Diagnosis and Knowledge Discovery of Turner Syndrome Based on Facial Images Using Machine Learning Methods

JIANQiang LI, (Senior Member, IEEE), LU LIU, JINGCHAO SUN, YAN PEI, (Senior Member, IEEE), JIJIANG YANG, HUI PAN, (Senior Member, IEEE), SHI CHEN, AND QING WANG

1School of Software Engineering, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
2School of Computer Science and Engineering, University of Aizu, Aizu 965-8580, Japan
3Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China
4Department of Endocrinology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100730, China

Corresponding author: Yan Pei (peiyan@u-aizu.ac.jp)

This work was supported by the National Key Research and Development Program of China under Project 2020YFB2104402.

ABSTRACT Turner syndrome (TS) is a chromosomal disorder disease that only affects the development of females. Based on TS facial images, we propose a new TS diagnosis model, which preserves the original ratio of face width to face height, extracts reliable features and conducts feature analysis based on support vectors (SVs) and principal components (PCs). The proposed TS model is composed of Image Preprocessing, Feature Extraction, Classifier Construction and Knowledge Discovery. For Image Preprocessing, by utilizing the techniques of face alignment, facial area intercept and brightness normalization, the original facial images are processed to the desired gray aligned facial area images, while the original ratio of face width to face height remains unchanged. For Feature Extraction, by employing the energy features of facial organ blocks and ratio features roughly and more finely, five reliable feature sets are extracted, i.e., Rough Energy Features, Finer Energy Features (FEF), Rough Ratio Features, Finer Ratio Features (FRF) and FRF2. For Classifier Construction, support vector machine (SVM), principal component analysis (PCA), kernel PCA (KPCA) and ensemble learning methods are used to establish 15 single classifiers (i.e., 5 SVM classifiers, 5 PCA\texttt{+}SVM classifiers and 5 KPCA\texttt{+}SVM classifiers) and 2 ensemble classifiers. The classifier established by the weighted voting method achieves the highest accuracy of 0.9127; FEF outperforms the other four feature sets. For Knowledge Discovery, the feature analysis based on SVs and PCs is carried out to discover important features. It is found that less energy of external canthus areas and a lower ratio of forehead height to forehead width often occur in TS patients through analyzing SVs, and the energy and ratio features of left zygoma area are important in identifying TS by analyzing PCs.

INDEX TERMS Knowledge discovery, machine learning, principal components analysis, support vector machine, turner syndrome.

I. INTRODUCTION

Turner Syndrome (TS) is a chromosomal disorder disease that affects the development of females, which should be diagnosed early due to their specific symptoms. It happens in approximately 1:(2000-4000) female live births [1]. A female diagnosed as TS may exhibit one or more clinical characteristics, such as short stature, primary amenorrhea, ovarian failure, facial appearance characteristics, ear infections, renal anomalies, cardiovascular malformations and liver anomalies, which are caused by the absence of structural abnormality of one X chromosome in some or all the cells [2]. Moreover, the symptoms of short stature and primary amenorrhea are most likely to occur in TS patients and require early treatment [1], [3]. For short stature, growth hormone treatment is applied to help TS patients achieve...
adult height within the lower range for female population standards. And it is advised to start the treatment early, i.e., about 4-6 years of age, and preferably before 12-13 years old. For primary amenorrhea, estrogen replacement therapy is advised to induce pubertal development and maintain female secondary sex characteristics, which should start between 11 and 12 years of age. Therefore, the early diagnosis of TS plays an important role for physicians in planning treatment, which is conducive to bringing a high-quality life to TS patients.

The diagnosis of TS must be determined by a standard 30-cell karyotype analysis [4]–[6], while TS screening should adopt other effective methods due to the expensive and time-consuming process of karyotyping. Many researchers turn to typical clinical characteristics at different life stages to do the preliminary diagnosis of TS. For instance, female fetuses with diffuse oedema, cystic hygroma and growth retardation [7]; newborns with a webbed neck and congenital lymphedema in the hands and feet at birth [8]; children with short stature and high follicle-stimulating hormone levels [9]; adolescents with primary amenorrhea and/or with short stature; adults with ovarian failure [10], should draw attention to be suspected TS patients. However, the above mentioned TS screening manners in clinical practice require several physicians with rich clinical experiences, which are lacking in underdeveloped cities and countryside. And these procedures can be lengthy and repetitive. Therefore, it is essential and urgent to explore a fast, cheap and automatic TS screening approach to assist physicians to achieve the early diagnosis and treatment of TS.

Machine Learning (ML) is a potentially effective tool for screening TS to address the issues mentioned above, which has been applied by some researchers to different datasets to construct TS diagnosis models. Leite [11] applied Bayesians network to the records of symptoms to detect TS. Dumanic et al. [12] built a logistic regression model utilizing characteristics of the craniofacial complex to identify TS patients. Catic et al. [13] established neural networks based on the first-trimester maternal serum screening data, ultrasonographic findings and patient demographics to classify TS cases. However, the acquisition of such data mentioned above is not an easy task. Song et al. [14] and Chen et al. [15] proposed using facial images to diagnose TS with the help of various ML techniques. A similar application has also been found in the works of Gao et al. [16] and Yao [17]. The facial image is an important diagnostic basis for TS, which is easy to acquire. Many potential characteristics existed in facial images could be applied to identify TS, such as the wide forehead, melanocytic nevus, epicanthus, higher-rooted nasal bridge and wide distance between eyes. However, these facial characteristics might be damaged or inaccurate in their works [14]–[17], because they adjusted the detected facial part to a same size 640*640 or 128*128 that may change the original ratio of face width to face height for most people. It results in that the features extracted from the above adjusted facial images could not exactly reflect the real facial characteristics of TS patients. Moreover, they didn’t analyze which of the extracted numerous features is more important in identifying TS. Thence, an appropriate TS diagnosis model should be proposed that preserves the original ratio of face width to face height, extracts reliable features representing the real facial characteristics, and carries out feature analysis from a large number of features to discover the knowledge of the potential crucial features in recognizing TS.

To solve the above- arisen problems, we propose a new TS diagnosis model based on facial images, which maintains the original ratio of face width to face height, gives the corresponding feature extraction method and shows the process of feature analysis in identifying the crucial facial features from a large number of extracted features for recognizing TS. The proposed TS diagnosis model is composed of four components, as shown in Fig.1, including Image Preprocessing, Feature Extraction, Classifier Construction and Knowledge Discovery. For Image Preprocessing, face alignment, facial area intercept and facial brightness normalization are applied to the original facial images to derive the desired gray aligned facial area images, and the ratio of face width to height remains unchanged. For Feature Extraction, because different people have different ratios of face width to face height, corresponding reasonable features should be carefully extracted. By dividing the processed facial images into several facial organ blocks roughly and more finely, respectively, energy features and ratio features are extracted to finally form five feature sets, which can better reflect the real facial characteristics compared to the energy features of average blocks and distance features utilized by Song et al. [14] and Chen et al. [15]. For Classifier Construction, Support Vector Machine (SVM) and Ensemble Learning are employed. Besides, before establishing classifiers, Principal Component Analysis (PCA) and Kernel PCA (KPCA) are also used to reduce dimensionality. For Knowledge Discovery, feature analysis based on support vectors (SVs) resulted from SVM classifiers and principal components (PCs) acquired from PCA are carried out respectively, to discover the potential important features in identifying TS.

The main contributions of the paper are listed below:

1. We propose a new TS diagnosis model based on facial images, which preserves the original ratio of face width to face height, extracts reliable features and carries on feature analysis based on SVs and PCs;

2. We find that less energy in external canthus areas and a lower ratio of forehead height to forehead width happens more frequently in TS patients by analyzing SVs; the energy and ratio features of left zygoma area are important in identifying TS by analyzing PCs;

3. Compared with the recently proposed methods [14]–[17], the performance of our model is also competitive, which achieves the accuracy of 0.9127.

The rest of the paper is organized as follows. Section II gives related works about utilizing facial images to diagnose TS. Section III presents our proposed TS diagnosis model, which consists of Image Preprocessing, Feature Extraction, Classifier Construction and Knowledge Discovery.
Section IV describes the employed image data, experiment design and evaluation metrics, compares the prediction performance of different classifiers and conducts knowledge discovery based on SVs and PCs. The conclusion and future work are given in the last section.

II. RELATED WORKS

Considering the fact that the facial image is easy to acquire and ML techniques have been served as an effective tool to diagnose disease, some researchers studied applying various ML algorithms to facial images to establish automatic TS diagnosis models, which could assist physicians to realize early TS diagnosis and intervention to improve the patients’ life.

