

Received October 25, 2020, accepted November 10, 2020, date of publication November 23, 2020,
date of current version December 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039867

An Automated Approach to Diagnose Turner Syndrome Using Ensemble Learning Methods

QING ZHAO¹, GUOHONG YAO¹, FAHEEM AKHTAR^{1,2},
JIANQIANG LI¹, (Senior Member, IEEE),
AND YAN PEI³, (Senior Member, IEEE)

¹Faculty of Information Technology, Beijing University of Technology, Beijing 100024, China

²Department of Computer Science, Sukkur IBA University, Sukkur 65200, Pakistan

³Computer Science Division, University of Aizu, Aizuwakamatsu 965-8580, Japan

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

This study is supported by the National Key R&D Program of China with Project no. 2017YFB1400803.

ABSTRACT This research proposes to use ensemble learning methods to diagnose and predict Turner syndrome using facial images. Turner syndrome, also known as congenital ovarian hypoplasia syndrome, is a common clinical chromosomal disorder. Without the aid of cytogenetic diagnostic results, the accuracy of diagnosis made by the paediatrician is unsatisfactory. Early diagnosis of the Turner syndrome requires the expertise of well-trained medical professionals, which may hinder early intervention due to a high potential cost. So far, most of the studies have reported the use of clinical chromosome detection to diagnose Turner syndrome. In this research, we are the first to use facial recognition technology to diagnose Turner syndrome using ensemble learning techniques. First, the features from each of the facial image are extracted by principal component analysis, kernel-based principal component analysis, and others. Second, we randomly selected samples and features to establish a basic learning model. Finally, we developed a combination of multiple basic learning models using majority voting and stacking for the facial image classification task. Experimental results show that the correct classification rate of the Turner syndrome detection was elevated up to 88.1%. The proposed method can be implemented to automatically diagnosis Turner syndrome patients that can facilitate clinicians during the prognosis process.

INDEX TERMS Turner syndrome, face recognition, ensemble learning, automatic diagnosis, machine learning.

I. INTRODUCTION

Turner is a person's name. In 1938, Dr H. H. Turner recorded the clinical signs from a group of patients. This group of patients has some common physical characteristics which include, females, pediatric hypoplasia, alar neck (also called neck groin), increased facial blepharospasm, elbow valgus, and the primary menstruation-free female disease. Later, the name of Turner was used to name this syndrome as the Turner syndrome (TS). The so-called syndrome in medicine refers to several phenomena that were not adequately understood and recognized at that time. Most of the patients in that group had a chromosomal abnormality. The Normal women's chromosomes should be 46XX, but most of the Turner's syndrome patients chromosomes were 45XO, and few of them

The associate editor coordinating the review of this manuscript and approving it for publication was Yunjie Yang.

were also 45XO and 46XX, which means that few of the patient's chromosomes were not critical at all. [1]. The Turner syndrome (TS) afflicts approximately 50 per 100,000 females and is characterized by retarded growth, gonadal dysgenesis, and infertility [2]. Patients with symptoms of short stature and lack of development are often diagnosed as TS patients. A karyotypic examination is used to recognize the said disease. Moreover, due to the delay in the prognosis, most of the TS patients use to miss the best developmental period of the body during the diagnosis process. The clinical studies have also discerned a decrease in the patient's abilities of spatial conformation, digital computing, and memory. Therefore, it is necessary to identify and diagnose Turner syndrome (TS) as early as possible to prevent and mitigate its adverse outcomes.

Even though the early diagnosis of Turner Syndrome (TS) can help patients to normal physical conditions. People in

underdeveloped regions cannot be treated promptly due to the unavailability of professional expertise and poor economic conditions. On the contrary, commonly used methods for TS diagnosis require chromosome tests. For example, cytogenetic detection is for diffuse oedema in clinical ultrasonography [3] or FISH is for *Y* chromosome detection [4]), which are complicated and expensive for the utmost patients. Therefore, to reduce prognosis cost and to simplify early diagnosis of TS patients, and to improve the genetic diagnosis services in underdeveloped areas, that can bring the gospel to the Turner Syndrome patients, requires the development of an efficient, cost-effective and reliable prognosis process on urgent basis.

