Information Sciences 547 (2021) 945–962

Contents lists available at ScienceDirect

Information Sciences

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

DsNet: Dual stack network for detecting diabetes mellitus and chronic kidney disease

Qi Zhang [a], Jianhang Zhou [a], Bob Zhang [a,*], Enhua Wu [b]

[a] PAMI Research Group, Dept. of Computer and Information Science, University of Macau, Macau SAR, People's Republic of China
[b] Faculty of Science and Technology, University of Macau, Macau SAR, People's Republic of China

ARTICLE INFO

Article history:
Received 6 February 2020
Received in revised form 7 July 2020
Accepted 23 August 2020
Available online 01 September 2020

Keywords:
Diabetes mellitus
Chronic kidney disease
Facial image
Medical biometrics
Traditional Chinese medicine
Stack network
Noninvasive disease detection

ABSTRACT

Diabetes mellitus and chronic kidney disease are two severe chronic diseases in the world, affecting the quality of a patient's life. However, detecting these two diseases often applies professional medical techniques such as a Fasting Plasma Glucose test and estimating the glomerular filtration rate (eGFR) measurement, which usually requires a blood test. Given the various inconveniences and risks in existing conventional diagnostic approaches, noninvasive healthcare systems based on intelligent electronic detection/prevention are preferred. To achieve this goal, we propose a progressively trainable network, i.e., dual stack network (DsNet), to distinguish patients with chronic kidney disease, diabetes mellitus from healthy people simultaneously through analyzing the facial images of candidates. The first stack subnetwork extracts high-level representative features from the facial images effectively. While the second stack subnetwork can further analyze the extracted high-level features from the first stack subnetwork, before classifying the two diseases from healthy individuals simultaneously. Extensive experiments on a dataset with 229 healthy samples, 236 diabetes, and 200 chronic kidney disease patients show that our proposed method generated the F1-score of 95.33%, 98.17%, and 94.67% for detecting chronic kidney disease, diabetes, and healthy samples respectively. Our proposed DsNet achieves significant improvements compared with other traditional noninvasive detection approaches.

1. Introduction

Diabetes mellitus is a fatal metabolic disease with no cure. It can impair a human body's ability to handle blood glucose, i.e., blood sugar, causing hyperglycemia due to a defect in insulin secretion or impairment [1]. Diabetes can lead to long-term hyperglycemia in the blood without careful management, which causes various serious complications, including heart disease, stroke, and so forth. [2]. According to the International Diabetes Federation, approximately 425 million people suffered from diabetes in 2017, accounting for around 8.8% of the adult population in the world. Diabetes also causes one of the highest numbers of deaths – an estimated 1.6 million deaths in 2016, which ranks as the 7th leading cause of death in the USA [3]. The chronic kidney disease (CKD) refers to the gradual and irreversible decline of kidney function over a period of several months or years [4]. Patients who have CKD often experience leg swelling, become easily tired, induce vomiting, and con-

* Corresponding author.
E-mail addresses: yc07485@connect.um.edu.mo (Q. Zhang), yc07424@connect.um.edu.mo (J. Zhou), bobzhang@um.edu.mo (B. Zhang), ehwu@um.edu.mo (E. Wu).

https://doi.org/10.1016/j.ins.2020.08.074
0020-0255/

fusion. Kidney failure, i.e., end-stage renal disease (ESRD), requires a kidney transplant to survive if the degree of CKD is severe [5]. According to relevant investigations, approximately 753 million people suffered from CKD in 2016, including 336 million males and 417 million females [6]. Nearly 1.2 million patients died directly from CKD in 2015, increasing from around 41 million in 1990 [7].

The most common approach to directly detect diabetes is by applying the Fasting Plasma Glucose test (FPG) [8]. To perform FPG on the examinees, blood samples are needed to analyze the glucose level by piercing a patient's fingertips, which can lead to some pain and discomfort for those being tested. Beyond that, the FPG requires a strict testing time – candidates have to wait around 12 h after eating before participating in this test. People who have CKD often do not feel obvious symptoms until their kidneys are severely damaged. At present, the most prevalent way to detect kidney disease is by using the estimated glomerular filtration rate (eGFR) measurement [9]. The eGFR can indicate the level of how well a candidate's kidney is cleaning his/her blood. To perform eGFR on the examinees, a blood test has to be implemented in order to find out the level of creatinine in the blood [10]. Then, a doctor can evaluate the eGFR value from the blood test and diagnose CKD eventually. However, these testing approaches generally take a long time to obtain results and may cause some distress to the patients.

Given these issues in diagnosing diabetes mellitus and CKD by using conventional medical techniques, developing a noninvasive and convenient digital healthcare system to detect and prevent these two diseases is necessary and desired. Recently, many researchers have attempted to implement and apply noninvasive approaches based on computerized analysis [11,12,13,14,15] to detect various diseases such as diabetes, brain disease, and heart disease by extracting color features from the facial images of candidates [16,17]. The idea of their work comes from medical biometrics [18], since unique and distinguishable features can be extracted from a particular disease. The motivation of analyzing facial images to detect diabetes and CKD in this study is inspired by two parts: One is Traditional Chinese Medicine (TCM), which considers various body regions and characters that reflect the different statuses of our organs [19]. Many researchers have investigated computerized TCM diagnosis showing its effectiveness recently [20,21,22]. For instance, in [23], the authors implemented an inception autoencoder model for clustering Chinese healthcare questions, which is effective in encouraging patients to obtain professional TCM consultation. H. Yang et al. investigated herbal prescriptions for patients by analyzing tongue images based on TCM knowledge and deep learning techniques [24]. The second part is some clinical research investigating diabetes mellitus and CKD based on skin characteristics. In [25], the authors investigated the color of a face in diabetes, indicating that most diabetes patients from their study in Jerusalem showed facial redness. O. Falodun et al. evaluated the pattern of skin disorders in 120 patients with CKD and concluded that pruritus, xerosis, and pigmentary were the most common skin characteristics in people with chronic kidney failure in Nigeria [26]. Although some favorable results were obtained by analyzing color features from facial images, to the best of our knowledge, there is little to none in the literature investigating the simultaneous detection of two or more diseases applying facial images. There are two reasons for this: the first is that the traditional color feature extraction approach cannot comprehensively retrieve useful features or relevant latent information in the facial images of the candidates, producing poor performances in detecting multiple diseases simultaneously; the second is the single classifier designed in previous studies may only be more sensitive and effective for binary classification, while not efficiently achieving optimal results for various diseases. For instance, researchers have attempted to apply convolutional neural networks for classifying diabetes by using facial key blocks directly. However, they did not investigate the relevant latent information or high-level features from the patients' facial images [27]. The corresponding average accuracy

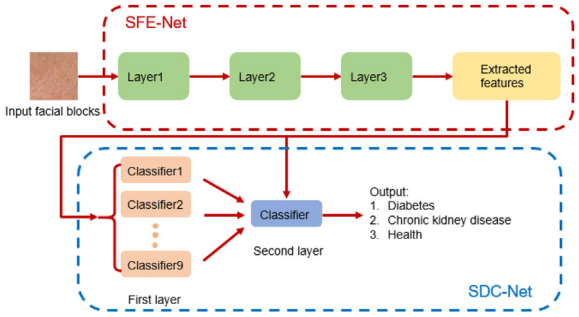

Fig. 1. The proposed Dual Stack Network (DsNet) classifies diabetes, chronic kidney disease and healthy candidates simultaneously in this study. The input data are facial blocks from the facial images, while the output is the label of the three classes. The SFE-Net is stacked with three layers (encoders) of SAEs, while the SDC-Net is constructed with two layers using a stacking framework.

reached 73%, which is not very ideal. In [21], the authors applied a joint similar and specific learning method for detecting diabetes and obtained 86.07% in accuracy, but did not explore multi-disease diagnosis simultaneously.

For those mentioned above two main problems in detecting multi-disease simultaneously, we hereby propose a dual stack network in this study. The motivation of our proposed method aims to extract more effective features from the facial images of the candidates, and obtain robust and reasonable detection performances by integrating different classifiers together. This network mainly contains two subnetworks (Fig. 1), i.e., stacked feature extraction network (SFE-Net), which can extract high-level representative features from the facial images effectively; stacked disease classification network (SDC-Net), which can further analysis the extracted features from the first subnetwork, before classifying the diseases from healthy individuals simultaneously. The SFE-Net is inspired from the basic sparse autoencoder (SAE) [28], which is popular and has been proven to be successful in image feature extraction, showing a better performance compared to traditional approaches [29]. We concatenated and implemented multi-SAEs together, i.e., stacked sparse autoencoder (SSAE) [30] to further improve the performance of extracting high-level representative features from the facial images (This is the first stack of DsNet). The SDC-Net is designed to integrate heterogeneous classifiers to obtain optimal results for multi-disease detection in this study. This subnetwork is proposed based on a stacking framework of ensemble learning theory [31], which addresses the same problem through multiple base models and combines them together to achieve better performances by forming a more accurate and robust new model [32] (This is the second stack of DsNet). The high-level extracted features generated from SFE-Net are regarded as the input for SDC-Net, where 9 different basic classifiers were fused together in the first layer by applying a stacking framework to generate the new predicted features for the classifier in the second layer, which outputs the final detection result. Besides this, the raw extracted features, i.e., *meta*-features from SFE-Net are embedded in the second layer of SDC-Net, which is aimed to capture effective relationships that may be ignored in the first layer [33]. This can further improve the final performance of detecting diabetes mellitus and CKD from healthy samples. Extensive experiments using our collected facial images (from hundreds of healthy candidates, diabetes, and chronic kidney disease patients) indicate that our proposed method, i.e., DsNet, generates superior performances than traditional feature engineering approaches.

The originality of our proposed method has three main parts. The first is that although many scientific works about diabetes detection have been achieved, few works attempted to investigate the CKD detection via a non-invasive method. Our work can provide more value in this area. Besides this, our proposed approach can detect two or more diseases from healthy samples simultaneously, which is much better and convenient compared to other conventional methods. At the same time, our computerized based method to detect disease using facial images can further make non-invasive disease detection more structured, standardized and increase its efficiency.

