49	78.64 ±2.10	80.51±2.60	80.91 ±3.33	82.53 ±2.03	83.36 ±1.12	**83.67**±1.82	83.21 ±2.76

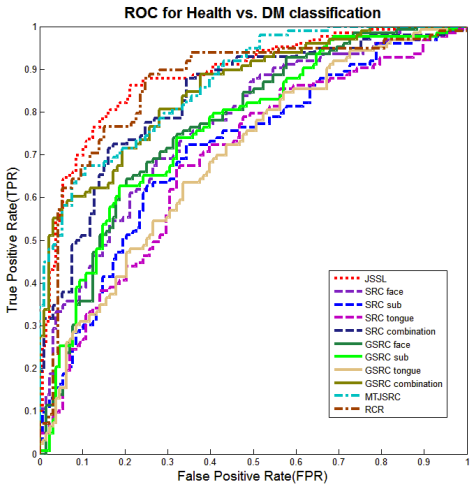

Figure 5: ROC curves of different methods and different features for DM detection.

Table 2: The area under curve (AUC) for the different methods in DM detection.

Methods	AUC	Methods	AUC
JSSL	**0.8842**	GSRC(tongue)	0.6944
SRC(face)	0.7696	SRC(combination)	0.8328
GSRC(face)	0.7686	GSRC(combination)	0.8512
SRC(sublingual)	0.7091	MTJSRC	0.8670
GSRC(sublingual)	0.7588	RCR	0.8633
SRC(tongue)	0.6885		

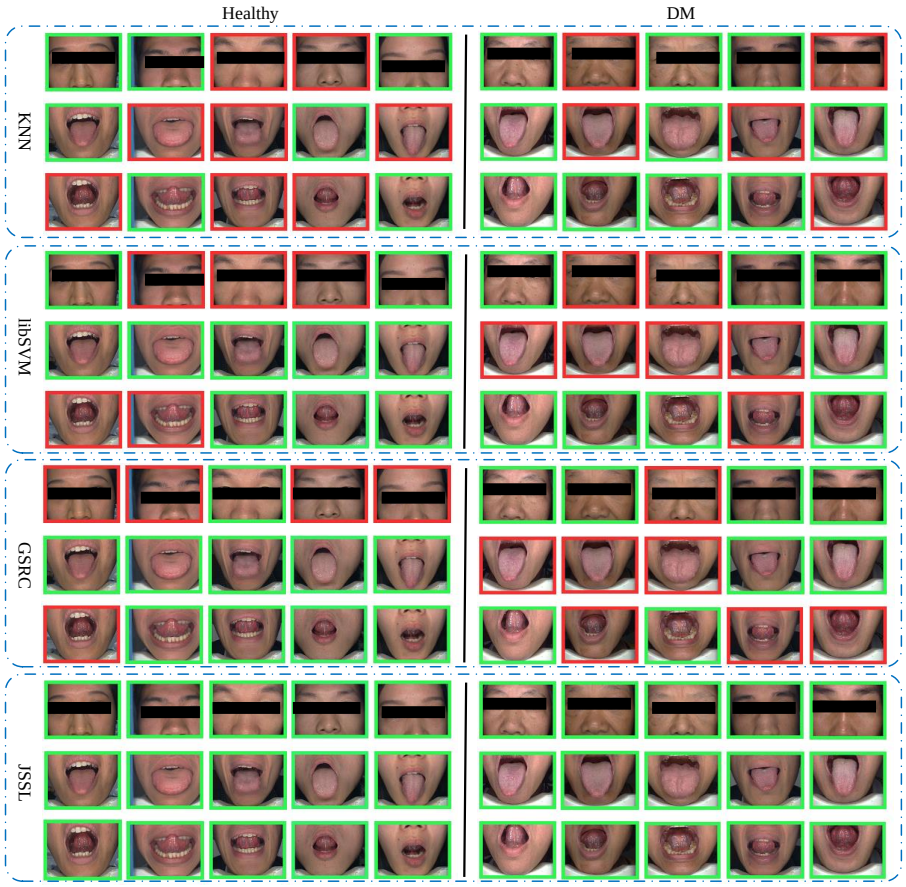

Figure 6: Example classification results of various classification methods on the tongue, facial, sublingual and their fusion features for healthy and DM diagnosis. For each image, the red border indicates incorrect classification, and the green border indicates correct classification. KNN, SVM and GSRC fail in some images. In contrast, JSSL has successfully classified all images. Particularly, JSSL can also detect some samples that can not be classified by other methods with the tongue, facial or sublingual vector.