Song et al. [14] and Chen et al. [15] proposed to establish TS diagnosis models by making use of ML techniques based on facial images. In their works, the facial images were intercepted and zoomed out to the size of 640*640, global geometrical features (GGF), global texture features (GTF) and multiple local features (LF) were extracted from these preprocessed images, and classifiers based on the three feature sets were constructed (i.e., GGF and GTF were sent to PCA and SVM to build classifiers, LF was used by AdaBoost to construct a classifier). Table 1 summaries their works. As shown in Table 1, Gao et al. [16] and Yao et al. [17] also exploited facial images and ML techniques to build TS diagnosis models, in which the intercepted facial image size has also been made to 640*640 or 128*128. However, the operation of adjusting all the intercepted facial images to the same size of 640*640 (or 128*128) may damage the inherent information of original intercepted facial images, for instance, the ratio of face width to face height may be changed in the adjusted images for most people. It results in that the features extracted from such morphed facial images could not accurately stand for the original facial characteristics, and even may change some facial characteristics. To reduce the damage of the original information of the facial image, we preserve the original ratio of face width to face height for each individual in our work.

In addition, none of them carried out feature analysis based on their large number of extracted features to discover potentially important features. However, identifying a smaller size of latent crucial features from such a large amount of features can provide physicians with more possibilities to discover underlying knowledge to diagnose TS. Thereby, in this paper, we carry out feature analysis (e.g., analyzing features based on PCs or SVs) for our extracted numerous features to discover the potential significant facial features in identifying TS.

Not well-defined features may be one reason to make the process of discovering knowledge difficult. For instance, the GGF and GTF features in [14] should be improved. The GGF was obtained by calculating the Euclidean distance features between two different feature points based on the 68 facial points model, as shown in Fig.2. However, the distance features may be invalid in the processed facial images, because the distances of different individuals measured in real life may have been scaled to different degrees in the processed...
TABLE 1. The related articles applying machine learning techniques to facial images.

| Article | Dataset | Intercepted Facial Image size | Feature | Classifier | The best performance | Knowledge Discovery |
|---------|---------|-------------------------------|---------|------------|----------------------|---------------------|
| Song et al [14] (2018) | Training set: 46 TS, 770 non-TS Testing set: 23 TS, 385 non-TS | Zoomed out to 640*640 | GGF, GTF, LF | GGF, GTF: PCA+SVM, LF: Adaboost | Accuracy: 83.4% (highest) Sensitivity: 75.6% Specificity: 91.2% Method: LF=Adaboost | No |
| Chen et al [15] (2018) | Training set: 16 TS, 48 non-TS Testing set: 16 TS, 48 non-TS | Zoomed out to 640*640 | GGF, GTF, LF | GGF, GTF: PCA+SVM, LF: Adaboost | Sensitivity: 68.8% (highest) Specificity: 87.5% Method: LF=Adaboost | No |
| Gao et al [16] (2019) | Dataset: 113 TS, 667 non-TS 5-fold cross validation | Zoomed out to 640*640 | Wavelet Gabor Edge | PCA+2DPCA+SVM | Accuracy: 91.3% (highest) Method: (Wavelet+Gabor+Edge)+PCA+2DPCA+SVM | No |
| Yao et al [17] (2019) | Dataset: 113 TS, 654 non-TS 3-fold cross validation | Zoomed out to 128*128 | Pixels of the facial images | PCA SVM PCA+SVM | Accuracy: 83.2% (highest) Sensitivity: 83.1% Specificity: 82.4% Method: PCA+SVM | No |
| Our method | Training set: 78 TS, 424 non-TS Testing set: 20 TS, 106 non-TS | 640* height, height is scaled by the fixed ratio for each image | RRF, FRF2 RRF, FRF2 | SVM, PCA+SVM KPCA+SVM Non-weighted Ensemble Weighted Ensemble | Accuracy: 91.27% (highest) Sensitivity: 50.0% Specificity: 99.05% Method: Weighted Ensemble | Yes |

The summary of works of the four most related articles and our method are shown here, including datasets used in the experiments, the intercepted facial image size, features extracted, classifiers constructed, the best performance they achieved and whether they conducted knowledge discovery. GGF, GTF and LF stand for global geometrical features, global texture features and multiple local features, respectively. RRF and FRF are rough and finer ratio features. PCA, KPCA and SVM represent Principal Component Analysis, Kernel PCA and Support Vector Machine, respectively. TS and non-TS are the Turner Syndrome cases and non-Turner Syndrome controls, respectively.

TABLE 2. Feature points number and representing areas.

| Areas Name | Points | Areas Name | Points |
|------------|--------|------------|--------|
| The outline of face | 0-16 | Right eyebrow | 22-26 |
| Left eyebrow | 17-21 | Left eye | 36-41 |
| Left eye | 36-41 | Right eye | 42-47 |
| Nose | 27-35 | Mouth | 48-67 |

The first and third columns are the area names, and the corresponding feature points are listed in the second and last columns.

FIGURE 2. An example of 68 feature-points model on a TS facial image. The yellow dots represent the positions of the 68 feature points, which are corresponding with their numbers. The eyes are hidden for privacy protection.

digital images. For example, the distances between points 0 and 16 of different individuals acquired in reality, no matter longer or shorter, have been changed to a similar length when the intercepted facial images were zoomed out to the same size of 640*640. Nevertheless, the ratio calculated between two distances are not changed, based on the assumption that the original ratio of face width to face height of the processed images has not been changed. In our work, we extract such ratio features while ensuring that the original ratio of face width to face height remains unchanged, so as to provide a reliable basis for unearthing knowledge that which ratio feature is more important in recognizing TS.

Other global features used in [14] is GTF, which may increase the difficulty of identifying key information for detecting TS. The GTF was acquired by averagely dividing the intercepted facial image into 64 blocks and computing the energy feature of each block. However, the averagely divided blocks bring trouble to discovering knowledge, because the blocks of the same position from different people do not always stand for the same facial areas. For example, a block of person A represents the left eyebrow, but due to different facial characteristics, the block of person B at the same position may represent a part of forehead nearby her left eyebrow. It results in that the information of the same organ of different people is not easy to be analyzed in identifying TS. Hence, in our work, we divide the intercepted facial image into several organ blocks rather than averagely divided blocks, which stand for their corresponding facial organs for each person, and compute their energy feature. It furnishes a relatively easy way to discover knowledge that the energy feature of which organ plays a more significant role in recognizing TS.
III. THE PROPOSED TURNER SYNDROME DIAGNOSIS MODEL

In this paper, we propose a new TS diagnosis model based on facial images, which preserves the original ratio of face width to face height, extracts the corresponding reliable features and analyzes the potential important features from a large number of features in identifying TS. As shown in Fig.1, the proposed TS diagnosis model consists of four components, i.e., Image Preprocessing, Feature Extraction, Classifier Construction and Knowledge Discovery. For each component, the utilized techniques are described below.

A. IMAGE PREPROCESSING

Image preprocessing is applied to the original facial images, due to the reason that the acquired original facial images are not always uniform and directly applying such images to construct TS diagnosis model would lead to inappropriate results with low classification accuracy. The original images are acquired by the same camera with fixed camera position, which guarantees the same image quality, photo distance and illumination intensity, but by different photographers and shooting angles, which results in the ununiform facial images. As shown in Fig.1, the original images contain some unwanted non-facial areas with less brightness and skewed faces. Such a low-quality image data set could not be directly employed to construct a TS diagnosis model. Multiple images preprocessing techniques should be utilized to handle these issues, including Face Alignment, Facial Area Intercept and Facial Brightness Normalization.

1) FACE ALIGNMENT

To guarantee the consistency of the direction of the faces, face alignment should be carried out. The faces can be aligned by making the position of both eyes be on the same horizontal line. To acquire the coordinates of eyes, the 68 feature-points face model [18] is used, as displayed in Fig.2. The left eye and right eye are detected by points 36 to 41 and the points 42 to 47, respectively, summarized in Table 2. The eye centre is used to indicate the position of each eye, which can be located by calculating the mean value of each eye’s points. Then the rotation angle for aligning two eyes horizontally is calculated based on the coordinates of both eyes’ centres. Finally, all the faces can be aligned automatically by rotating the original facial image around the left eye centre with the calculated rotation angle.