To sum up, we produce the following contributions in this work:

- We propose a dual stack network architecture for detecting diabetes and CKD from healthy individuals simultaneously. This proposed method makes use of two stack techniques in the first sub-network, i.e., the SFE-Net, and the second sub-network, i.e., the SDC-Net, respectively.
- SFE-Net is proposed based on a stacked sparse autoencoder (SSAE) to extract high-level features. The SSAE architecture can comprehensively retrieve effective features and latent information in the facial images of the candidates.
- SDC-Net based on the stacking framework of the ensemble method fuses various heterogeneous classifiers to find optimal results for the various diseases.
- We further implement SDC-Net by inserting raw extracted features in the second layer to enhance the robustness and accuracy of our proposed method.

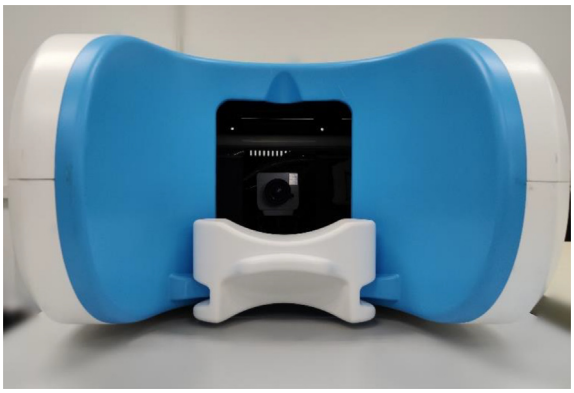

Fig. 2. The specially designed device to capture the facial images of candidates.

2. Materials and methods

2.1. Facial image dataset

To capture suitable facial images of the candidates while decreasing the effects of the environment, position, illumination, and angle of the photograph, a specially designed device with color correction was invented and applied [34]. This device can fix a candidate's chin on a chin rest, before taking his/her facial image (Fig. 2). The original facial image of the candidate taken by this device is 768*576 pixels. In order to avoid the interference of facial organs such as the eyes and mouth during detection, four facial key blocks—Forehead Block (FHB), Nose Bridge Block (NBB), Left Cheek Block (LCB), and Right Cheek Block (RCB), each with 64*64 pixels were extracted from one facial image [16] in this study (Fig. 3). Besides this, samples from diabetes, CKD, and healthy are shown in Fig. 4.

In order to train and evaluate our proposed method, the dataset containing 665 facial images was established, including 236 diabetes patients obtained from the Hong Kong Foundation for Research and Development in Diabetes, Prince of Wales Hospital in 2012, 200 patients with CKD and 229 healthy samples captured from the Guangdong Provincial Hospital of Traditional Chinese Medicine in 2016, respectively. Patients with Diabetes and CKD were diagnosed by applying the conventional medical examinations mentioned in Section 1. Please note that all patients collected in this study were confirmed not to have these two diseases simultaneously. The healthy candidates were determined through a blood test and other commonly used assessments. All of the collected data was obtained under the ethical standards explained in the Declaration of Helsinki and approved by the Science and Technology Development Fund (FDCT) of the Macau SAR. More details about the related dataset and its corresponding acquisition device's software can be found in [35].

To validate our proposed approach's superiority, our previous work in detecting diseases by analyzing color features extracted from facial images was compared in this study (refer to Section 3.2). The corresponding color features of an individual's facial key blocks were produced by applying the color gamut [16]. For further information about the usage of the facial key blocks and traditional color feature extraction method, please refer to our previous study [17,36].

2.2. Stacked feature extraction network (SFE-Net)

As shown in Fig. 1, our proposed model contains two subnetworks, i.e., Stacked Feature extraction network (SFE-Net) and stacked disease classification network (SDC-Net). The SFE-net was designed to address the problem existing in traditional color feature extraction from our previous work, which is unable to retrieve comprehensively effective features or relevant latent information in facial images when detecting multiple diseases simultaneously. Here, we implemented SFE-Net based on SSAE to process input data, i.e., raw facial key blocks, before generating high-level representative features for the subsequent SDC-Net to do further analysis.

2.2.1. Sparse autoencoder

The basic sparse autoencoder (SAE) is an unsupervised learning approach, commonly used to learn useful features from the input data [28]. The general architecture of the sparse autoencoder to obtain effective features is expressed in Fig. 5.

It can be noticed that the encoder contains the input layer and hidden layer in the sparse autoencoder transforming the input sample, i.e., facial key blocks $\times$ into a representative matrix h, which is also regarded as a new representative feature of

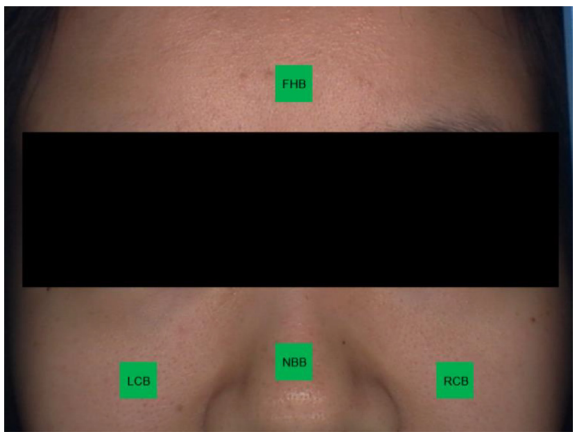

Fig. 3. The four facial key blocks: Forehead Block (FHB), Nose Bridge Block (NBB), Left Cheek Block (LCB), and Right Cheek Block (RCB) extracted from the facial image captured by our imaging device.

Fig. 4. The selected sample of (a) healthy, (b) diabetes, and (c) CKD candidates.

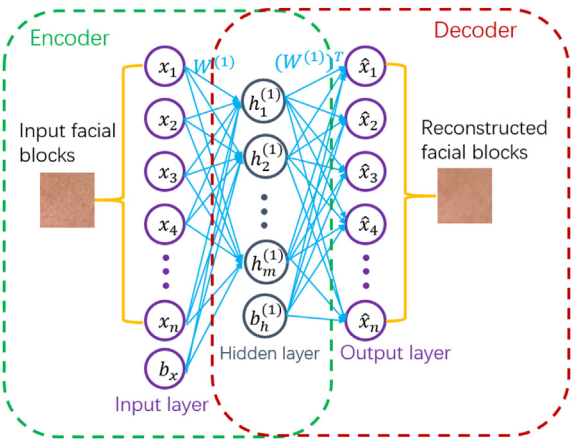

Fig. 5. The principle of the basic sparse autoencoder (SAE), consisting of an encoder and decoder for extracting high-level representative features from the facial blocks.

the input sample in the subsequent hidden layer. The decoder's output layer can reconstruct an approximated sample $\hat{x}$ based on the hidden representative feature h compared to the input sample. The loss can be minimized and calculated with an input sample $\times$,and its corresponding reconstructed sample $\hat{x}$, which is applied to retrieve the optimal parameters for training this autoencoder. Generally, the loss can be reduced by the back-propagation algorithm through learning the encoder and decoder simultaneously, before generating the corresponding weights W and biases b. Thus, the cost function of a basic SAE can be defined as:

$$\mathcal{L}_{SAE} = \frac{1}{N}\sum_{k=1}^{N}(L(x(k), d_{\widehat{\theta}}(e_{\widehat{\theta}}(x(k))))) + \alpha\sum_{j=1}^{m} KL(\rho||\widehat{\rho}_j) + \beta\|W\|_2^2 \tag{1}$$

The first term of Eq. (1) denotes the average sum-of-squares error obtained by calculating the difference between the reconstructed sample $\tilde{x}(k)$ and input sample $x(k)$. The m and index j indicate the number of units for the hidden layer and the sum of all hidden units respectively in the second term of Eq. (1). The $KL(\rho||\widehat{\rho}_j)$ represents the Kullback-Leibler (KL) divergence for the desired activations ρ, and $\widehat{\rho}_j$, which is the average activation of the hidden unit j. The KL divergence can be expressed as:

$$KL(\rho|\big|\widehat{\rho}_j\big) = \rho log\frac{\rho}{\widehat{\rho}_j} + (1-\rho)log\frac{1-\rho}{1-\widehat{\rho}_j} \tag{2}$$

where $\widehat{\rho}_j$ is shown as:

$$\widehat{\rho}_j = \frac{1}{n}\sum_{i}[a_j(x^{(i)})] \tag{3}$$

and $x^{(i)}$ signifies the nth training sample.

Finally, the third term of Eq. (1) indicates a weight decay function, in order to decrease the magnitude of the weight, and avert the overfitting problem. The weight decay term can be expressed as:

$$\|W\|_2^2 = tr\left(W^T W\right) = \sum_{l=1}^{n_l} \sum_{i}^{S_{l-1}} \sum_{j}^{S_l} \left(w_{i,j}^{(l)}\right)^2 \tag{4}$$

where S_{l-1} and S_l represent the number of neurons in layer l-1 and l respectively, while n_l indicates the number of layers. The $w_{i,j}^{(l)}$ is generally applied to associate the i^{th} neuron in layer l-1 with the j^{th} neuron in layer l.

2.2.2. Stacked sparse autoencoder

The stacked sparse autoencoder (SSAE) is implemented based on the simple SAE, concatenating multiple hidden layers of basic SAEs, which is more powerful at extracting effective high-level features from the input samples compared with the simple SAE. The output features of the encoder in each basic SAE is seen as the input data to the subsequent encoder of the following SAE, where multiple encoders are formed successively (briefly described SFE-Net in Fig. 1) [37]. Here, we propose an SFE-Net with 3 hidden layers, i.e., associating with 3 encoders of the basic SAE together to process raw facial key blocks, as shown in Fig. 6.

The input samples are raw facial key blocks described in Section 2.1. In order to train SFE-Net in this study, the optimal parameters can be defined as:

$$\theta = (\mathrm{W}, b_h, b_x) \tag{5}$$

where the corresponding weights W and biases b_h, b_x should be determined simultaneously by minimizing the loss between the input samples and the corresponding reconstructed samples with the conjugate gradient method [38]. The first new features, i.e., $h^{(1)}$ can be generated from the facial key blocks through the first hidden layer, i.e., the first encoder in SFE-Net, which can be described as:

$$f : R^{d_x} \rightarrow R^{d_{h^{(1)}}} \tag{6}$$

$$h^{(2)} = f(x) \in R^{d_{h^{(1)}}} \tag{7}$$

where the function f in Eq. (6) aims to transform the input samples into new representative features. The second new features, i.e., $h^{(2)}$ can be generated in Eq. (7) in the same way, by processing $h^{(1)}$ as the input sample for the second hidden layer. The corresponding third new features $h^{(3)}$ are also similarly yielded from the third hidden layer of SFE-Net.