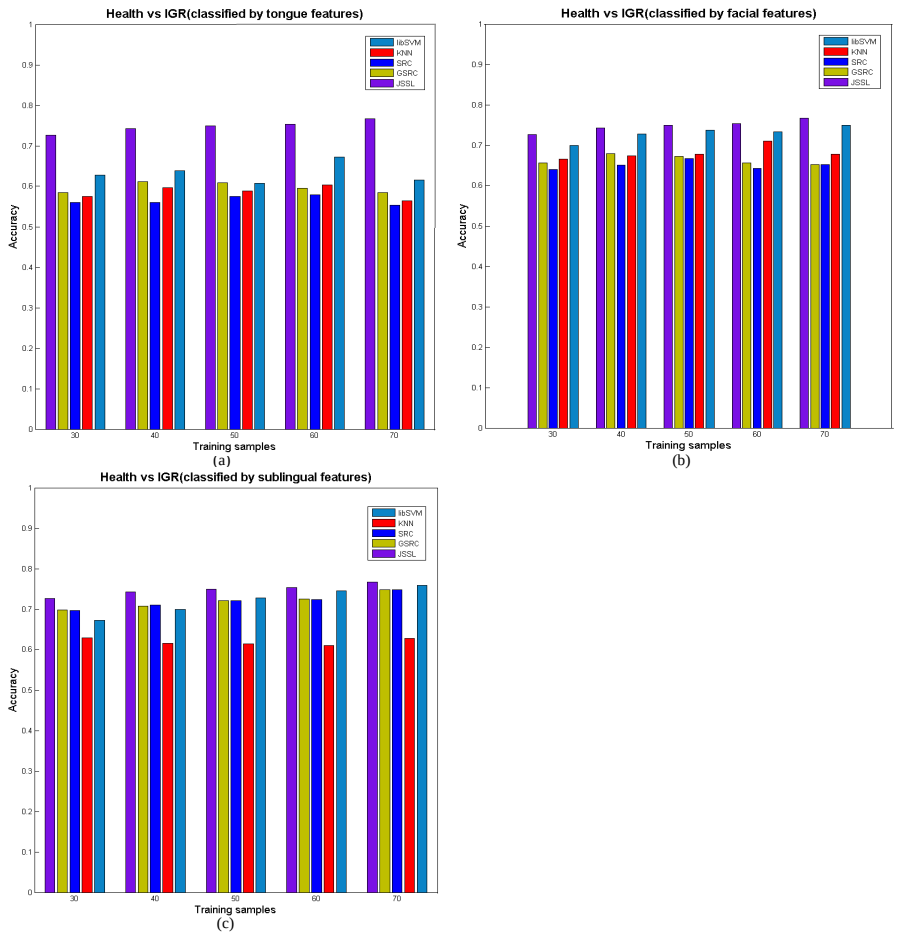

Figure 7: Comparison of Healthy Vs IGR performance of the tongue, facial, sublingual feature and their fusion based classification methods. (a) Comparison of our method with tongue image based feature. (b) Comparison of our method with facial image based feature. (c) Comparison of our method with sublingual image based feature.