2) FACIAL AREA INTERCEPT

Multi-task Cascaded Convolutional Networks (MTCNN) is used to intercept the facial areas from the aligned facial images. It is an effective tool, which can detect and intercept a desirable face area from an image. As a result, the intercepted facial image contains the entire face while removing non-facial areas. In addition, the width of all intercepted facial images is set to 640 pixels based on the locked aspect ratio that guarantees the original ratio of the face width to face height unchanged. It reduces the damage of the inherent information of the original images, which benefits for the subsequent reliable feature extraction.

3) FACIAL BRIGHTNESS NORMALIZATION

The grayscale image is used to increase the brightness of an image. The following equation is used to convert the RGB value of a pixel into its grayscale value, as shown below:

\[
g_{\text{gray}} = 0.2989 \times r + 0.5870 \times g + 0.1140 \times b
\]

where red, green, blue and gray correspond to the color of a pixel, respectively. It is provided by the classical MATLAB function rgb2gray [19] and produces excellent results over images in general. These weights in Equation 1 are obtained empirically and are proportional to the sensitivity of the human eye to each of the trichromat colours. The human eye is most sensitive to green while it is least sensitive to blue. Finally, all the RGB images are converted to their corresponding grayscale images.

B. FEATURE EXTRACTION

Reasonable features should be carefully extracted from the processed facial images having the original ratio of face width to face height. In order to extract reliable features and benefit for the subsequent knowledge discovery, energy features of facial organ blocks and ratio features are considered, as analyzed in the last paragraph in Section II. The energy features of facial organ blocks are acquired in two ways, i.e., the rough energy features (REF) based on rough area segmentation simply using five-row split and six-column split, and the finer energy features (FEF) based on finer area segmentation defined carefully. The ratio features are obtained in three manners, i.e., the rough ratio features (RRF) calculated between two distances acquired based on 68 feature-points, the finer ratio feature (FRF) calculated between two distances acquired based on 68 feature-points, and another FRF2 calculated between two defined distances. The five feature sets are extracted as follows.

1) ROUGH ENERGY FEATURES

In order to acquire REF, we carried out rough area segmentation to the facial images to obtain the facial organ blocks and computed the energy of each organ block. Specifically,

(1) Blocks: Rough area segmentation is used to divide the facial images into several facial organ blocks, which includes row split, column split and entire split, as shown in Fig.3. The row split is utilized to get the row parts of forehead, eyebrows, eyes, nose, cheeks, mouth and chin, as shown in Fig.3(a). The column split is employed to get the column parts of the left cheek, right cheek, left eye, right eye and nose, as exhibited in Fig.3(b). The entire split is a combination of row split and column split, which segments the facial images into several organ blocks, as displayed in Fig.3(c). As a result, 6, 5 and 30 blocks are acquired, respectively. The specific split lines used in rough area segmentation are listed in Table 3.
FIGURE 3. Facial image segmented by the rough area segmentation. (a) Row Split (red lines), get six row areas; (b) Column Split (green lines), get five column areas; (c) Entire Split (blue lines), get thirty blocks, a combination of (a) and (b). As a result, a total of 41 parts have been acquired. The white numbers shown in the three images are feature names.

TABLE 3. The specific lines utilized in row split and column split.

| Lines used in Row Split | Lines used in Column Split |
|-------------------------|---------------------------|
| y1 = 0                  | x1 = 0                    |
| y2 = upbound([left eyebrow, right eyebrow]) | x2 = leftbound([left eye]) |
| y3 = lowbound([left eye, right eye]) | x3 = rightbound([left eye]) |
| y4 = upbound(mouth)     | x4 = leftbound(right eye) |
| y5 = lowbound(mouth)    | x5 = rightbound(right eye) |
| y6 = upbound(mouth)     | x6 = img_width-1          |

$upbound(\cdot)$ and $lowbound(\cdot)$ mean the minimum and maximum $y$ coordinates of a list of points, respectively. $leftbound(\cdot)$ and $rightbound(\cdot)$ mean the minimum and maximum $x$ coordinates of a list of points. Left eyebrow and other areas denote a list of points listed in Table 2. $img\_height$ and $img\_width$ are the height and width of an image, respectively.

(2) Energy: For facial organ blocks acquired by the rough area segmentation, the energy of each block is calculated by the following equation:

$$e(B) = \text{avg}(\sum_{i=1}^{\text{height}} \sum_{j=1}^{\text{width}} |p(B)|^2)$$

(2)

where $B$ stands for a block, which is a matrix made up by pixels with value of (0,255), $e(B)$ denotes calculating a block $B$’s energy, $p(B)$ means normalizing each pixel value in $B$ to (0,1), the width and height are the block $B$’s, and $\text{avg}(\cdot)$ represents the average value.

(3) REF: REF is the joint energy of all blocks, which is represented by a vector:

$$JE(F) = (E1(e(0), e(1), \ldots, e(5)), E2(e(6), e(7), \ldots, e(10)), E3(e(11), e(12), \ldots, e(40)))$$

(3)

where $JE(F)$ denotes calculating a facial image $F$’s joint energy, $e(\cdot)$ is the energy of a block, and $E1(\cdot)$, $E2(\cdot)$ and $E3(\cdot)$ indicate the energy vector of blocks segmented by row split, column split and entire split as shown in Fig.3(a-c), respectively. Hence, REF is the joint energy of all 41 blocks acquired by the rough area segmentation.

(2) FINER ENERGY FEATURES

Though the rough area segmentation obtains the main centre parts of each facial organ, the information of more precise and detailed facial organs is also significant for TS diagnosis and analysis. Accordingly, a finer area segmentation is carried out, as shown in Fig.4. All resulting 9 segmented parts are displayed in Fig.4 (b)-(j), respectively. Fig.4 (a) displays all the 9 parts in one facial image. The corresponding detected areas belong to each segmented part are shown behind them, i.e., the second and fourth columns in Fig.4. As a result, a total of 34 areas are detected. Just like REF, FEF is the joint energy of the detected 34 areas, which may reflect the facial characteristics more precisely.

(3) ROUGH RATIO FEATURES

For each facial image, the reliable ratio features are extracted rather than distance features, as analyzed in the last paragraph in Section II. To acquire RRF, the Euclidean distances between each pair of points in 68 facial feature-points model as displayed in Fig.2 should be calculated first. As a result, a total of 2278 (2278 = 68 × 67/2) distances are obtained. If directly calculating all ratio features between two different distances, there will be a “curse of dimensionality”, as the number of ratio features is more than 2.5 million. To avoid the high dimensionality of ratio features, the face width of each image is used to normalize the distances, i.e., only the...
ratios between distances and face width are calculated. The face width is the distance between points 0 and 16. Finally, a total of 2277 RRF, not 2278 features are obtained, because one RRF feature acquired by distance(0,16)/distance(0,16) is removed.

4) FINER RATIO FEATURES
In spite that RRF includes a large number of ratio features calculated between Euclidean distances, the ratios between some vertical distances or horizontal distances haven’t been involved, such as the ratio of forehead_width and forehead_height as shown in Fig.4 (b_1). Thereby, we define 46 distances for the organ blocks acquired by the finer area segmentation, which are marked in blue as revealed in Fig.4. Based on these distances, there are two kinds of ratio feature sets are obtained, i.e., FRF and FRF2. FRF with 52 features is defined in Table 4. From this table, we can find that the calculation of FRF features utilizes not only the distances listed in Fig.4, but also the height and width of the face. The face_width is the distance between points 0 and 16 shown in Fig.2, and the face_height is the height of the intercepted facial image.

To make full use of the defined distances displayed in Fig.4, FRF2 is also calculated between all the defined distances. As a result, a total of 1035 FRF2 features are obtained, which include some features of FRF, but not all, due to not involved face_width and face_height.

