Information Sciences 472 (2019) 1–14

Contents lists available at ScienceDirect

Information Sciences

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

Body surface feature-based multi-modal Learning for Diabetes Mellitus detection

Jinxing Li[a], Bob Zhang[b], Guangming Lu[c], Jane You[a], David Zhang[c,d,*]

[a] Biometrics Research Center, Department of Computing, The Hong Kong Polytechnic University, 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
[d] School of Science and Engineering, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China

ARTICLE INFO

Article history:
Received 25 January 2018
Revised 3 July 2018
Accepted 2 September 2018
Available online 4 September 2018

Keywords:
Diabetes mellitus
Multi-modal
Body surface feature
Gaussian process

ABSTRACT

In recent year, the number of people who are suffering from the Diabetes Mellitus (DM) has increased remarkably and the detection of DM disease has attracted much attention. Different from some existing methods which are invasive, Traditional Chinese Medicine (TCM) provides a non-invasive strategy for DM diagnosis by exploiting some features in the body surface, including the tongue, face, sublingual vein, pulse and odor. Since a combination of these modalities would contribute to improving detection performance, a novel multi-modal learning method is proposed to learn a shared latent variable among the tongue, face, sublingual, pulse and odor information, which efficiently exploits the correlation. In detail, the raw images or signals of five modalities are first captured through our non-invasive devices. Their corresponding features are then extracted, respectively. Finally, a shared auto-encoder gaussian process latent variable model (SAGP) is introduced to learn a latent variable for various modalities in a non-linear and generative way. An efficient algorithm is designed to optimize the proposed model. The DM detection experiments are conducted on a dataset composed of 548 Healthy and 356 DM samples collected by us and the results substantiate the superiority of the proposed method.

1. Introduction

In recent decades, diabetes mellitus (DM) [4,18,20] has attracted increasing attention around the world. According to the statistic report in [21], the number of people who suffer from DM will reach as many as 366 million by 2030. Thus, an accurate diagnosis of DM is becoming more and more important. A general approach for DM detection in most hospitals is the so called fasting plasma glucose (FPG) test. Specifically, FPG detects the DM according to the patient's blood glucose level. Furthermore, before making an analysis, the patient must have gone about 12 hours without taking any food. Despite the effectiveness of this method, the main limitation is that its convenience is unsatisfying since it is invasive and slightly painful (piercing process), and even has a risk of infection.

In order to tackle this problem, some recent works based on the Traditional Chinese Medicine (TCM) [36,37] have focused on the non-invasive method by exploiting some features in the body surface, including the tongue, face, sublingual vein,

* Corresponding author.
E-mail addresses: csjxli@comp.polyu.edu.hk (J. Li), bobzhang@umac.mo (B. Zhang), luguangm@hit.edu.cn (G. Lu), csyjia@comp.polyu.edu.hk (J. You), csdzhang@comp.polyu.edu.hk, yank10@mails.tsinghua.edu.cn (D. Zhang).

https://doi.org/10.1016/j.ins.2018.09.010
0020-0255/

pulse and odor. Generally speaking, DM would have an influence on a patient's microcirculation ability due to the change of blood glucose level. There would be a clustering of red corpuscles and platelets, which may subsequently result in the endothelial injury and the basement membrane thickening. As a great number of microvascular vessels exist around the tongue and face, there is a reflection in these two modalities. Additionally, the change of the microcirculation capability may cause the peripheral vessel vascular blockage, reflected in the sublingual vein. Furthermore, suffering from DM will also influence the heart beat, and subsequently make an effect on the pulse signal. For odor, it has been proved that acetone in the breath gas will become much higher than the normal level if a person suffers from DM. Therefore, various works [8,10,14,15,25,34,38] based on these modalities have been done for disease detection.