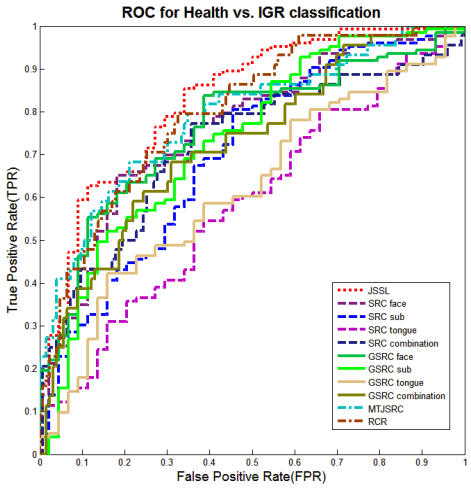

Figure 8: ROC curves of different methods and different features for IGR detection.

Table 3: The average accuracy and error bar (percentage) in 5 independent experiments for IGR detection.

	Training samples				
Methods	30	40	50	60	70
JSSL	72.63±1.64	**74.23**±2.46	74.88±1.32	**75.37**±1.85	**76.68**±3.45
K-NN (sub)	62.87±2.85	61.54±2.24	61.35±3.49	60.96±3.05	62.81±3.63
libSVM (sub)	67.21±0.93	69.95±1.50	72.80±1.17	74.49±1.52	75.87±1.51
SRC (sub)	69.68±2.96	71.01±4.32	72.08±2.08	72.41±4.19	74.85±3.79
GSRC (sub)	69.80±4.62	70.71±2.75	72.13±3.25	72.57±3.94	74.79±3.21
K-NN (combine)	66.64 ± 3.35	66.96 ±8.78	65.41 ±3.99	68.02 ±3.58	66.59 ±4.68
libSVM (combine)	**72.87** ±1.69	73.30 ±2.80	72.37 ±5.63	72.83 ± 4.38	75.81 ±1.00
SRC (combine)	59.43 ±4.33	61.76 ±7.59	62.51 ±5.77	57.43 ±6.55	67.19 ±8.24
GSRC (combine)	69.64 ±0.95	71.89 ±3.52	72.66 ±5.38	71.76 ±3.90	73.65 ± 4.96
MTJSRC	71.74 ±4.84	70.57 ±2.60	73.8 ±2.54	74.76 ± 3.17	74.13 ± 4.70
RCR	72.53 ±4.55	71.81 ±3.55	**74.94** ±2.34	73.37± 2.78	73.14 ±1.63

and make a comparison between JSSL and other existing methods. Similar with the experiment on DM detection, the number of training samples from 30 to 70 are randomly selected with 5 times, and the rest samples are used for testing. Fig. 7 illustrates the averaged experimental results of our JSSL approach, compared with the strategies based on individual task. Our presented fusion strategy accomplishes a prominent rise in classification accuracy compared with other methods based on tongue and facial. For comparison with SVM, SRC and GSRC on sublingual features, the proposed method also get a slight enhancement in the averaged accuracy. Specially, as shown in Table 3, the rate arrives at 76.68% after combination with tongue, facial and sublingual tasks, while SVM with the sublingual task, which obtains the best result in all individual tasks, only performs 75.87%. Additionally, values of our error bar are almost totally much lower than that of K-NN, SRC and GSRC, which means that our fusion method is more stable in practical applications.

Table 3 also lists average classification rates along with the error bar in 5 independent experiments for IGR diagnosis, being compared to feature concatenation methods and other fusion methods. It is easy to see that JSSL achieves an obvious improvement in contrast to K-NN, SRC and GSRC based combination methods. Similar with the results in the DM detection experiment, a simply concatenation may suffer a large performance drop, e.g. SRC. Although SVM (combination) gets a higher rate when the training number is equal to 30, our proposed algorithm gains better results on other cases. In comparison to MTJSRC and RCR, JSSL has a noticeable increase in average accuracy except when the number of training samples is 50. Particularly, our proposed fusion algorithm gets about 2.5% improvement when the number of training samples is 40 and 70.

The ROC curves of different classification methods for IGR diagnosis when the number of training samples reaches 70 is plotted in Fig. 8, followed by their corresponding covered areas in Table 4. Similarly, we only show the ROC curves and area values of SRC, GSRC, MTJSRC and RCR to compare with our approach. The ROC curves and values of covered area demonstrate that the

Table 4: The area under curve (AUC) for the different methods in IGR detection.