C. CLASSIFIER CONSTRUCTION
For the extracted five feature sets, i.e., REF, FEF, RRF, FRF and FRF2, the proper TS detection classifier should be
TABLE 4. All the defined ratio features in 9 parts split by the finer area segmentation.

| Parts   | Defined Ratio Features                                                                 | Number |
|---------|---------------------------------------------------------------------------------------|--------|
| forehead| forehead_ratio = forehead_height / face_height; forehead_wratio = forehead_width / face_width; forehead hvoratio = forehead_height / forehead_width | 3      |
| eyebrows| left_eyebrow_wratio = left_eyebrow_width / face_width; right_eyebrow_wratio = right_eyebrow_width / face_width; ophrion_wratio = ophrion_width / face_width; ophrion_eyebrows_wratio = ophrion_width / eyebrows_area_width; eyebrows_hratio = eyebrows_height / face_height | 5      |
| eyes    | outercanthal_wratio = outercanthal_dis / face_width; innercanthal_wratio = innercanthal_dis / face_width; inner_outer_canthal_wratio = innercanthal_dis / outercanthal_dis; left_palpebral_fissure_wratio = left_palpebral_fissure_width / face_width; right_palpebral_fissure_wratio = right_palpebral_fissure_width / face_width; left_palpebral_fissure_hratio = left_palpebral_fissure_height / face_height; right_palpebral_fissure_hratio = right_palpebral_fissure_height / face_height; lr_palpebral_fissure_wratio = lr_palpebral_fissure_width / right_palpebral_fissure_width | 11     |
| nose    | nasal_appro_ratio = nasal_height / face_height; nasal_ridge_appro_ratio = nasal_ridge_appro_length / face_height; alar_base_wratio = alar_base_width / face_width; lr_alar_base_hratio = left_alar_base_length / right_alar_base_length; alarbasew_nasalheight_whratio = alar_base_width / nasal_height; nasal_tip_protrusion_hratio = nasal_tip_protrusion_height / nasal_height; left_alar_base_wlocos = left_alar_base_width / left_alar_base_length; right_alar_base_wlocos = right_alar_base_width / right_alar_base_length | 8      |
| mouth   | mouth_wratio = mouth_width / face_width; mouth_ratio = mouth_height / face_height; philtrum_ratio = philtrum_height / face_height; up_lip_ratio = up_lip_height_mid / face_height; down_lip_ratio = down_lip_height_mid / face_height | 5      |
| chin    | chin_nratio = chin_height / face_height; chin_wratio = chin_width / face_width | 2      |
| cheek   | lccheck_wratio = lccheck_width / face_width; lccheck_hratio = lccheck_height / face_height; lccheck_whratio = lccheck_width / lccheck_height; rccheck_wratio = rccheck_width / face_width; rccheck_hratio = rccheck_height / face_height; rccheck_whratio = rccheck_width / rccheck_height | 6      |
| zygoma  | lzygoma_wratio = lzygoma_width / face_width; lzygoma_hratio = lzygoma_height / face_height; lzygoma_whratio = lzygoma_width / lzygoma_height; rzygoma_wratio = rzygoma_width / face_width; rzygoma_hratio = rzygoma_height / face_height; rzygoma_whratio = rzygoma_width / rzygoma_height | 6      |
| lowercheck | llowercheck_wratio = llowercheck_width / face_width; llowercheck_hratio = llowercheck_height / face_height; llowercheck_whratio = llowercheck_width / llowercheck_height; llowercheck_wratio = llowercheck_width / face_width; llowercheck_hratio = llowercheck_height / face_height; llowercheck_whratio = llowercheck_height / llowercheck_width; rlowercheck_wratio = rlowercheck_width / rlowercheck_height; rlowercheck_hratio = rlowercheck_height / rlowercheck_width; rlowercheck_whratio = rlowercheck_height / rlowercheck_width | 52     |

The first column lists the names of the 9 segmented parts shown in Fig. 4. The second column gives the defined ratio names. And the last one counts the number of defined ratio features for each part. “1” in lccheck (lzygoma and llowercheck) means the left cheek (zygoma and lower area of cheek) area. “r” denotes the corresponding right areas. “hratio” represents the ratio of the specific height to the face height. “wratio” is the ratio of specific width to the face width. “lratio” is the specific length to the face height. “hratio” (“whratio”, “wwratio”, “wlcos” and “llratio”) is the ratio of specific height to specific width (width to height, width to width, length to length, and length to length). The face width is the distance between point 0 and 16, and the face height is the height of the intercepted image.

To establish an effective TS diagnosis model, SVM is considered due to its stability, which has been widely used in medical diagnosis [20]. Moreover, ensemble learning is investigated to improve the classification performance. Accordingly, single classifiers for each feature set and ensemble classifiers combining all single classifiers are built.

1) SINGLE CLASSIFIERS

For the five feature sets, SVM is employed to establish single classifiers respectively. Considering that some redundant and irrelevant features may affect the classifiers’ performance, especially for the classifiers established by a large number of features, e.g., RRF (2277 features) and FRF2 (1035 features), PCA and KPCA are applied to the extracted feature sets to reduce dimensionality. Single classifiers including base SVM classifier, PCA+SVM classifier and KPCA+SVM classifier for each feature set are established. As a result, a total of 15 single classifiers are constructed, i.e., 5 SVM classifiers, 5 PCA+SVM classifiers and 5 KPCA+SVM classifiers. Details of the adopted SVM, PCA and KPCA methods are described as below.

(1) Support Vector Machine

SVM is a popular supervised classification and regression tool. The motivation of SVM is to solve the two-class pattern recognition problem. SVM makes an attempt to minimize the classification error and maximize the geometric margin (i.e., the maximum distance from the nearest samples from two classes). Thus, SVM is also called maximum margin classifier and the nearest training samples are formed as the SVs. Moreover, to solve the classification problem of non-linear data, SVM can perform non-linear classification through kernel tricks. The kernel tricks map the input into high-dimensional feature spaces, which are linearly separable.

For the TS recognition, given a set of labeled training data D, let \((x_i, y_i)\), denote the training pairs, where \(x_i\) is a vector of features for a TS facial image and \(y_i\) is the corresponding label 1 or \(-1\), i.e., TS or non-TS. The SVM algorithm tries to seek for the maximum-margin hyperplane and the solution is obtained through the method of Lagrange multipliers:

\[
\max \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \cdot \phi(x_j)
\]

s.t. \(0 \leq \alpha_i \leq C\)

\[
\sum_i \alpha_i y_i = 0
\]  

(4)

where the variables \(\alpha_i\) and \(\alpha_j\) denote the Lagrange multipliers; \(C\) is the regularization parameter; \(\phi(\cdot)\) is a non-linear
mapping, which usually maps $x_i$ into a higher-dimensional space. This formula can be solved based on the Karush-Kuhn-Tucker theorem. The solution is a linear combination of the SVs, which are the examples lie closest to the decision boundary. Notably, the kernel function $K$ is defined as $K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$, which is exploited to transform the original data space to a high-dimensionality and linearly separable space. In our work, we investigate multiple kernels (i.e., Gaussian kernel, Laplace kernel, Polynomial kernel and Polyplus kernel) for the classification task.

$$Gaussian: K(x, y) = e^{-((x-y)^2)/2\sigma^2} \quad (5)$$

$$Laplace: K(x, y) = e^{-|x-y|/\tau} \quad (6)$$

$$Polynomial: K(x, y) = (x \ast y)^d \quad (7)$$

$$Polyplus: K(x, y) = (x \ast y + r)^d \quad (8)$$

For each kernel, tuning process is initiated to filter out the optimal parameter, which could achieve the highest accuracy. The kernel with the optimal parameter is applied to construct the prediction model.

(2) Principal Component Analysis and Kernel PCA

PCA method is applied to reduce the dimensionality and expose the internal structure of data. It is one of the most widely used statistical techniques for dimensionality reduction [21]–[23]. PCA uses all of the original variables to acquire a smaller set of new variables (i.e., PCs), which can be utilized to approximately reconstruct the original variables. It is an orthogonal linear transformation which converts original data into a new coordinate system such that the greatest variance by a certain projection of data comes to lie on the first coordinate, i.e., the first PC, the second greatest variance on the second coordinate, and so on. The more relevant the original variables are, the fewer the number of new variables required. PCs are ranked based on the magnitude of their corresponding eigenvalues, which results in the first few PCs retaining most of the variations presented in the original sets.

Given a data matrix $M$ with $n$ features, each of which has $q$ variables denoted as $f_1, \ldots, f_q$, PCA computes a new variable $z_1$ as the first PC. The new variable accounts for the greatest variance subject to $\gamma_1^T \gamma_1 = 1$. Similarly, the second PC also can be acquired by $z_2 = \gamma_2^T f$, where $f$ is $f_1, \ldots, f_q$. To achieve the greatest variance, $\gamma_2$ meets the two conditions, $\gamma_2^T \gamma_2 = 1$ and $\gamma_2^T \gamma_1 = 0$. Further, the $v_{th}$ PC is $z_v = \gamma_v^T f$, which gets greatest variance subject to $\gamma_v^T \gamma_j = 0$ for $j < v$. To find the coefficients $\gamma_v$ for $v_{th}$ PC, we should solve an optimization problem:

$$\max z_v = \gamma_v^T f$$

s.t. $\gamma_v^T \gamma_v = 1$

$$\gamma_v^T \gamma_j = 0 \quad (0 \leq j \leq v)$$

$$\gamma_0 = 0 \quad (9)$$

The solution can be obtained by $\text{Var}(z_v) = \gamma_v^T H \gamma_v$, where $H$ is the co-variance matrix of original variables. The solution of $\gamma_v$ is the eigenvector of $H$ which is corresponding to the largest eigenvalue. The eigenvalues can be acquired from the equation, $|H - \lambda E| = 0$, where $E$ is a unitary matrix and $\lambda$ is some unknown eigenvalue of $H$. Then we get the ordered eigenvectors based on the eigenvalues. The eigenvectors with larger eigenvalues are corresponding to the PCs that achieve more variance of the original data. In addition, before using PCA, the original data should be scaled to unit variance, which makes the analysis meaningful.