The facial key blocks of each candidate $x(k)$ in this study can eventually be expressed as a high-level representative feature through the third hidden layer, as shown in Fig. 6. Therefore, all facial key blocks of the candidates can be represented as:

$$\left\{h^{(3)}(k), y(k)\right\}_{k=1}^{N} \tag{8}$$

where N indicates the number of facial key blocks from the candidates, while $y(k), h^{(3)}(k)$ are the class label and corresponding high-level features, respectively. The class label can be defined as:

$$y(k) \in \{0, 1, 2\} \tag{9}$$

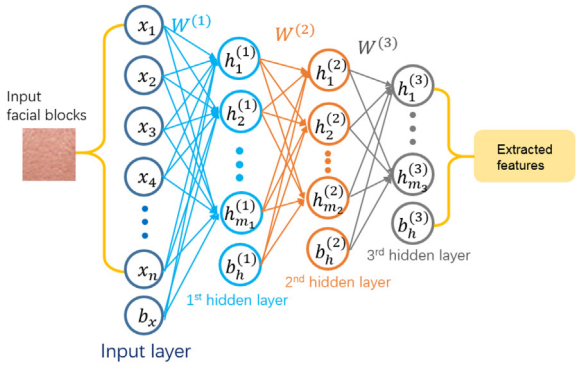

Fig. 6. The architecture of SFE-Net, implemented based on SSAE with 3 hidden layers to extract high-level features from the input facial key blocks.

where 0, 1, and 2 indicate three classes, i.e., healthy, diabetes, and chronic kidney disease, respectively, in this study. Please note that the SFE-Net is based on SSAE used here as an unsupervised learning method. Since the label information $y(k)$ is not used in this subnetwork, the corresponding high-level features are regarded as the input data for SDC-Net.

From Fig. 5, it can be seen that the facial blocks are regarded as the input for the basic sparse autoencoder (SAE), while the output is the corresponding reconstructed facial blocks. Thus, the extracted high-level features are the corresponding parameters learned in the hidden layer. For the SSAE, the learned parameters are contained in several hidden layers, which are cascaded to generate the final extracted features for the SDC-Net (shown in Fig. 6).

2.2.3. Stacked disease classification network (SDC-Net)

As is commonly known, the fusion of multiple machine learning models often improves the robustness and overall prediction ability. The stacked disease classification network (SDC-Net) is inspired by this idea, where the fusion of multiple machine learning models often improves the robustness and overall prediction ability. The stacking framework is a popular ensemble method at present [31,32], where by applying this framework, SDC-Net can learn the effective weights of the various sub-models. In this way, the proposed SDC-Net integrates heterogeneous sub-models to locate optimal results for multiple diseases in this study.

The stacking attempts to combine multiple sub-classifiers generated through different basic-level classifiers on a single dataset S, which can be expressed as:

$$S : s_i = (x_i, y_i) \tag{10}$$

where x_i represents the feature of each sample, i.e., s_i in the whole dataset S, while y_i is the corresponding label of each sample. In the first phase, a series of basic-level classifiers, i.e., $L_1, \cdots, L_N$ is employed as:

$$C_m = L_m(S) \tag{11}$$

where C_m is the prediction result by applying one basic-level classifier L_m on dataset S. Afterwards, a *meta*-level classifier is used to combine all outputs $(C_1, \cdots, C_N)$ of the basic level classifiers in the second phase.

Generally, the training dataset for learning the *meta*-level classifier can be generated based on two approaches: a leave-one-out or a cross validation process [39]. The basic-level classifiers are applied to almost the entire dataset S, while leaving one example for testing in leave-one-out can be expressed as:

$$\forall i = 1, \cdots, n : \forall k = 1, \cdots, N : C_k^i = L_k(S - s_i) \tag{12}$$

Then, the learned classifiers can be applied to generate a prediction, which can be represented as:

$$s_i : \widehat{y}^k = L_k^i(x_i) \tag{13}$$

Thus, the *meta*-level dataset would contain samples expressed as:

$$\left(\left(\widehat{y}_i^1, \cdots, \widehat{y}_i^n\right), y_i\right) \tag{14}$$

where the features, i.e., $\left(\widehat{y}_i^1, \cdots, \widehat{y}_i^n\right)$ represent the corresponding predictions by the basic-level classifiers, while the label is the correct label of the sample. For k-fold cross validation, subsets of one-kth of the original dataset are left out, instead of leaving out one sample once, before predicting the class label through basic-level classifiers. The algorithm of general stacking can be summarized as the following:

Algorithm 1. Stacking in leave-one-out approach

Input: Training dataset $\boldsymbol{S - s_i}$, Entire dataset $S = \{\boldsymbol{x_i}, \boldsymbol{y_i}\}_{\boldsymbol{i=1}}^{\boldsymbol{n}}$, the number of basic-level classifiers N.
Output: An ensemble classifier E

```
Step 1: Learn first-level classifiers
for m ← 1 to N do
    Learn a base classifier L_m based on S − s_i
end for
Step 2: Construct new dataset from S − s_i
for j ← 1 to n do
    Construct a new dataset containing {X'_j, y_j}, where
    X'_j = (ŷ_j^1, ..., ŷ_j^n)
end for
Step 3: Learn a second-level (meta-level) classifier
Learn a new classifier L' based on the newly constructed dataset
return E = L'(L_1(x), L_2(x), ··· L_m(x))
```

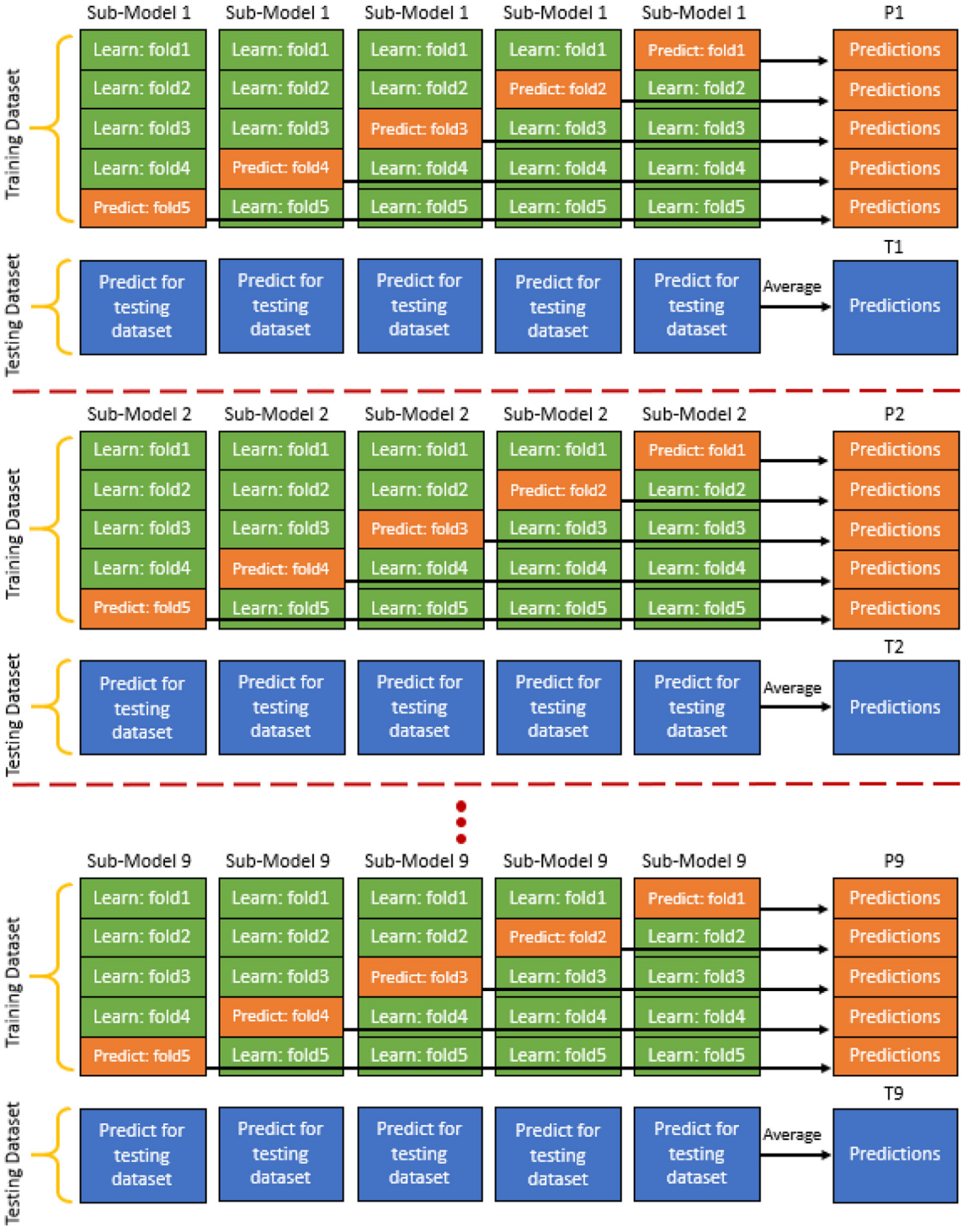

Fig. 7. The stacking procedures to generate new predicted training/testing datasets in SDC-Net.

For the SDC-Net used in this study, the dataset of extracted high-level features with corresponding labels from SFE-Net was randomly split to a testing dataset (30% of all the dataset, i.e., 200 samples) and a training dataset (the rest of all the dataset, i.e., 465 samples), where 5-fold training was applied as is shown in Fig. 7.