Methods	AUC	Methods	AUC
JSSL	**0.8278**	GSRC(tongue)	0.6155
SRC(face)	0.7579	SRC(combination)	0.7203
GSRC(face)	0.7596	GSRC(combination)	0.7243
SRC(sublingual)	0.6984	MTJSRC	0.7870
GSRC(sublingual)	0.7321	RCR	0.7982
SRC(tongue)	0.5850		

Table 5: The contributions of different features on the classification in DM and IGR detection experiments.

	Features						
Experiments	C&T&G	C&T	C&G	T&G	C	T	G
DM	86.07	81.99	82.30	60.21	79.79	62.30	67.02
IGR	76.68	70.66	71.26	61.08	64.91	50.90	62.88

JSSL has a further performance compared with single task based approaches. In addition, compared with MTJSRC and RCR, our algorithm is also competitive.

4.4. The Contribution of Different Features

In this subsection, we test the contribution on the classification of different features. The experimental results are displayed in Table 5 when the number of training samples is 100 and 70 in DM and IGR detections, respectively. Note that the symbols C, T, G denote the color, texture and geometric feature, respectively; C&T means that we remove the geometric feature and only take the color and texture features of each task into account. From Table 5 we can see that the color feature provides the most information to the classification both in DM and IGR diagnosis, followed by the geometric feature and the texture feature. The combination of color and texture or color and geometric features benefits the diagnosis, while the combination of the texture and geometric features leads to a decrease. However, a total integration of these three types of features provides a remarkable enhancement in accuracy.

5. Conclusion

In this paper, a joint sparse representation and fusion method for the Diabetes Mellitus and Impaired Glucose Regulation detection is proposed. The tongue, face and sublingual images are first captured by using a non-invasive capture device. Different features of these three types of images are then extracted. In order to exploit the correlation among them, we propose a novel fusion method to learn the similar components and specific components of tongue, facial and sublingual feature vectors. Two types of experiments in identification of DM (or IGR) from healthy controls are conducted. The experimental results substantiate the effectiveness and superiority of our fusion method, compared with the case of using a single task and other fusion strategies.

Acknowledgment

The work is partially supported by the GRF fund from the HKSAR Government, the central fund from Hong Kong Polytechnic University, the NSFC fund (61332011, 61272292, 61271344) and Shenzhen Fundamental Research fund (JCYJ20150403161923528, JCYJ20140508160910917).

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

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

[3] R. Deja, W. Froelich, G. Deja, A. Wakulicz-Deja, Hybrid approach to the generation of medical guidelines for insulin therapy for children, Information Sciences.

[4] Y. Li, C. Bai, C. K. Reddy, A distributed ensemble approach for mining healthcare data under privacy constraints, Information sciences 330 (2016) 245–259.

[5] J. Morente-Molinera, I. Pérez, M. Ureña, E. Herrera-Viedma, Creating knowledge databases for storing and sharing people knowledge automatically using group decision making and fuzzy ontologies, Information Sciences 328 (2016) 418–434.

[6] W. H. Organization, Prevention of blindness from diabetes mellitus, World Health Organization, 2006.

[7] B. Kirschbaum, Atlas of Chinese tongue diagnosis, Vol. 1, Eastland Press, 2000.

[8] D. Zhang, B. Pang, N. Li, K. Wang, H. Zhang, Computerized diagnosis from tongue appearance using quantitative feature classification, The American Journal of Chinese Medicine 33 (06) (2005) 859–866.

[9] Z. Yongtao, L. Rong, W. Zhaoping, et al., Analysis of the color characteristics of tongue digital images of 884 cases from the personsreceived a general physical examination [j], Journal of Beijing University of Traditional Chinese Medicine 1.

[10] W. Su, Z.-y. Xu, Z.-q. Wang, J.-t. Xu, Objectified study on tongue images of patients with lung cancer of different syndromes, Chinese Journal of Integrative Medicine 17 (2011) 272–276.