The kernel tricks also can be introduced to the PCA, named as Kernel PCA (KPCA) [24]. The idea of KPCA is to firstly map the original input vectors $x_i$ into a high-dimensional feature space $\phi(x_i)$ and then calculate the linear PCA in $\phi(x_i)$. By mapping $x_i$ into $\phi(x_i)$ whose dimension is assumed to be larger than the number of training samples.

2) ENSEMBLE CLASSIFIERS

In order to improve the accuracy of the TS diagnosis, we use the ensemble learning technique to construct effective TS prediction models. An ensemble method often can combine multiple different weak classifiers to form a stronger model based on a certain strategy. The higher diversity the weak classifiers have, the better results the ensemble model generates. In our study, 15 different single classifiers are constructed and combined together to form the final stronger TS classifier. The 15 single classifiers are trained on five well-defined feature sets (described in the above section).

Two classical ensemble methods, i.e., non-weighted voting method and weighted voting method, are employed to build powerful TS models. The non-weighted voting method utilizes the conventional majority voting principle for the combination of the weak classifiers, which decides the final class for a sample according to the predictive label of the weak classifiers in the majority. The weighted voting method adds weight to each weak classifier and the label which acquires majority votes is regarded as the final prediction result.

D. KNOWLEDGE DISCOVERY

To discover potential important features from the extracted numerous features in identifying TS, feature analysis based on SVs and PCs are carried out respectively. The process of knowledge discovery is shown below.

1) FEATURE ANALYSIS BASED ON SUPPORT VECTORS

The SVs acquired from SVM classifiers could be utilized to conduct feature analysis to dig out the latent significant features, due to the reason that SV is a subset of training data containing the most informative patterns. For each feature set, the corresponding SVM classifiers are constructed. The maximum margins of different SVM classifiers trained using different kernels and parameters could provide valuable information for the feature analysis of TS. By ranking the maximum margins built by different SVM classifiers, we select the best SVM model with the largest maximum.
margin. Accordingly, for each feature set, SVs derived from the best SVM model are used for the feature analysis.

To identify the crucial features based on the SVs, three widely used feature ranking methods are employed, including chi-square test, information gain (IG) and analysis of variance (ANOVA) F-test. For each ranking method, the top 10 features with the highest scores are chosen. And the common features selected at least twice are investigated to be the important features in identifying TS.

2) FEATURE ANALYSIS BASED ON PRINCIPAL COMPONENTS
The PCs derived from the PCA should be employed to analyze which feature is more important in recognizing TS. Each PC is reconstructed based on all of the original variables. PCs are ranked based on their corresponding eigenvalues, i.e., the larger the eigenvalue, the more important the PC. The first few PCs retains most variance of the original data. To discover the crucial significant features conveniently, the number of PCs that represents about 93% of the variance is chosen due to its small size. Since there are 41 (34, 2277, 52, 1035) variables in REF (FEF, RRF, FRF, FRF2) required to be considered for each PC, only the variable with the maximum absolute coefficient value in each PC is reported. In summary, the variable with the maximum absolute coefficient value in the first PC is thought to be the most important feature in identifying TS, meanwhile, other significant PCs are also presented.

IV. EXPERIMENTS AND RESULTS
A. IMAGE DATA
The facial images used in this paper are acquired from the Peking Union Medical College Hospital. Written informed consent was obtained from all participants or guardians and we also obtained the approval of our institutional ethical review board for the research. The facial images are collected from TS females ranging from 3 to 17 years of age. 90% of them are females with ages ranging from 9-14 years that are usually suspected with short stature and/or delayed puberty. Meanwhile, the controls are also selected from a large number of non-turner individuals in the hospital within the same age group. Their frontal images are employed to automatically identify TS with the help of various ML techniques. From July 2016 to December 2017, a total of 98 cases and 530 controls are obtained by excluding the unqualified images, such as blurry data and images with too many rich facial expressions. Generally, in ML, cases (i.e., TS individuals) and controls (i.e., non-TS individuals) are labelled as ‘+1’ and ‘−1’, respectively.

B. EXPERIMENT SETTINGS
The original dataset is divided into the training set and testing set, where training set includes 80% of cases and 80% of controls, and testing set consists of the rest of the dataset. For the training set, 3-fold cross validation is employed to decide the optimal parameters of classifiers. Based on the acquired optimal parameters, the whole training set is reused to establish the final classifiers. The testing set is utilized to evaluate the performance of the final classifiers built by different techniques.

The number of components is decided by 99.5% of information retained in each feature set, which brings good performance for classifiers. The information amount is the cumulated percentages of eigenvalues occupying the total eigenvalues.

For each kernel used for SVM and KPCA, the parameter $t$ of Gaussian kernel (the parameter $t$ of Laplace kernel, the parameter $d$ of Polynomial kernel and the parameter $d$ of Polyplus kernel with $r=1$) is set to the range from $10^{−5}$ to $10^{15}$ with a step size of $10^{(i+1)}$ (from $10^{−5}$ to $10^{5}$ with a step size of $10^{(i+1)}$, from 2 to 15 with a step size of 1, and from 2 to 15 with a step size of 1). The grid search method is applied for the selection of optimal parameters for KPCA+SVM classifiers.

C. EVALUATION METRICS
To assess the performance of TS classifiers, three widely used evaluation metrics are considered, i.e., accuracy (acc), sensitivity (se) and specificity (sp). Accuracy indicates the judgment ability that the TS prediction model can make the correct decision for the detection of TS. Sensitivity represents the ability that the model classifies the positive TS images correctly. Specificity refers to the ratio of the number of negative images with correctly predicted results to the number of all negative images. The three criteria are calculated based on Eq.(10), (11) and (12), as shown below:

$$\text{acc} = \frac{TP + TN}{TP + FN + FP + TN} \quad (10)$$
$$\text{se} = \frac{TP}{TP + FN} \quad (11)$$
$$\text{sp} = \frac{TN}{TN + FP} \quad (12)$$

where TP (FN) is the number of positive samples that are predicted as positive (negative) samples, and TN (FP) is the number of negative samples that are classified as negative (positive) samples. Here, TS images are specified as positive samples, and the non-TS images are specified as the negative samples. The higher the values of these metrics, the better the prediction performance.

D. COMPARISON OF PREDICTION PERFORMANCE OF DIFFERENT CLASSIFIERS
1) COMPARISON OF PREDICTION PERFORMANCE OF SVM SINGLE CLASSIFIERS
Fig. 5 reports the performance of SVM classifiers built by the five feature sets respectively. From this figure, we can observe that the SVM classifier constructed by FEF achieves the highest accuracy of 0.881. SVM classifier built by FEF works better than the one built by REF, which indicates that using the finer area segmentation to get more entire and detailed organs is truly valuable. The similar result is also observed...
between RRF and FRF. The feature set FRF2 obtains the lowest accuracy. By analyzing the features in FRF2, we find that the ratios between face height or width and other defined distances have not been included; whereas, parts of this kind of ratios are included in FRF. Therefore, it can be concluded that this kind of ratio features may have a significant role in the prediction of TS. Meanwhile, all the classifiers have a high specificity with low sensitivity. One possible reason may be the imbalanced problem between the TS cases and controls, in which the number of components is decided by 99.5% of the variance retained in each feature set. Comparing to the results shown in Fig.5, we observe that not all the classifiers have an improvement in accuracy by using the dimensionality reduction methods. In Fig.6, the PCA+SVM classifier built by FEF achieves the highest accuracy. There are apparent improvements in accuracy for RRF and FRF2, but the accuracies for FEF and FRF remains unchanged, and it even decreases for REF. Similar findings could be observed in Fig.7. As shown in Fig.7, the highest accuracy is obtained by the KPCA+SVM classifier built by FEF. For REF, RRF and FRF2, the accuracies are better than the corresponding results in Fig.5, but for FEF and FRF, there is a decline in accuracy. Besides, we find that the improvement of accuracy for RRF and FRF2 is more significant than that for REF, FEF, and FRF. The different change in accuracy may be caused by the fact that the original RRF and FRF2 feature sets are high-dimensional, in which several ineffective features are included. Applying PCA/KPCA to the original RRF and FRF2 could exclude some redundant and irrelevant features to improve the quality of the two feature sets; however, for small feature sets, i.e., REF, FEF and FRF, some informative features may be removed from the original feature sets by using PCA/KPCA, which results in a decrease in classification accuracy.