It can be noted that Sub-Model 1 is repeated five times for training and prediction in each round of 5-fold training. Each time, the whole training dataset can generate 372 new samples for a sub-training dataset (containing 4 folds of the original training dataset) and 93 new samples for a sub-testing dataset (containing 1 fold of the original training dataset). We applied Sub-Model 1 for training with the new sub-training dataset, before predicting the 93 samples of the sub-testing dataset.

After repeating the procedures mentioned above five times, we can obtain a series of predicted samples, i.e., 93*5 = 465 that coincides with the size of the original training dataset. These predicted samples were generated by Sub-Model 1, which can be regarded as the source of training for the second layer of SDC-Net. Please note that the predicted samples here can be transformed into a vector (465*1), before being recorded as P1 for further processing in the following steps.

For the testing dataset (the blue boxes) shown in Fig. 7, Sub-Model 1 with a sub-training dataset (containing 372 samples) was trained to predict all of the testing dataset, before generating the predicted testing results with 200 samples. Repeating the procedures mentioned above procedures five times, we can produce a series of predicted samples from the testing dataset, i.e., 200*5. Then, by averaging the predicted values based on the rows, we can obtain an average predicted testing vector (200*1), which can be recorded as T1. Here, the Sub-Model 1 has accomplished its mission—generating predicted vectors: P1, T1.

Please note that we applied nine classifiers, i.e., nine sub-models: random forest, extra tree classifier, gradient boosting classifier, *k*-NN, SVM, quadratic discriminant analysis (QDA), linear discriminant analysis (LDA), naïve bayes, and decision trees [40,41,42] in the first layer of SDC-Net to transform the original training and testing datasets. There are other sub-models, such as Sub-Model 2, in the first layer. Similarly, we can obtain the corresponding P2 and T2 of Sub-Model 2. Thus, we can produce the predicted training vectors: P1, . . ., P9, and the predicted testing vectors: T1, . . ., T9. These will be used to construct a predicted training matrix P-all (465*9) and a predicted testing matrix T-all (200*9) for further processing in the second layer of SDC-Net.

The second layer of SDC-Net only contains a *meta*-level classifier, i.e., logistic regression classifier [40], in order to evaluate all outputs from the sub-models of the first layer. The 5-fold predicted training matrix (P-all) and testing matrix (T-all) are regarded as the new training data and testing data for training/testing this *meta*-level classifier, respectively. The overall diagram of the stacking procedures in SDC-Net is expressed in Fig. 8. Here, we implemented the second layer of SDC-Net with raw extracted features from SFE-Net as well as the predicted P-all and T-all, as shown in the red line directing to the *meta*-level classifier in Fig. 8. This intuition is inspired by another researcher's work [33], which used additional *meta*-features to address the degradation problem and improve the performance of their proposed model. A similar methodology was applied to our proposed SDC-Net. It is believed to rescan effective relationships that may be ignored in generating new predicted training/testing data from the first layer, aiming to improve the final performance of detecting the different classes. Besides this, the SDC-Net aims to integrate various classifiers and fuse the effective information together. Thus, the input is the extracted high-level features from the SFE-Net, with the learned parameters are the predictions, which can be found in Figs. 7 and 8.

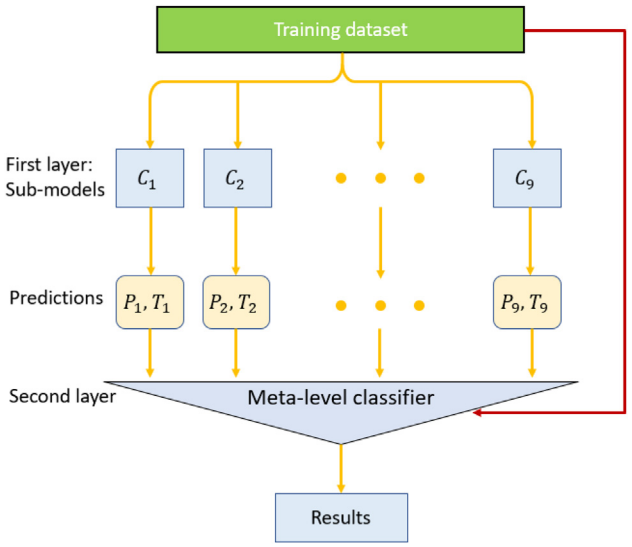

Fig. 8. The diagram of classifying different classes using SDC-Net. The training dataset is the extracted features from SFE-Net, while the results are the predicted labels of diabetes, CKD, and healthy samples.

3. Experimental results

3.1. Experimental settings

Four facial key blocks, i.e., FHB, NBB, LCB, and RCB, which were introduced in Section 2.1, were used to represent the whole facial image of a candidate. Here, each single block of one candidate was resized to 32*32 pixels, before we combined all four blocks together (FHB + NBB + LCB + RCB) to generate a single input image (64*64 pixels) for our proposed DsNet.

In SFE-Net, all the samples (665 samples with a size of 64*64) in our dataset were regarded as input data for training. As SEF-Net is based on the SSAE framework, which is an unsupervised learning approach. All input facial samples are regarded as input training data and used to generate the corresponding high-level extracted features for SDC-Net. The first to third hidden layers of SFE-Net contains 4000, 1000, and 400 hidden units, while the output corresponds to high-level representative features for further analysis in SDC-Net (refer to Fig. 1). The input samples, i.e., facial key blocks, can be considered as x in the Eq. (1). Moreover, the process of extracting features can be denoted in the Eqs. (6) and (7), also shown in Fig. 6, while the final extracted high-level features can be regarded as $h^{(3)}$, which is introduced in Section 2.2.2 Stacked sparse autoencoder. The penalty coefficient for the KL-divergence term is $v = 0.1$, while the penalty for the average activation is $\mu = 4$. For the SDC-Net, this subnetwork is a supervised learning method. Thus, the extracted high-level representative features with corresponding labels from SFE-Net were randomly split into a training dataset with 465 samples and a testing dataset with 200 samples, respectively (Totally 665 samples from our proposed dataset, containing three classes, i.e., diabetes, CKD and healthy. Refer to Section 2.1 Facial Image Dataset). All sub-models in the first (9 different classifiers) and second (*meta*-level classifier) layers of SDC-Net (as mentioned in Section 2.2.3 Stacked Disease Classification Network) were fine-tuned to obtain the best performance in this study. The x_i, y_i in Eq. (10) indicate the high-level extracted features from SFE-Net and its corresponding label, while $C_1 - C_9$ denote the nine sub-models used in Fig. 8, which are aimed to generate corresponding predictions P and T for a *meta*-level classifier, before producing the predicted labels of various diseases from healthy samples. All sub-models in the first (9 different classifiers) and second (*meta*-level classifier) layers of SDC-Net (as mentioned in Section 2.2. 3 Stacked Disease Classification Network) were fine-tuned to obtain the best performance in this study.

Experiments for the proposed model were randomly repeated 12 times on our collected dataset (please see Section 2.1). All experiments were implemented on a PC with a quad-core Intel(R) i7-4770 (3.40 GHz) CPU and 32 GB RAM. To avoid the slight imbalance and overfitting problem in this dataset, nine metrics, i.e., Accuracy, Specificity, Error Rate, Precision, Macro-Precision, Recall, Macro-Recall, F1 score, Macro-F1 score, and a Confusion matrix [43,44] were evaluated on the final results of DsNet. The confusion matrix is a tool for visualization, commonly used in supervised learning, which is efficient in determining whether a classifier fails to differentiate two different classes. Furthermore, Specificity, Error Rate, Precision, Recall, F1 score, Macro-Precision, Macro-Recall, and Macro-F1 score can be calculated as:

$$Specificity = \frac{TN}{TN + FP} \tag{15}$$

$$ErrorRate = \frac{FP + FN}{TP + TN + FP + FN} \tag{16}$$

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

$$F1 - score = \frac{2TP}{2TP + FN + FP} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{19}$$

$$Macro - Precision = \frac{\sum_{i=1}^{n} Precision_i}{n} \tag{20}$$

$$Macro - Recall = \frac{\sum_{i=1}^{n} Recall_i}{n} \tag{21}$$

$$Macro - F1score = \frac{\sum_{i=1}^{n} F1 - score_i}{n} \tag{22}$$

where TP is defined as the number of correctly identified samples in each class, i.e., true positive. FP, TN, and FN are expressed as false positive, true negative, and false negative errors, respectively.

3.2. Ablation study

We aim to evaluate the effectiveness of different components of our proposed model and to compare with the traditional color feature extraction method in this section. The same training/testing dataset settings were run 12 times for the following configurations:

1) $DsNet_{ab}$: DsNet without the additional *meta*-features in SDC-Net
2) $SFE-Net_{ab}$: Only applying SFE-Net to extract high-level features. The nine classifiers in the first layer and one classifier in the second layer of SDC-Net, i.e., ten different classifiers were evaluated separately.
3) $Color_{ab}$: Only using conventional color features [36] from the facial images. Ten different classifiers mentioned before and some state-of-the-art models, such as the convolutional neural network (CNN) and Sparse representation classifier (SRC) [45,46] were evaluated separately. The CNN model used contained five convolutional layers and two fully connected layers to generate the final results using the Adam optimizer [47].

All of the experimental results for the above configurations were evaluated by nine metrics mentioned in Section 3.1.

3.3. Extensive results

3.3.1. Proposed DsNet

Table 1 shows the corresponding results of Precision, Recall, and F1 score on the three classes, i.e., healthy, diabetes, and chronic kidney disease applying our proposed method. The average Accuracy and average Error Rate of detecting the three classes is 96.17 ± 0.88% and 3.61 ± 0.91%. In addition, the Macro-Precision, Macro-Recall, and Macro-F1 score are 96.05 ± 0.91%, 96.10 ± 0.83%, and 96.04 ± 0.88%, respectively. The corresponding confusion matrix is plotted as Fig. 9(a) in order to evaluate the quality of DsNet. It can be seen that most of the predicted outputs are correct with reference to true labels.