Wang and Zhang et al. have focused on studying the disease diagnosis based on these body surface features in many works [8,10,14,15,25,27,28,32–34,38], substantiating the superiority and reasonability of the non-invasive approaches. In detail, a novel device was designed by Wang et al. [30] to precisely capture a human's tongue and face images for diagnosis. This device consists of two chamber with the LED light placed symmetrically on two sides to ensure an uniform illumination, and a SONY 3-CCD camera is used to obtain the image with high resolution. Wang et al. also [30] presented a novel tongue color space, which is capable of statistically representing the tongue image with 12 types of colors. Their corresponding experiments on disease detection demonstrate the effectiveness of these colors. Moreover, Kim et al. [9] proposed a heart disease diagnostic system by exploiting the color distributions around the face image. Similarly, another work based on the face image was done for hepatitis detection in [19], achieving 73.6% in average accuracy. Also, a breath analysis system was presented for DM diagnosis by Yan et al. in [33] according to the acetone level. Additionally, a pulse acquisition system proposed in [28] achieved a satisfactory performance in disease detection.

According to the aforementioned analysis, we proposed a non-invasive DM detection method based on five modalities, including tongue, face, sublingual vein, pulse and odor. Different from related works which only consider a single modality, we aim to jointly represent these data and exploit the correlation among them which contributes to an improvement for the overall classification performance. As shown in Fig. 1, we can easily to observe that some samples including both health and DM fail to be detected if only tongue images are available. Similarly, the healthy and DM samples cannot be classified by using only face or sublingual images, as shown in the first and third rows. Therefore, it is significant to take different modalities into account simultaneously.

The remaining content of this paper is organized as follows. In Section 2, the tongue, face, sublingual vein image capture device, and pulse and odor signal extraction devices described in our previous works are briefly introduced, followed by their feature extraction. We then introduce the fusion strategy, named Shared Auto-encoder Gaussian Process latent variable model (SAGP) in Section 3. The experimental results compared with some existing methods are then demonstrated in Section 4. The paper is finally concluded in Section 5.

2. Feature extraction

In this section, an image capture device and two signal capture devices are first briefly introduced, followed by the feature extraction of aforementioned five types of modalities.

2.1. Device

2.1.1. Tongue, face and sublingual vein image capture device

In our work, three types of images are gained through a same device. As shown in Fig. 2, the device is composed of a SONY 3-CCD video camera and two D65 fluorescent tubes. The camera is located in the center and tubes are symmetrically placed on either sides of it to make a stable and uniform illumination, as shown in Fig. 2(a). To obtain different types of images, a changeable chin rest can be changed to display three types of images to the camera (shown in Fig. 2(b)). To make a color correction for the image, the image of a Munsell Colorchecker [24] is first captured by using the camera and lighting, which are the same to the acquisition of the tongue, face and sublingual images. According to the Munsell Colorchecker image, the colorchecker RGB values in the source RGB color space can be extracted. A sRGB value provided by the color company is then regarded as the target RGB color space. Here, the sRGB color space is the objective device-independent system color space and based on this space we can correctly evaluate the color value no matter whether the device is changed. At the next step, these two types of colorchecker values are inputted into the designed regression algorithm [31] to estimate the correction parameters. According to these learned parameters, we can successfully correct the images. Details of the color correction can be found in [31]. Finally, the image is saved in JPEG format and its size is $640 \times 480 \times 3$.

2.1.2. Pulse extraction device

The device for pulse signal extraction is shown in Fig. 3 [28]. From Fig. 3(a) we can see that three probes are used to extract the signals from three channels including Cun, Guan, and Chi (this was designed based on the traditional Chinese pulse diagnosis (TCPD)). A pedestal located on the bottom is also installed. In this way, the sensor can gain the pulse signal at an accurate position. As shown in Fig. 3(b), the styloid process is the end of radius bone. In fact, although our initial purpose of designing this device is to capture three types of signals including Cun, Guan, and Chi, we find that it is difficult to extract them simultaneously. In other words, Cun or Chi signals meet a missing in some cases. Additionally, we further find that Guan is more effective for disease detection compared with Cun and Chi. Thus, we only use Guan in this paper.