3) COMPARISON OF PREDICTION PERFORMANCE OF ENSEMBLE CLASSIFIERS
The prediction results of two ensemble methods (i.e., non-weighted voting and weighted voting methods) are presented in Fig.8. Both ensemble classifiers achieve better performance than the 15 single classifiers. Compared with the non-weighted voting method, the weighted approach obtains a higher accuracy of 0.9127. However, the se value is still low, which may be caused by the imbalanced data and the not revealed special facial appearances in some TS cases.
TABLE 5. The ranked top 10 features based on three ranking methods.

| Features | Ranking method | Ranked features |
|----------|----------------|-----------------|
| REF | Chi Square | 36.40 35.16 29.30 10.23 13.11 |
| | IG | 30.16 29.18 32.21 1 33.34 |
| | ANOVA F-test | 36.35 29.23 40.9 10.39 34.28 |
| RRF | Chi Square | 6 51.0 17.6 30.6 52.6 62.6 67.6 66.7 65.6 49.7 51 |
| | IG | 6.19 12.28 6.34 6.27 6.53 7.28 5.40 6.52 6.18 6.35 |
| | ANOVA F-test | 59.67 7.9 9.7 8.5 6.5 7.65 6.53 6.52 7.66 6.9 |
| FEF | Chi Square | rowercheek_area righ_extmearcanthus rcheek_area rzygoma_area left_eye_white_area left_eye_white_area left_extmearcanthus eye_brow_area right_eye_area right_iris |
| | IG | philtrum_area lyzygoma_area left_extmearcanthus ophryon_area nosetip_area eyes_area left_eyebrow_area central chin_area rcheek_area rzygoma_area |
| | ANOVA F-test | rrowercheek_area rcheek_area rzygoma_area right_extmearcanthus right_iris left_eye_white_area right_eye_area right_eye_area left_eyebrow_area right_eyepicanthus |
| FRF | Chi Square | forehead_hratio lyzygoma_whratio rzygoma_whratio forehead_hratio down_lip_hratio rcheek_whratio rzygoma_whratio rowercheek_whratio left_palpebral_fissure_whratio mouth_hratio |
| | IG | lrowercheek_whratio rzygoma_whratio lyzygoma_whratio rowercheek_whratio lcheek_hratio rcheek_whratio right_palpebral_fissure_hratio left_palpebral_fissure_hratio forehead_hratio lr_alae_base_lratio |
| | ANOVA F-test | lrowercheek_hratio rcheek_hratio rzygoma_whratio forehead_hratio down_lip_hratio rcheek_hratio left_palpebral_fissure_hratio |
| FRF2 | Chi Square | forehead_height/eyebrows_height forehead_width/eyebrows_height eyebrows_area_width/eyebrows_width forehead_height/down_lip_height_left forehead_height/down_lip_height_forehead_height/down_lip_height_right eyebrows_area_width/down_lip_height_forehead_height/down_lip_height_left forehead_height/up_lip_height_left |
| | IG | nasal_ridge_atop_length/up_lip_height_mid ophryon_width/up_lip_height_right right_eyebrow_width/effect_alae_base_length eyebrows_area_width/up_lip_height_right mouth_height/mouth_height left_eyebrow_width/down_lip_height_forehead_height/up_lip_height_right eyebrows_area_width/down_lip_height_forehead_height/down_lip_height |
| | ANOVA F-test | forehead_width/chin_width eyebrows_area_width/chin_width innercanthal_dis/chin_width outercanthal_dis/chin_width right_eyebrow_width/chin_width ch_width/lowercheek_width ch_width/lowercheek_width innercanthal_dis/lcheek_height/chin_width/zygoma_width |

The features selected at least twice are considered to be important to discover the underlying knowledge.

4) PERFORMANCE COMPARISON BETWEEN THE PROPOSED TS MODEL AND OTHERS

Comparing the performance of our proposed method with the related works mentioned in Section II, the weighted ensemble classifier established by our method achieves the highest specificity of 0.9905 and the second higher accuracy of 0.9127, which is very close to the highest accuracy of 0.913 acquired by Gao et al. [16]. However, the sensitivity is lower than others acquired by Song et al. [14], Chen et al. [15] and Yao et al. [17]. We conjecture that the difference may be caused by the different quality of datasets. In addition, the imbalance of our dataset may also have a negative impact on the proposed model.

Moreover, Chen et al. [15] also reported the diagnosis result of their dataset by clinical workers, which included 32 cases and 96 controls. However, only sensitivity and specificity were presented, which were 57.4 ± 21.9% and 75.4 ± 17.3%, respectively. From the result, we find the diagnosis of clinical workers has a fluctuation based on the facial image. Therefore, comparing with clinical workers, it can be seen that the proposed model is more stable and effective.

E. KNOWLEDGE DISCOVERY FROM SUPPORT VECTORS AND PRINCIPAL COMPONENTS

1) FEATURE ANALYSIS BASED ON SUPPORT VECTORS

Three feature ranking methods (i.e., chi-square test, IG and ANOVA F-test) are applied to SVs to select the discriminative features. For each ranking method, the top 10 features are listed in Table 5. The features selected at least twice are thought to be significant features in identifying TS, which are discussed below.

(1) For REF, the energy features of part 10, 16, 23, 29, 30, 34, 35, 36 and 40 have been selected at least twice. These important energy features should attract the attention of physicians to discover the potential facial characteristics of TS patients. For example, for cases and controls in SVs, the mean values of energy feature of part 23 are 0.2251 and 0.2370 respectively, where part 23 stands for the nasal root area. The energy feature of part 23 in cases is smaller than the one in controls. For part 35 (the lower part of the right cheek), the energy has a higher mean value of 0.1387 in cases than 0.1238 in controls. These findings could further be analyzed by physicians to discover useful information.

VOLUME 8, 2020 214877
TABLE 6. Summary of PCs for REF.

| PCs   | Eigenvalues | Percentage of eigenvalues | Cumulated percentage | Coefficient | Variable name |
|-------|-------------|---------------------------|----------------------|-------------|---------------|
| 1     | 0.2239      | 61.16                     | 61.16                | 0.319        | 26            |
| 2     | 0.0438      | 11.96                     | 73.12                | 0.3221       | 29            |
| 3     | 0.0526      | 8.91                      | 82.03                | 0.5260       | 36            |
| 4     | 0.0206      | 5.64                      | 87.67                | 0.3793       | 16            |
| 5     | 0.0179      | 4.88                      | 92.55                | -0.4818      | 36            |
| 6     | 0.0047      | 1.29                      | 93.84                | -0.6091      | 40            |

The first two columns are the PCs ranked by eigenvalues and the corresponding percentages (%). The third and fourth columns represent the percentage of each PC and the corresponding cumulated percentage. The last two columns display the maximum absolute coefficient for each PC and the corresponding variable name.

TABLE 7. Summary of PCs for RRF.

| PCs   | Eigenvalues | Percentage of eigenvalues | Cumulated percentage | Coefficient | Variable name |
|-------|-------------|---------------------------|----------------------|-------------|---------------|
| 1     | 0.6799      | 37.93                     | 37.93                | -5.6453e-02 | 7.23          |
| 2     | 0.3564      | 19.88                     | 57.81                | 5.8631e-02  | 11.30         |
| 3     | 0.3071      | 17.13                     | 74.94                | -6.1285e-02 | 16.30         |
| 4     | 0.0825      | 4.60                      | 79.54                | 7.9429e-02  | 48.54         |
| 5     | 0.0654      | 3.65                      | 83.19                | 9.5709e-02  | 30.57         |
| 6     | 0.0523      | 2.92                      | 86.11                | 6.0542e-02  | 48.54         |
| 7     | 0.0382      | 2.13                      | 88.24                | -7.1765e-02 | 4.12          |
| 8     | 0.0334      | 1.86                      | 90.1                 | 6.3406e-02  | 18.26         |
| 9     | 0.0233      | 1.30                      | 91.4                 | -7.7838e-02 | 60.54         |
| 10    | 0.0198      | 1.10                      | 92.5                 | 6.3533e-02  | 30.51         |
| 11    | 0.0161      | 0.90                      | 93.4                 | -6.9775e-02 | 8.26          |

The meanings of the column headers are the same with Table 6.