3.3.2. Proposed $DsNet_{ab}$

Here, we applied $DsNet_{ab}$, which is mentioned in Section 3.2, to evaluate the effectiveness of the additional *meta*-features. Table 2 represents the corresponding results of Precision, Recall, and F1 score for the three classes, i.e., healthy, diabetes, and chronic kidney disease using $DSnet_{ab}$. The average Accuracy and Error Rate of detecting the three classes along with Macro-Precision, Macro-Recall and Macro-F1 score are 95.58 ± 0.67%, 4.15 ± 0.69%, 95.45 ± 0.68%, 95.54 ± 0.60%, and 95.46 ± 0.65%, respectively. Here, the results are a little bit lower than the above-mentioned DsNet, validating the validity of additional *meta*-features. The corresponding confusion matrix is shown in Fig. 9(b), with lower correctly predicted outputs, compared to DsNet, further proving the performance of our proposed approach.

3.3.3. Proposed$SFE-Net_{ab}$

To demonstrate the effectiveness of the stacking technique in SDC-Net, we selected the extracted high-level features from SFE-Net as input for ten different classifiers instead of using SDC-Net directly. Table 3 depicts the corresponding results of Precision, Recall, and F1 score for the three classes via random forest, extra tree classifier, gradient boosting classifier, *k*-NN, SVM, QDA, LDA, naïve bayes, decision tree, and logistic regression (LR) separately. Furthermore, the average Accuracy of detecting the three classes and related Macro-Precision, Macro-Recall, and Macro-F1 score with the ten classifiers are also presented in this table. It can be observed that all of these metrics for evaluating the proposed $SFE-Net_{ab}$ are lower with varying degrees compared to the proposed DsNet and the ablation study in $DsNet_{ab}$, confirming the importance of SDC-Net in this study. The corresponding confusion matrix of $SFE-Net_{ab}$ with higher wrongly predicted outputs compared to DsNet and $DsNet_{ab}$ is shown in Fig. 10.

3.3.4. $Color_{ab}$

To verify the performance of extracting effective high-level features from SFE-Net, we used a color feature extraction method from our previous work [34,36] as input to the ten different classifiers. Table 4 illustrates the corresponding results of Precision, Recall, F1 score, average Accuracy, Macro-Precision, Macro-Recall, and Macro-F1. Here, we can easily notice that all these metrics are much lower than the aforementioned experiments, proving the significance of SFE-Net, which can cap-

Table 1
The Specificity, Precision, Recall and F1 score with standard deviation using our proposed DsNet for detecting healthy, diabetes, and chronic kidney disease, respectively.

Class	Specificity	Precision	Recall	F1 score
Healthy	94.32 ± 1.35%	96.17 ± 1.17%	93.50 ± 2.07%	94.67 ± 1.03%
Diabetes	97.97 ± 1.01%	97.83 ± 1.33%	98.49 ± 1.64%	98.17 ± 0.75%
CKD	94.51 ± 1.49%	94.50 ± 3.02%	96.50 ± 1.05%	95.33 ± 1.37%

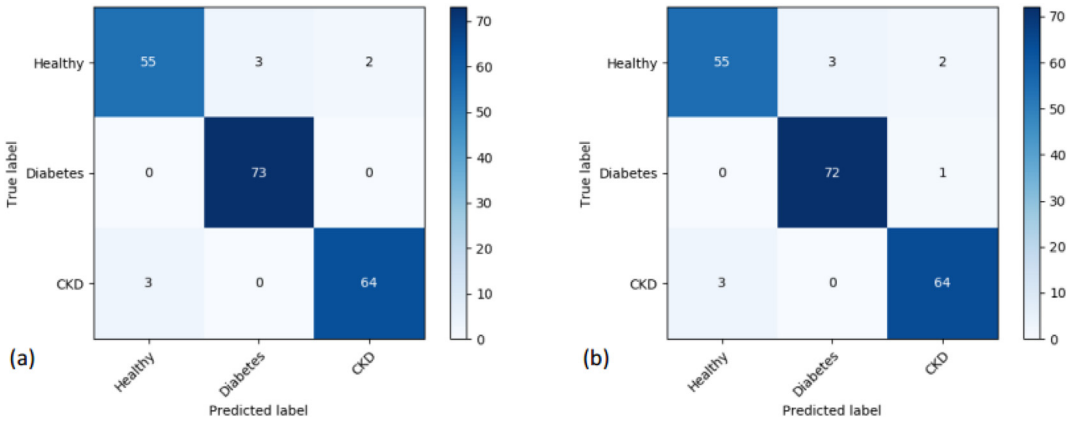

Fig. 9. The sample confusion matrices of applying (a) proposed DsNet and (b) proposed $DsNet_{ab}$.

Table 2
The Specificity, Precision, Recall and F1 score with standard deviation using our proposed $DsNet_{ab}$ for detecting healthy, diabetes, and chronic kidney disease, respectively.

Class	Specificity	Precision	Recall	F1 score
Healthy	93.13 ± 1.52%	95.33 ± 1.03%	92.50 ± 1.64%	93.83 ± 0.75%
Diabetes	97.70 ± 1.74%	97.67 ± 1.51%	97.83 ± 2.14%	97.67 ± 1.03%
CKD	94.19 ± 1.81%	93.83 ± 2.40%	96.17 ± 1.17%	94.83 ± 0.98%

ture more useful features than the traditional color feature extraction approach. Besides this, the corresponding confusion matrix shown in Fig. 11 also supports this conclusion.

As our proposed model and other conventional models are repeated 12 times to generate the final results evaluated by various metrics in this study. Thus, we select the accuracy to perform the paired t-test with two-tails [48] to verify the validity and reliability of our results. The LDA and CNN models obtained the best performances among the $SFE - Net_{ab}$ and $Color_{ab}$, respectively. Therefore, the LDA, CNN, and $DsNet_{ab}$ are used to compare with our proposed DsNet by performing a t-test.

The p-value of the LDA, CNN, and $DsNet_{ab}$ versus our proposed DsNet are $4.91 \times 10^{-4}, 2.70 \times 10^{-12}$, and 2.46×10^{-2}, respectively. Consequently, the statistically significant differences can be found in our proposed method compared to other state-of-the-art approaches in terms of accuracy.

4. Discussion

We proposed a progressive network, i.e., DsNet, which can extract high-level representative features to perform the detection of diabetes mellitus and chronic kidney disease from healthy samples simultaneously by only using facial images. The highest average accuracy of our proposed method is 96.17 ± 0.88%, which compares to 95.58 ± 0.67%, 93.33 ± 0.89%, and 75.94 ± 0.79% for $DsNet_{ab}$, $SFE - Net_{ab}$, and $Color_{ab}$, respectively. The other metrics in Tables 2 and 4 and Figs. 9–11, such as F1 score, Macro-F1 score, Recall, and so forth, also support the effectiveness of different components in our proposed model. Besides this, compared to the other works in non-invasive disease detection, [27] obtained an average accuracy of 73% in identifying diabetes, [21] achieved 86.07% in the detecting accuracy, which is lower than our proposed method (96.17%), showing the superiority of our model. Here, we have also applied our proposed method for detecting heart disease and brain disease with the same samples and compared it to our previous work [17,36]. Employing the same experimental settings as the previous method, our proposed DsNet obtained an average accuracy of 93.47% and a specificity of 94.03%, compared to the average accuracy of 88.01% and 91.07% for [36] in classifying heart disease. In addition, our method achieved an average accuracy of 96.28% and a specificity of 96.11%, compared to the average accuracy of 95.00% and 95.67% in [17] in classifying brain disease, which also proves the effectiveness and robustness of our proposed approach. Here, we also have applied our proposed method for detecting heart disease and brain disease with the same samples and compared to our previous work [16,36]. The experiment settings are the same as the previous method. Our proposed DsNet obtains the average accuracy of 93.47% and specificity of 94.03%, compared to the average accuracy of 88.01% and 91.07% of [36] in classifying heart disease. And our method achieves the average accuracy of 96.28% and specificity of 96.11%, compared to the average accuracy of 95.00% and 95.67% of [16] in classifying brain disease, which also proves the effectiveness and robustness of our proposed approach.

Table 3
The Specificity, Precision, Recall, F1 score, Accuracy, Error Rate (ER), Macro-Precision (MP), Macro-Recall (MR), and Macro-F1score (MF) with standard deviation using our proposed $SFE - Net_{ab}$ for detecting healthy (H), diabetes (DM), and chronic kidney disease (CKD), respectively.