[11] B. Huang, J. Wu, D. Zhang, N. Li, Tongue shape classification by geometric features, Information Sciences 180 (2) (2010) 312–324.

[12] B. Li, Q. Huang, Y. Lu, S. Chen, R. Liang, Z. Wang, A method of classifying tongue colors for traditional chinese medicine diagnosis based on the cielab color space, in: Medical Biometrics, Springer, 2008, pp. 153–159.

[13] C. H. Li, P. C. Yuen, Tongue image matching using color content, Pattern Recognition 35 (2) (2002) 407–419.

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

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

[16] M. Luo, K. Zhang, A hybrid approach combining extreme learning machine and sparse representation for image classification, Engineering Applications of Artificial Intelligence 27 (2014) 228–235.

[17] Y. Xu, W. Zuo, Z. Fan, Supervised sparse representation method with a heuristic strategy and face recognition experiments, Neurocomputing 79 (2012) 125–131.

[18] D. Zhang, P. Liu, K. Zhang, H. Zhang, Q. Wang, X. Jing, Class relatedness oriented-discriminative dictionary learning for multiclass image classification, Pattern Recognition.

[19] L. An, X. Chen, S. Yang, B. Bhanu, Sparse representation matching for person re-identification, Information Sciences 355 (2016) 74–89.

[20] R. He, J. Tang, P. Gong, Q. Hu, B. Wang, Multi-document summarization via group sparse learning, Information Sciences 349 (2016) 12–24.

[21] B.-h. Kim, S.-h. Lee, D.-u. Cho, S.-Y. Oh, A proposal of heart diseases diagnosis method using analysis of face color, in: Advanced Language Processing and Web Information Technology, 2008. ALPIT'08. International Conference on, IEEE, 2008, pp. 220–225.

[22] M. Liu, Z. Guo, Hepatitis diagnosis using facial color image, in: Medical Biometrics, Springer, 2008, pp. 160–167.

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

[24] Z. Lin, R. Liu, Z. Su, Linearized alternating direction method with adaptive penalty for low-rank representation, in: Advances in neural information processing systems, 2011, pp. 612–620.

[25] T.-Y. Ji, T.-Z. Huang, X.-L. Zhao, T.-H. Ma, G. Liu, Tensor completion using total variation and low-rank matrix factorization, Information Sciences 326 (2016) 243–257.

[26] Z. Lin, M. Chen, Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, arXiv preprint arXiv:1009.5055.

[27] B. Ghasemishabankareh, X. Li, M. Ozlen, Cooperative coevolutionary differential evolution with improved augmented lagrangian to solve constrained optimisation problems, Information Sciences.

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

[29] G. Maciocia, The foundations of Chinese medicine, Churchill Livingstone, 1989.

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

[31] H.-Z. Zhang, K.-Q. Wang, X.-S. Jin, D. Zhang, Svr based color calibration for tongue image, in: Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on, Vol. 8, IEEE, 2005, pp. 5065–5070.

[32] C.-C. Chiu, C.-Y. Lan, Y.-H. Chang, Objective assessment of blood stasis using computerized inspection of sublingual veins, Computer methods and programs in biomedicine 69 (1) (2002) 1–12.

[33] M. Yang, L. Zhang, D. Zhang, S. Wang, Relaxed collaborative representation for pattern classification, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2224–2231.

[34] L. Rosasco, A. Verri, M. Santoro, S. Mosci, S. Villa, Iterative projection methods for structured sparsity regularization.

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

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

[37] C.-J. Hsieh, K.-W. Chang, C.-J. Lin, S. S. Keerthi, S. Sundararajan, A dual coordinate descent method for large-scale linear svm, in: Proceedings of the 25th international conference on Machine learning, ACM, 2008, pp. 408–415.

[38] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, C.-J. Lin, Liblinear: A library for large linear classification, The Journal of Machine Learning Research 9 (2008) 1871–1874.

[39] S. Bengio, F. Pereira, Y. Singer, D. Strelow, Group sparse coding, in: Advances in neural information processing systems, 2009, pp. 82–89.