TABLE 8. Summary of PCs for FEF.

| PCs   | Eigenvalues | Percentage of eigenvalues | Cumulated percentage | Coefficient | Variable name     |
|-------|-------------|---------------------------|----------------------|-------------|-------------------|
| 1     | 0.1581      | 70.65                     | 70.65                | 0.3352      | lrygoma_area      |
| 2     | 0.0335      | 14.96                     | 85.61                | 0.3432      | rzygoma_area      |
| 3     | 0.0093      | 4.14                      | 89.75                | 4.4325e-01  | right_externalcanthus |
| 4     | 0.0072      | 3.20                      | 92.95                | -0.5773     | forehead_area     |
| 5     | 0.0037      | 1.66                      | 94.61                | -0.4676     | left_eye_lwblote_area |

The meanings of the column headers are the same with Table 6.

TABLE 9. Summary of PCs for FRF.

| PCs   | Eigenvalues | Percentage of eigenvalues | Cumulated percentage | Coefficient | Variable name |
|-------|-------------|---------------------------|----------------------|-------------|---------------|
| 1     | 0.0837      | 49.42                     | 49.42                | 7.5916e-01  | rzygoma_whratio |
| 2     | 0.0378      | 21.06                     | 70.48                | -6.7713e-01 | lzygoma_whratio |
| 3     | 0.0277      | 15.46                     | 85.94                | 0.6264      | llowercheek_whratio |
| 4     | 0.0067      | 3.71                      | 89.65                | -6.9258e-01 | forehead_whratio |
| 5     | 0.0040      | 2.22                      | 91.87                | -5.7146e-01 | right_palpebral_fissure_hwratio |
| 6     | 0.0025      | 1.41                      | 93.28                | 0.4378      | alarbaseal_natalheight_whratio |

The meanings of the column headers are the same with Table 6.

TABLE 10. Summary of PCs for FRF2.

| PCs   | Eigenvalues | Percentage of eigenvalues | Cumulated percentage | Coefficient | Variable name                      |
|-------|-------------|---------------------------|----------------------|-------------|------------------------------------|
| 1     | 205.1575    | 64.59                     | 64.59                | 0.2308      | forehead_width/up_lip_height_mid   |
| 2     | 36.4458     | 11.47                     | 76.06                | 0.3728      | forehead_width/right_palpebral_fissure_height |
| 3     | 23.3773     | 7.36                      | 83.42                | 0.2299      | eyebrows_area_width/down_lip_height_mid |
| 4     | 12.3291     | 3.88                      | 87.3                 | -0.2234     | forehead_height/up_lip_height_mid   |
| 5     | 7.8926      | 2.48                      | 89.78                | 0.3736      | eyebrows_area_width/eyebrows_height |
| 6     | 6.9368      | 2.18                      | 91.96                | 0.1944      | eyebrows_area_width/eyebrows_height |
| 7     | 4.4969      | 1.42                      | 93.38                | 0.2480      | eyebrows_area_width/eyelid_width   |

The meanings of the column headers are the same with Table 6.

(3) For FEF, the features selected by the chi-square test are just selected at least twice. It is found that the energy of right_externalcanthus and left_externalcanthus, corresponding to both ends of the rectangle areas shown in Fig. 4 (d), are deemed to have potential value. For cases and controls in SVs, the mean energy value of the area left_externalcanthus is 0.2738 and 0.2980, respectively, and the mean value of area right_externalcanthus is 0.1518 and 0.1529, respectively.
Therefore, TS patients have less energy in their external canthus areas. The lowercheek_area represents the area of the right lower cheek, which gets a higher mean energy value of 0.1401 for cases than 0.1358 for controls. This finding is similar to the one found in part 35 (the lower part of the right cheek) in REF.

(4) For FRF, the features of forehead_hwratio (i.e., the forehead height to width ratio), left_palpebral_fissure_hwratio (i.e., the left palpebral fissure height to width ratio), down_lip_hratio (i.e., the down lip height to face height ratio), lcheek_whratio (i.e., the left cheek width to height ratio), rcheek_whratio (i.e., the right cheek width to height ratio), lzygoma_whratio (i.e., the left zygoma width to height ratio), rzygoma_whratio (i.e., the right zygoma width to height ratio), llowercheek_whratio (i.e., the left lower cheek width to height ratio) and rlowercheek_whratio (i.e., the right lower cheek width to height ratio) are selected at least twice. For cases and controls, the mean values of the feature forehead_hwratio are 0.3883 and 0.3935, respectively. Forehead_hwratio is the ratio of the forehead height and width. The smaller values in cases may indicate that the facial features of the smaller forehead heights and the larger forehead widths exist more frequently in TS patients.

(5) For FRF2, we find that none of the features is selected twice or more. It is guessed that a single feature may not have strong predictive power, but their combination perhaps contributes to the classification decision.

2) FEATURE ANALYSIS BASED ON PRINCIPAL COMPONENTS

To discover the potential important features, the number of PCs that represents more than 93% of the variance is chosen. Since there are 41 (2277, 34, 52, 1035) variables in REF (RRF, FEF, FRF, FRF2) required to be considered, only the variable with the maximum absolute coefficient value in each PC is reported. Tables 6-10 show the summary of PCs in the five feature sets.

As shown in Table 6, for the feature set REF, by using PCA, the first PC occupies 61.16% information of all the resulting PCs. The variable 26 exhibited in Fig. 3(c), i.e., the average energy of left zygoma part, has the maximum absolute coefficient value, which contributes the most important information to the first PC.

As displayed in Table 7, for RRF, the first three PCs occupy 74.94% information in all the resulting PCs. The variables of the ratio of distance calculated between the point pairs (7.23)/(11.30)/(16.30) exhibited in Fig.2 to face width, have the maximum absolute coefficient values for the first three PCs, respectively.

As presented in Table 8, for FEF, the first PC occupies 70.65% information in all the resulting PCs. From this table, we find that the average energy of lzygoma_area (the left zygoma area) shown in Fig.4(i) is of significant value, which is similar to the finding observed from Table 6.

As shown in Table 9, for FRF, the first two PCs occupy 70.48% information in all the resulting PCs. The variables of rzygoma_whratio and lzygoma_whratio (i.e., the ratios of width to height of the right and left zygoma areas) are important for the first two PCs, respectively.

As listed in Table 10, for FRF2, the first PC occupies 64.59% information in all the resulting PCs. The variable of forehead_width/up_lip_height_mid (i.e., the ratio of forehead width to the height of up lip) has the maximum absolute coefficient value for the first PC.

In summary, the information (energy or ratio) of left zygoma area are involved in three first PCs (i.e., the first PC of REF, FEF and FRF), which implies that the left zygoma is informative to the TS diagnosis. Other informative variables are also listed in Tables 6-10, which are helpful for physicians to find special facial features for TS patients.

V. CONCLUSION

In this paper, we propose a new TS diagnosis model based on facial images preserving the original ratio of face width to face height, extracting reliable features and carrying out feature analysis based on SVs and PCs. The proposed TS diagnosis model consists of Image Preprocessing, Feature Extraction, Classifier Construction and Knowledge Discovery. Through the Image Preprocessing, the facial area is intercepted and the original ratio of face width to face height is preserved. By the Feature Extraction, the reliable feature sets are extracted, i.e., REF, FEF, RRF, FRF and FRF2. For the classifier construction, 15 single classifiers and 2 ensemble classifiers are constructed. The weighted voting method achieves the highest accuracy of 0.9127. FEF outperforms the other four feature sets, however, FRF is also a competitive feature set, which always obtain relatively good performance for the TS prediction task. For knowledge discovery, by analyzing the SVs acquired from SVM classifiers, multiple significant features are observed, such as less energy in external canthus areas and lower ratio of forehead height to forehead width in TS patients. By analyzing features derived from the PCs, we find that the energy and ratio features in left zygoma area of TS patients are important, which could be further studied by physicians.

In the future, we will try to propose more effective and reliable feature extraction methods to benefit the feature analysis of TS disease, which trends to extract the deep potential facial features involved in TS. Furthermore, deep learning techniques will also be applied to the TS dataset for constructing more accurate and reliable TS prediction models.