Class	Random forest	Extra tree	Gradient boosting	*k*-NN	SVM	QDA	LDA	Naïve bayes	Decision tree	LR
Specificity using the 10 classifiers separately										
H	90.17 ± 0.72%	89.30 ± 2.31%	88.82 ± 1.89%	87.83 ± 4.47%	79.60 ± 2.63%	78.49 ± 4.30%	88.36 ± 2.12%	86.53 ± 4.37%	88.31 ± 4.16%	85.13 ± 3.51%
DM	94.52 ± 0.73%	94.13 ± 1.29%	93.80 ± 1.72%	92.32 ± 3.03%	78.32 ± 3.40%	82.15 ± 4.91%	90.68 ± 1.63%	87.14 ± 3.60%	90.19 ± 3.83%	86.31 ± 2.30%
CKD	94.03 ± 1.55%	94.15 ± 1.63%	93.24 ± 1.89%	93.21 ± 2.43%	80.19 ± 4.36%	83.10 ± 4.64%	90.35 ± 2.70%	86.60 ± 3.87%	90.45 ± 4.08%	89.45 ± 3.10%
Precision using the 10 classifiers separately										
H	89.68 ± 0.69%	88.92 ± 3.01%	89.07 ± 1.90%	87.52 ± 7.29%	99.43 ± 0.60%	74.27 ± 5.61%	89.57 ± 3.09%	88.71 ± 5.39%	87.77 ± 4.09%	94.90 ± 2.63%
DM	95.44 ± 0.77%	95.01 ± 1.46%	94.14 ± 1.77%	92.46 ± 2.51%	74.88 ± 4.45%	83.76 ± 8.85%	96.61 ± 1.56%	86.00 ± 2.85%	90.8 ± 4.17%	79.24 ± 3.54%
CKD	93.81 ± 2.46%	94.78 ± 1.70%	93.82 ± 2.53%	94.17 ± 2.59%	89.20 ± 4.18%	92.44 ± 0.78%	92.94 ± 4.69%	86.35 ± 4.62%	91.2 ± 5.05%	89.69 ± 3.47%
Recall using the 10 classifiers separately										
H	91.36 ± 2.33%	91.89 ± 1.53%	90.24 ± 2.84%	87.55 ± 6.94%	57.92 ± 6.67%	87.47 ± 4.38%	91.00 ± 2.22%	72.64 ± 5.66%	86.1 ± 4.17%	64.97 ± 5.86%
DM	96.51 ± 2.33%	96.29 ± 0.98%	95.01 ± 1.06%	97.40 ± 2.26%	99.13 ± 0.75%	77.97 ± 12.35%	99.59 ± 0.71%	91.27 ± 1.94%	92.2 ± 2.74%	97.11 ± 1.74%
CKD	90.78 ± 3.42%	89.85 ± 4.79%	91.41 ± 3.29%	88.85 ± 6.33%	95.62 ± 0.76%	81.26 ± 4.33%	88.02 ± 3.01%	97.45 ± 3.08%	91.2 ± 3.26%	96.26 ± 2.15%
F1 score using the 10 classifiers separately										
H	90.51 ± 1.46%	90.35 ± 1.42%	89.64 ± 2.06%	87.41 ± 3.42%	73.20 ± 5.32%	80.30 ± 4.87%	90.23 ± 0.59%	79.78 ± 4.68%	86.8 ± 3.26%	77.04 ± 4.38%
DM	95.96 ± 0.96%	95.64 ± 5.74%	94.56 ± 0.69%	94.83 ± 0.88%	85.26 ± 2.72%	80.70 ± 10.72%	98.07 ± 0.69%	88.55 ± 2.16%	91.5 ± 3.36%	87.26 ± 2.75%
CKD	92.21 ± 0.71%	92.18 ± 2.14%	92.55 ± 1.61%	91.29 ± 2.27%	92.25 ± 1.95%	86.46 ± 2.71%	90.31 ± 0.86%	91.46 ± 1.66%	91.1 ± 1.30%	92.83 ± 2.17%
Accuracy, Macro-Precision, Macro-Recall and Macro-F1 score (Multi-class classification)										
Accuracy	93.03 ± 0.50%	92.83 ± 1.26%	92.33 ± 1.26%	91.33 ± 2.08%	84.50 ± 8.67%	82.50 ± 5.77%	93.33 ± 0.89%	87.03 ± 1.32%	90.0 ± 1.73%	86.33 ± 1.16%
ER	6.52 ± 0.71%	6.81 ± 0.73%	7.14 ± 0.89%	8.43 ± 1.04%	15.10 ± 2.39%	19.01 ± 5.86%	7.18 ± 0.90%	12.45 ± 1.55%	9.23 ± 1.60%	13.63 ± 1.91%
MP	92.96 ± 0.89%	92.90 ± 1.26%	92.34 ± 1.52%	91.38 ± 1.88%	88.03 ± 1.12%	83.49 ± 4.79%	92.91 ± 0.96%	87.02 ± 1.46%	89.9 ± 1.79%	87.95 ± 0.97%
MR	92.88 ± 4.56%	92.67 ± 1.25%	92.22 ± 1.36%	91.27 ± 2.17%	84.22 ± 2.34%	82.23 ± 6.09%	92.87 ± 0.92%	87.12 ± 2.09%	89.8 ± 1.79%	86.11 ± 2.30%
MF	92.90 ± 0.64%	92.72 ± 1.31%	92.25 ± 1.44%	91.18 ± 2.16%	83.57 ± 1.75%	82.49 ± 5.68%	93.04 ± 0.98%	86.60 ± 1.79%	89.8 ± 1.74%	85.71 ± 1.76%

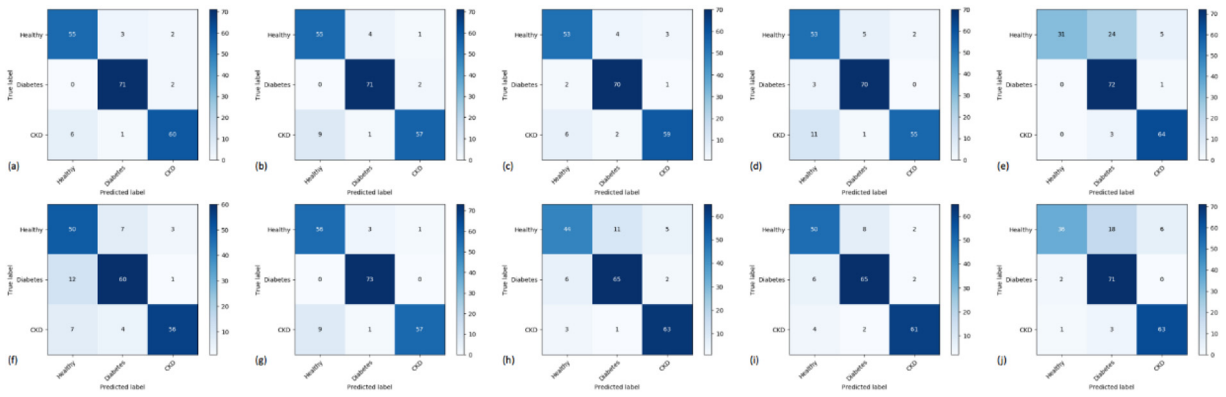

Fig. 10. The sample confusion matrices of (a) random forest, (b) extra tree classifier, (c) gradient boosting classifier, (d) *k*-NN, (e) SVM, (f) QDA, (g) LDA, (h) naïve bayes, (i) decision tree, and (j) logistic regression applied separately using $SFE - Net_{ab}$.

The subnetwork: SFE-Net is implemented based on SSAE to obtain high-level features in this study. Here, we synthesized three basic SAEs, i.e., three encoders (three hidden layers with 4000, 1000, and 400 hidden units) by concatenating them together to capture effective features from the input facial key blocks. We also attempted to combine two, four, or five SAEs together in this study to evaluate its performance. The results, when compared to three SAEs were slightly worse. Due to the limited space of this paper, we did not post all the related results. The reason why we selected three SAEs is due to the compression ratio, i.e., the ratio of the number of different hidden units in connecting with the upper/lower basic SAE. The compression ratio in our proposed SFE-Net is 4 and 2.5 for the first/second and second/third hidden layers (refer to Section 3.A), which is effective at capturing useful high-level features. Too large a compression ratio in the two hidden layers of SFE-Net may lose numerous useful latent information, generating poor high-level features compared to the three layers approach. Moreover, too small a compression ratio with four or five hidden layers of SFE-Net may generate invalid noisy information between different hidden layers, disturbing effective information distillation, and reducing the performance of high-level feature extraction. We also assessed different compression ratios in the proposed SFE-Net with three hidden layers. The related results indicate that the compression ratio between two connected hidden layers ranging from 2 to 6 is reasonable.

The subnetwork: SDC-Net is constructed based on a stacking framework with additional *meta*-features in this study. This subnetwork can process the extracted high-level features from SFE-Net, before producing the results of detecting diabetes and CKD as well as healthy samples. The stacking framework used here is a representation learning one, which can further find effectively different information through heterogeneous classifiers automatically [39]. Therefore, the classifiers in the stacking framework should be as different as possible while still maintaining an ideal performance. With this requirement, we selected nine classifiers based on six different types: 1. random forest, extra tree classifier, decision tree; 2. *k*-NN; 3. SVM; 4. QDA, LDA; 5. naïve bayes; 6. gradient boosting classifier, respectively, to improve the final performance as much as possible. From the experiments of $SFE - Net_{ab}$ in the ablation study, the performance of each selected classifier was acceptable, satisfying the basic requirement of an ideal stacking framework. The second layer of SDC-Net used only one classifier, i.e., logistic regression, to generate the final results. The reason we used this classifier was aimed to decrease the risk of overfitting. The relatively simple classifier, such as logistic regression [40] may be a good choice. As we have applied complex nonlinear transformations in the first layer of SDC-Net, the output layer does not need a model with very complicated functions. The logistic regression is a simple classifier, with which we can apply L1 regularization in preventing the overfitting problem and select effective features from the first layer. The results in Section 3.C: also support these views.

It can be inferred that the ideal performance of SDC-Net does not originate from the effect of the multi-layer stacking approach, while coming from the integrated learning ability of different classifiers for different aspects of the input data. This is why we only choose two layers for SDC-Net—more layers may increase the corresponding risk of overfitting, obstructing our proposed model's whole performance. Actually, we also investigated adding more stacking layers in SDC-Net to evaluate its performance. Extended results indicated that the relevant results were the same or slightly lower than the above-mentioned method.

In $Color_{ab}$ part, several state-of-the-art models such as CNN or SRC were evaluated by applying the color features from the facial key blocks. The highest performance (Accuracy: 75.94 ± 0.79%, F1-score: 75.20 ± 0.90%) was obtained by applying the CNN model, followed by an LDA classifier. It can be noted that the performance of only applying color features from the facial images with the single classifier is worse than our proposed method (Accuracy: 96.17% vs. 75.94%), which validates the effectiveness and robustness of our approach.

The reason of designing our proposed DsNet with two sub-networks, i.e., SFE-Net and SDC-Net, is: the SFE-Net is good at extracting high-level representative features from facial images, without requiring hand-crafted feature extraction; the SDC-Net can synthesize the results of various classifiers by using a stacking framework. These two sub-networks can perform

Table 4
The Specificity, Precision, Recall, F1 score, Accuracy, Error Rate (ER), Macro-Precision (MP), Macro-Recall (MR), and Macro-F1score (MF) with standard deviation using our proposed $Color_{ab}$ for detecting healthy, diabetes, and chronic kidney disease, respectively.

Class	Random forest	Extra tree	Gradient boosting	*k*-NN	SVM	QDA	LDA	Naïve bayes	Decision tree	LR	CNN	SRC
Specificity using the 12 classifiers separately												
H	69.97 ± 2.78%	75.32 ± 3.35%	77.57 ± 2.43%	57.28 ± 5.36%	67.10 ± 2.32%	55.37 ± 6.89%	76.19 ± 3.06%	43.55 ± 7.12%	62.32 ± 5.02%	72.13 ± 2.28%	77.03 ± 2.92%	72.34 ± 2.73%
DM	68.41 ± 3.32%	71.14 ± 4.15%	72.22 ± 3.14%	58.31 ± 4.10%	69.14 ± 3.08%	57.23 ± 7.80%	77.80 ± 3.69%	60.14 ± 6.39%	64.18 ± 4.21%	73.59 ± 2.10%	77.51 ± 2.23%	73.53 ± 2.60%
CKD	68.20 ± 3.85%	74.20 ± 4.01%	71.46 ± 3.72%	57.01 ± 4.72%	66.91 ± 3.81%	51.37 ± 8.12%	71.21 ± 3.93%	52.18 ± 7.09%	58.30 ± 5.92%	69.36 ± 3.12%	73.05 ± 3.80%	69.98 ± 3.39%
Precision using the 12 classifiers separately												
H	72.66 ± 3.61%	74.19 ± 5.10%	78.59 ± 1.93%	53.79 ± 4.25%	65.39 ± 3.71%	62.90 ± 7.13%	73.32 ± 3.23%	73.46 ± 5.67%	65.94 ± 6.97%	71.05 ± 1.77%	75.16 ± 2.09%	73.09 ± 2.81%
DM	67.10 ± 4.40%	70.28 ± 5.09%	69.17 ± 4.09%	58.01 ± 5.78%	71.47 ± 4.57%	51.45 ± 6.74%	78.17 ± 4.73%	44.21 ± 3.44%	62.06 ± 7.05%	75.33 ± 5.81%	78.84 ± 2.17%	77.27 ± 3.30%
CKD	72.31 ± 3.97%	75.63 ± 4.42%	71.99 ± 5.50%	66.97 ± 5.20%	68.84 ± 4.92%	57.9 ± 18.38%	72.39 ± 4.78%	63.65 ± 8.28%	52.54 ± 4.79%	65.56 ± 5.30%	72.74 ± 3.15%	69.81 ± 4.03%
Recall using the 12 classifiers separately												
H	81.84 ± 6.34%	83.07 ± 7.89%	82.05 ± 3.03%	68.84 ± 7.12%	75.60 ± 11.95%	39.04 ± 15.71%	84.52 ± 5.74%	23.18 ± 8.09%	65.92 ± 6.69%	75.51 ± 5.81%	86.13 ± 2.01%	75.96 ± 3.29%
DM	73.09 ± 3.36%	75.66 ± 4.24%	73.22 ± 3.35%	62.70 ± 9.53%	74.28 ± 5.61%	76.73 ± 17.75%	75.96 ± 3.47%	94.51 ± 1.18%	61.24 ± 3.64%	75.87 ± 3.58%	77.13 ± 2.33%	74.88 ± 3.09%
CKD	54.61 ± 5.86%	60.69 ± 2.84%	64.15 ± 5.65%	42.08 ± 6.32%	56.68 ± 2.72%	36.14 ± 25.95%	63.70 ± 1.04%	24.84 ± 6.38%	53.51 ± 3.70%	61.24 ± 5.17%	65.26 ± 3.27%	63.51 ± 3.60%
F1 score using the 12 classifiers separately												
H	76.82 ± 2.51%	78.10 ± 3.23%	80.23 ± 0.45%	60.09 ± 1.94%	69.83 ± 5.86%	46.93 ± 12.90%	78.36 ± 0.87%	34.50 ± 8.82%	65.59 ± 3.64%	73.10 ± 2.28%	79.14 ± 1.48%	74.48 ± 2.13%
DM	69.82 ± 1.28%	72.68 ± 1.32%	71.02 ± 1.12%	59.72 ± 3.81%	72.64 ± 1.88%	60.24 ± 2.55%	76.90 ± 0.70%	60.19 ± 3.18%	61.61 ± 5.23%	75.42 ± 2.02%	76.43 ± 1.99%	76.79 ± 2.37%
CKD	62.05 ± 3.45%	67.27 ± 3.50%	67.72 ± 4.0.47%	51.55 ± 5.56%	62.45 ± 1.17%	38.23 ± 11.10%	67.73 ± 2.63%	35.01 ± 4.69%	52.90 ± 3.03%	63.31 ± 5.14%	68.54 ± 3.38%	66.32 ± 3.23%
Accuracy, Macro-Precision, Macro-Recall and Macro-F1 score (Multi-class classification)												
Accuracy	70.33 ± 0.76%	73.00 ± 2.29%	73.17 ± 1.61%	57.83 ± 2.26%	68.67 ± 2.89%	52.17 ± 3.22%	74.83 ± 0.58%	49.33 ± 2.93%	60.33 ± 4.07%	71.17 ± 1.89%	75.94 ± 0.79%	73.85 ± 2.07%
ER	29.52 ± 0.89%	26.81 ± 1.95%	27.01 ± 1.70%	44.00 ± 2.49%	31.42 ± 2.68%	48.04 ± 3.40%	25.19 ± 0.82%	49.32 ± 3.07%	39.36 ± 3.49%	28.98 ± 1.93%	24.57 ± 0.86%	26.31 ± 1.80%
MP	70.69 ± 1.07%	73.31 ± 2.40%	73.25 ± 1.76%	59.59 ± 2.76%	68.90 ± 3.16%	57.43 ± 4.55%	74.53 ± 1.03%	60.44 ± 1.84%	60.18 ± 3.64%	70.65 ± 1.96%	75.82 ± 1.12%	73.22 ± 2.17%
MR	69.85 ± 0.14%	73.14 ± 2.75%	73.14 ± 2.13%	57.88 ± 2.88%	68.85 ± 3.19%	50.64 ± 5.98%	74.72 ± 1.07%	47.50 ± 1.18%	60.22 ± 3.81%	70.87 ± 2.01%	75.97 ± 1.16%	72.19 ± 2.38%
MF	69.57 ± 0.16%	72.68 ± 2.59%	72.99 ± 2.01%	57.12 ± 2.49%	68.31 ± 2.77%	48.47 ± 6.54%	74.43 ± 0.81%	43.23 ± 2.09%	60.03 ± 3.76%	70.61 ± 2.03%	75.20 ± 0.90%	72.91 ± 2.11%

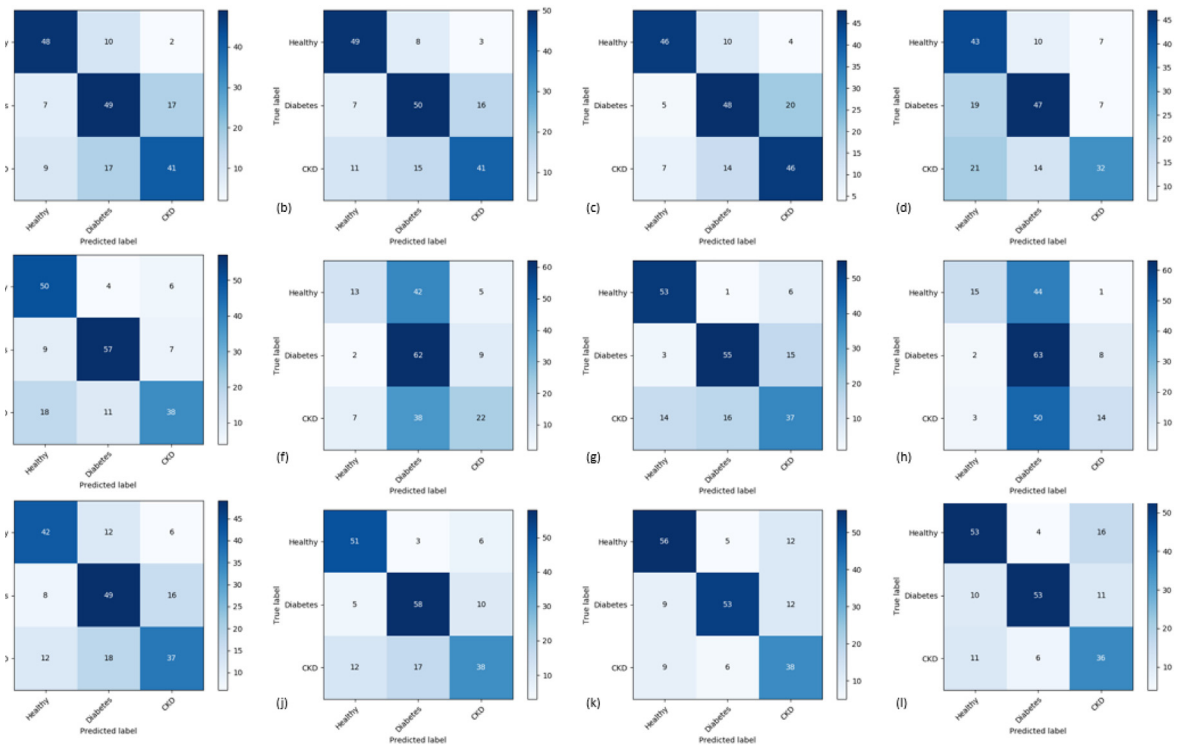

Fig. 11. The sample confusion matrices of (a) random forest, (b) extra tree classifier, (c) gradient boosting classifier, (d) *k*-NN, (e) SVM, (f) QDA, (g) LDA, (h) naïve bayes, (i) decision tree, (j) logistic regression, (k) CNN, and (l) SRC applied separately using.$Color_{ab}$

their own functions, and do not interfere with each other. Therefore, we have constructed this intelligent healthcare system based on this idea (shown in Fig. 1). This smart system can be made into an end-to-end model, which can utilize the facial images captured by our unique imaging device [34], before automatically generating the healthy status of each candidate. The superior experimental results (Table 1) of DsNet compared to other state-of-the art methods can also validate the effectiveness of our design for this healthcare system. There are two main originalities of our proposed DsNet. The first originality has been expressed in the Introduction, which elaborated several contributions of the system design. The second originality is that this model can provide a positive and useful effect in the prevention of diabetes mellitus and CKD by using the facial images, which is much more convenient and efficient compared to conventional medical examinations. At the same time, our proposed method can be used to detect two different diseases (DM and CKD) occurring in different parts of the body.