REFERENCES

[1] C. H. Gravholt, N. H. Andersen, G. S. Conway, O. M. Dekkers, M. E. Geffner, K. O. Klein, A. E. Lin, N. Maurus, C. A. Quigley, K. Rubin, D. E. Sandberg, T. C. J. Sas, M. Silberbach, V. Söderström-Anttila, K. Stochholm, J. A. van Alfen-van Derveld, J. Woelfle, and P. F. Backeljauw, “Clinical practice guidelines for the care of girls and women with turner syndrome: Proceedings from the 2016 cincinnati international turner syndrome meeting,” Eur. J. Endocrinol., vol. 177, no. 3, pp. G1–G70, 2017.

[2] C. A. Bondy, “Turner’s syndrome and X chromosome-based differences in disease susceptibility,” Gender Med., vol. 3, no. 1, pp. 18–30, Mar. 2006.

[3] R. J. Blashcke and G. A. Rappold, “SHOX: Growth, Léri–Weill and turner syndromes,” Trends Endocrinol. Metabolism, vol. 11, no. 6, pp. 227–230, Aug. 2000.
J. Li et al.: Diagnosis and Knowledge Discovery of TS Based on Facial Images Using Machine Learning Methods

[4] K. H. Nicolaides, G. Azar, R. J. M. Snijders, and C. M. Gosden, “Fetal nuchal oedema: Associated malformations and chromosomal defects,” Fetal Diagnosis Therapy, vol. 7, no. 2, pp. 123–131, 1992.

[5] C. Ruiz, F. Lamm, and P. S. Hart, “Turner syndrome and multiple-marker screening,” Clin. Chem., vol. 45, no. 12, pp. 2259–2261, Dec. 1999.

[6] D. J. Wolff, D. L. Van Dyke, and C. M. Powell, “Laboratory guideline for turner syndrome,” Genet. Med., vol. 12, no. 1, pp. 52–55, Jan. 2010.

[7] B. Polivka and K. L. Merideth, “Sonographic prenatal diagnosis of turner syndrome,” J. Diagnostic Med. Sonogr., vol. 31, no. 2, pp. 99–102, March 2015.

[8] M. D. P. Jung, J. L. D. Amaral, R. G. Fontes, A. T. D. Costa, S. M. Wuillaume, and M. H. C. D. A. Cardoso, “Diagnóstico da síndrome de turner: A experiência do instituto estadual de diabetes e endocrinologia rio de janeiro, de 1970 a 2008,” Revista Brasileira de Saúde Materno Infantil, vol. 10, no. 1, pp. 117–124, 2010.

[9] S. Carpini, A. B. Carvalho, S. H. V. de Lemos-Marini, G. Guerra-Junior, and A. T. Maciel-Guerra, “FSH may be a useful tool to allow early diagnosis of turner syndrome,” BMC Endocrine Disorders, vol. 18, no. 1, pp. 1–7, Dec. 2018.

[10] C. A. Bondy, “New issues in the diagnosis and management of turner syndrome,” Rev. Endocrine Metabolic Disorders, vol. 6, no. 4, pp. 269–280, Dec. 2005.

[11] H. Pereira, “Applicability of data mining technique using bayesian networks in diagnosis of genetic diseases,” Int. J. Adv. Comput. Sci. Appl., vol. 4, no. 1, pp. 1–4, 2013.

[12] J. Dumancic, Z. Kaic, M. L. Varga, T. Lauz, M. Domic, S. A. Milosevic, and H. Brkic, “Characteristics of the craniofacial complex in turner syndrome,” Arch. Oral Biol., vol. 55, no. 1, pp. 81–88, Jan. 2010.

[13] A. Catic, L. Gurbeta, A. Kurtovic-Kozaric, S. Mehmedbasic, and A. Badnjevic, “Application of neural networks for classification of patau, edwards, down, turner and klinefelter syndrome based on first trimester maternal serum screening data, ultrasonographic findings and patient demographics,” BMC Med. Genomics, vol. 11, no. 1, pp. 1–12, Dec. 2018.

[14] W. Song, Y. Lei, S. Chen, Z. Pan, J.-J. Yang, H. Pan, X. Du, W. Cai, and Q. Wang, “Multiple facial image features-based recognition for the automatic diagnosis of turner syndrome.” Comput. Ind., vol. 100, pp. 85–95, Sep. 2018.

[15] S. Chen, Z.-X. Pan, H.-J. Zhu, Q. Wang, J.-J. Yang, Y. Lei, J.-Q. Li, and H. Pan, “Development of a computer-aided tool for the pattern recognition of facial features in diagnosing turner syndrome: Comparison of diagnostic accuracy with clinical workers,” Sci. Rep., vol. 8, no. 1, pp. 1–9, Dec. 2018.

[16] X. Gao, J. Li, Y. Pei, F. Akhtar, Q. Wang, T. Yang, K. Huang, J. Li, and J.-J. Yang, “Turner syndrome diagnosis with facial features extraction and selection schemes,” in Proc. Int. Conf. Frontier Comput. Singapore: Springer, 2019, pp. 72–78.

[17] G. Yao, J. Li, Y. Pei, F. Akhtar, and B. Liu, “An automatic turner syndrome identification system with facial images,” in Proc. Int. Conf. Frontier Comput. Singapore: Springer, 2019, pp. 105–112.

[18] V. Kazemi and J. Sullivan, “One millisecond face alignment with an ensemble of regression trees,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 1867–1874.

[19] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, Digital Image Processing Using MATLAB. Noida, India: Pearson Education India, 2004.

[20] S. Tambade, M. Somvanshi, P. Chavan, and S. Shinde, “SVM based diabetic classification and hospital recommendation,” Int. J. Comput. Appl., vol. 167, no. 1, pp. 40–43, Jun. 2017.

[21] K. Pearson, “On lines and planes of closest fit to systems of points in space,” London, Edinburgh, Dublin Phil. Mag. J. Sci., vol. 2, no. 11, pp. 559–572, 1901, doi: 10.1080/14786440109462720.

[22] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” J. Educ. Psychol., vol. 24, no. 6, pp. 417–441, 1933.

[23] H. Hotelling, “Relations between two sets of variates,” in Breakthroughs in Statistics. New York, NY, USA: Springer, 1992, pp. 162–190.

[24] B. Schölkopf, A. Smola, and K.-R. Müller, “Nonlinear component analysis as a kernel eigenvalue problem,” Neural Comput., vol. 10, no. 5, pp. 1299–1319, Jul. 1998.
JIJIANG YANG received the B.S. and M.S. degrees from Tsinghua University and the Ph.D. degree from the National University of Ireland, Galway. His research areas involve e-health, e-government/e-commerce, privacy preservation, information resource management, data mining, and cloud computing. He is currently an Associate Professor at Tsinghua University. He worked for the Computer Integrated Manufacturing System/Engineering Research Center (CIMS/ERC), Tsinghua University, from 1995 to 1999. He joined or led different projects funded by the State Hi-Tech program (863 programs), NSF (China), and the European Union. Since 2009, his main focus has been e-health and medical service. He has undertaken a few projects in the National Science & Technology Supporting Program on the Digital Medical Service model and key technologies. He is also collaborating with many medical institutions and hospitals. He is a member of the expert committee of the Internet of Things (IoT) in health at the Chinese Electronic Association, the Expert Committee of remote medicine and cloud computing at the Chinese Medicine Informatics Association. He has published more than 60 articles in professional journals and conferences.

HUI PAN (Senior Member, IEEE) received the Doctor of Medicine degree from the Peking Union Medical College, in 2004. He is currently an Endocrine Professor with the Department of Endocrinology, Peking Union Medical College Hospital. He is engaged in research on medical education and endocrinology.

SHI CHEN received the Doctor of Medicine degree from Peking Union Medical College, in 2007. He is currently an Attending Doctor with the Department of Endocrinology, Peking Union Medical College Hospital. He is engaged in research on medical education and endocrinology.

QING WANG received the Ph.D. degree from Tsinghua University. He became a Researcher with the Research Institute of Information Technology, Tsinghua University. His research interests include web service technology, data mining, and machine learning, among other interests, especially in the healthcare field. Recently, his research has focused on big data technology applied in medical services.

* * *

HUI PAN (Senior Member, IEEE) received the Doctor of Medicine degree from the Peking Union Medical College, in 2004. He is currently an Endocrine Professor with the Department of Endocrinology, Peking Union Medical College Hospital. He is engaged in research on medical education and endocrinology.