The main limitation of our proposed DsNet is using SFE-Net to extract effective features before applying SDC-Net to evaluate the extracted features and generate the disease detection results. The procedures are a bit more complex than the traditional color feature extraction method. As $Color_{ab}$ only analyzes the color features of the facial key blocks and predicts a disease with one classifier. Although the performance of our proposed method is much better than other aforementioned methods, finding some ways to simplify the procedures is desired. Another limitation is that we only evaluated 9 different classifiers in SDC-Net, while there are many other classifiers such as multilayer perceptron or different ensemble methods such as boosting, voting techniques. Further analysis of more classifiers and ensemble methods for detecting various diseases are warranted in the future.

Some researchers attempted to synthesize a stacking framework with a genetic algorithm to solve specific problems and receive satisfactory results [49]. Their practices inspire us as we can implement a stacking framework with other state-of-the-art models such as a convolution neural network [45] and other swarm intelligence [50] to improve the whole performance of our proposed model. It is an interesting topic that can be further investigated in the future.

5. Conclusion

In this paper, we presented a dual stack network for detecting diabetes mellitus and CKD from healthy candidates simultaneously. In particular, we proposed the subnetwork: SFE-Net, along with a modified SSAE architecture. The subnetwork SDC-Net is based on a stacking framework with additional meta features. Extensive experimental evaluations on our pro-

posed model with other ablation studies demonstrate the superior performance of DsNet. The proposed model can extract high-level features from facial key blocks efficiently, obtaining the highest results by applying nine metrics on our collected dataset with 229 healthy candidates, 236 diabetes, and 200 CKD patients, respectively. Moreover, our future work on this model includes collecting more samples from different types of diseases, such as fatty liver and brain disease, further evaluating its effectiveness. Besides this, more classifiers, ensemble methods, or other feature fusion methods for detecting different diseases simultaneously will be investigated in the future.

CRediT authorship contribution statement

Qi Zhang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft. **Jianhang Zhou:** Software, Formal analysis, Data curation, Investigation. **Bob Zhang:** Resources, Writing - review & editing, Supervision. **Enhua Wu:** Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the University of Macau (File no. MYRG2019-00006-FST).

References

[1] Centers for Disease Control and Prevention, National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States, 2011, Atlanta, GA: US department of health and human services, centers for disease control and prevention, vol. 201, no. 1, 2011, pp. 2568-2569.
[2] World Health Organization, Diabetes, Fact sheet N 312, 2011.
[3] N. Cho et al, IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045, Diabetes Res. Clin. Pract. 138 (2018) 271–281.
[4] A.S. Levey et al, National Kidney Foundation practice guidelines for chronic kidney disease: evaluation, classification, and stratification, Ann. Intern. Med. 139 (2) (2003) 137–147.
[5] S.I. Hallan et al, International comparison of the relationship of chronic kidney disease prevalence and ESRD risk, J. Am. Soc. Nephrol. 17 (8) (2006) 2275–2284.
[6] B. Bikbov, N. Perico, G. Remuzzi, Disparities in chronic kidney disease prevalence among males and females in 195 countries: analysis of the Global Burden of Disease 2016 Study, Nephron 139 (2018) 313–318.
[7] H. Wang et al, Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015, Lancet 388 (10053) (2016) 1459–1544.
[8] A. Tirosh et al, Normal fasting plasma glucose levels and type 2 diabetes in young men, N. Engl. J. Med. 353 (14) (2005) 1454–1462.
[9] A.S. Levey et al, A new equation to estimate glomerular filtration rate, Ann. Intern. Med. 150 (9) (2009) 604–612.
[10] L.A. Stevens, J. Coresh, T. Greene, A.S. Levey, Assessing kidney function—measured and estimated glomerular filtration rate, N. Engl. J. Med. 354 (23) (2006) 2473–2483.
[11] J.C. Guzmán, I. Miramontes, P. Melin, G. Prado-Arechiga, Optimal genetic design of type-1 and interval type-2 fuzzy systems for blood pressure level classification, Axioms 8 (1) (2019).
[12] E. Ramirez, P. Melin, G. Prado-Arechiga, Hybrid model based on neural networks, type-1 and type-2 fuzzy systems for 2-lead cardiac arrhythmia classification, Expert Syst. Appl. 126 (2019) 295–307.
[13] P. Melin, I. Miramontes, G. Prado-Arechiga, A hybrid model based on modular neural networks and fuzzy systems for classification of blood pressure and hypertension risk diagnosis, Expert Syst. Appl. 107 (2018) 146–164.
[14] R.A. Gil, Z.C. Johanyák, T. Kovács, Surrogate model based optimization of traffic lights cycles and green period ratios using microscopic simulation and fuzzy rule interpolation, Int. J. Artif. Intell. 16 (1) (2018) 20–40.
[15] E.L. Hedrea, R.E. Precup, C.A. Bojan-Dragos, Results on tensor product-based model transformation of magnetic levitation systems, Acta Polytechnica Hungarica 16 (9) (2019).
[16] B. Zhang, D. Zhang, Noninvasive diabetes mellitus detection using facial block color with a sparse representation classifier, IEEE Trans. Biomed. Eng. 61 (4) (2013) 1027–1033.
[17] T. Shu, B. Zhang, Y. Tang, Novel noninvasive brain disease detection system using a facial image sensor, Sensors 17 (12) (2017) 2843.
[18] Z. David, Z. Wangmeng, Medical Biometrics: Computerized TCM Data Analysis, World Scientific, 2016.
[19] B. Zhu, H. Wang, Diagnostics of traditional Chinese medicine, Singing, Dragon, 2011.
[20] X. Wang, D. Zhang, A high quality color imaging system for computerized tongue image analysis, Expert Syst. Appl. 40 (15) (2013) 5854–5866.
[21] J. Li, D. Zhang, Y. Li, J. Wu, B. Zhang, Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection, Inf. Sci. 384 (2017) 191–204.
[22] J. Ma, G. Wen, C. Wang, L. Jiang, Complexity perception classification method for tongue constitution recognition, Artif. Intell. Med. 96 (2019) 123–133.
[23] D. Dai, J. Tang, Z. Yu, H.S. Wong, J. You, W. Cao, Y. Hu, C.P. Chen, An inception convolutional autoencoder model for chinese healthcare question clustering, IEEE Trans. Cybern. (2019).
[24] Y. Hu, G. Wen, H. Liao, C. Wang, D. Dai, Z. Yu, Automatic construction of chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics, IEEE Trans. Cybern. (2019).
[25] S. Gitelson, N. Wertheimer-Kaplinski, Color of the face in diabetes mellitus: observations on a group of patients in Jerusalem, Diabetes 14 (4) (1965) 201–208.
[26] O. Falodun, A. Ogunbiyi, B. Salako, A.K. George, Skin changes in patients with chronic renal failure, Saudi J. Kidney Dis. Transpl. 22 (2) (2011) 268.
[27] L. Zhang, B. Zhang, Non-Invasive Multi-Disease Classification via Facial Image Analysis Using a Convolutional Neural Network, in: 2018 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR), 2018, pp. 66–71.
[28] A. Ng, Sparse autoencoder, CS294A Lecture notes 72 (2011) (2011) 1–19.

[29] W. Jia, M. Yang, S.H. Wang, Three-category classification of magnetic resonance hearing loss images based on deep autoencoder, J. Med. Syst. 41 (10) (2017) 165.
[30] W. Liu, T. Ma, D. Tao, J. You, HSAE: a Hessian regularized sparse auto-encoders, Neurocomputing 187 (2016) 59–65.
[31] D.H. Wolpert, Stacked generalization, Neural Networks 5 (2) (1992) 241–259.
[32] X. Bao, L. Bergman, R. Thompson, Stacking recommendation engines with additional meta-features, in: Proceedings of the third ACM conference on Recommender systems, 2009, pp. 109–116.
[33] X. Bao, Applying machine learning for prediction, recommendation, and integration. [Online]. Available at: https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/cf95jd85f, Accessed on: 2009
[34] T. Shu, B. Zhang, Y.Y. Tang, An improved noninvasive method to detect Diabetes Mellitus using the Probabilistic Collaborative Representation based Classifier, Inf. Sci. 467 (2018) 477–488.
[35] HK PolyU, Biometric Research Center – HK PolyU. [Online]. Available at: http://www4.comp.polyu.edu.hk/~biometrics/, Accessed on: 2014
[36] T. Shu, B. Zhang, Y. Tang, Effective heart disease detection based on quantitative computerized traditional chinese medicine using representation based classifiers, Evidence-Based Complementary and Alternative Medicine 2017 (2017).
[37] J. Xu et al, Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images, IEEE Trans. Med. Imaging 35 (1) (2015) 119–130.
[38] M.F. Moller, A scaled conjugate gradient algorithm for fast supervised learning, Neural Networks 6 (1993) 525–533.
[39] S. Džeroski, B. Ženko, Is combining classifiers with stacking better than selecting the best one?, Mach Learn. 54 (3) (2004) 255–273.
[40] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[41] O. Maier, M. Wilms, J. von der Gablentz, U.M. Krämer, T.F. Münte, H. Handels, Extra tree forests for sub-acute ischemic stroke lesion segmentation in MR sequences, J. Neurosci. Methods 240 (2015) 89–100.
[42] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[43] M. Zancanaro, B. Lepri, F. Pianesi, Automatic detection of group functional roles in face to face interactions, in: Proceedings of the 8th international conference on Multimodal interfaces, ACM, 2006, pp. 28–34.
[44] X. Deng, Q. Liu, Y. Deng, S. Mahadevan, An improved method to construct basic probability assignment based on the confusion matrix for classification problem, Inf. Sci. 340 (2016) 250–261.
[45] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.
[46] J. Yang, D. Chu, L. Zhang, Y. Xu, J. Yang, Sparse representation classifier steered discriminative projection with applications to face recognition, IEEE Trans. Neural Networks Learn. Syst. 24 (7) (2013) 1023–1035.
[47] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980.
[48] G.D. Ruxton, The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test, Behav. Ecol. 17 (4) (2006) 688–690.
[49] A. Ledezma, R. Aler, A. Sanchis, D. Borrajo, GA-stacking: evolutionary stacked generalization, Intell. Data Anal. 14 (1) (2010) 89–119.
[50] J. Kennedy, Swarm intelligence, in: Handbook of nature-inspired and innovative computing, Springer, 2006, pp. 187–219.