During the past several decades, machine learning and deep learning schemes are envisioned to solve problems at various application domains [5], [6] and similarly different researchers have used facial images to recognize various diseases in numerous application domains as well. It has been used in recognizing endocrine diseases such as acromegaly [7], [8], Cushing Syndrome [9], hereditary syndromes such as Down syndrome [10], Derange fever syndrome [11], and in several other diagnoses processes. Where the recognition rate of the system is not lower than the empirical diagnosis of the clinician in some of the reports [7], [11]. Moreover, it is valuable to use facial features for the recognition of patients with endocrine diseases at an early stage. This technique is expected to be used for the detection of endocrine diseases and genetic syndromes to shorten the disease diagnosis time and to assist in the staging of endocrine diseases [12]. However, facial features to diagnose TS is rarely been employed [13]. On the contrary, a single learning model was just used to diagnose the said disease.

With the rapid development of science and technology, the classification tasks that have to be solved in reality (i.e., on real-world medical applications) have become more challenging [14]. Therefore, to develop a classifier with a strong generalization ability has become an arduous task. Ensemble learning is practised to accomplish the said task. The charm of ensemble learning is that it can integrate weak classifiers, which are better than random guesses, to build a robust classifier that can achieve a higher accuracy [15], [16]. The authors in [17], [18] used stacking ensemble model with direct prediction strategy to predict the daily number of cardiovascular diseases admissions and large for gestational age fetuses using data obtained from various sources. Moreover, the application of ensemble learning using facial features can facilitate researchers in solving various medical related problems. A different set of classifiers are used for facial features recognition during the classification step. The experimental results of different colour face data sets show that the facial image has the highest impact on results, and a higher precision can be achieved [19]. Moreover, using the motivation of ensemble learning, the algorithm for facial recognition achieves better results than those of a single classifier [20]. As per different clinical studies, TS patients have several special facial features such as facial paralysis,

wide eye distance, internal epidermis, cervical collar, low hairline, and high arch height. Therefore, based on the advantages of ensemble learning, we propose an ensemble learning framework that uses facial features to recognize and diagnose a TS patient.

It is generally accepted that ensemble learning, i.e., the combination of multiple learning models has a higher possibility than any its constituent learning model to achieve efficient and accurate classification results [15]. This research describes the use of an ensemble learning framework to improve the diagnosis of Turner syndrome through facial features recognition. To achieve classifier diversity (the key to high-quality classification results using ensemble learning techniques [19]), we randomly selected samples and characteristics of the data set. For the training set, a fundamental learning model is constructed using 30 basic learning models, namely support vector machines (SVMs). Later, we used a combination of ensemble methods, majority voting, and stacking with multiple basic classifiers to implement the final face image classification. The contribution of this paper can be summarized as follows: (1) This research is novel in a sense that it uses facial features besides collection of tuned ensemble learning schemes for the establishment of an efficient TS Prognosis process, whereas, previous reproaches lacks in applying machine learning and deep learning schemes. (2) An ensemble learning method is proposed to classify images, which includes studying the learning model for the construction of two basic classifiers. Two widely used ensemble methods, stacking and majority voting, are used to combine multiple basic classifiers with improving the diagnostic of the Turner syndrome. (3) Stacking and random forest methods were analyzed by the combination of perturbation mechanisms using attribute and data samples.

Following this introduction section, section II introduces related work and techniques. We make a brief review of the study of TS and its diagnosis. In section III, we present our ensemble learning method to diagnose and predict the TS. Image processing methods, feature selection, and ensemble strategy are involved in our proposed methods. We evaluate the proposed method in section IV. Section V discusses the experimental results and section VI summarizes the application of ensemble learning methods in the diagnosis of the TS.

II. RELATED WORK AND TECHNIQUES

In recent years, facial image recognition technology has extensively been used for the detection of various syndromes. In medical diagnostic applications, facial recognition software analyzes the patient's facial features and compares it with measurement data of the specific landmarks on the patient's face in the disease database to determine the patient's disease type.

In 2011, the authors Harald J *et al.* used the facial image diagnosis aid (FIDA), which is a face classification software to differentiate in-between patients and normal populations. The software analyzed images using the Gabor

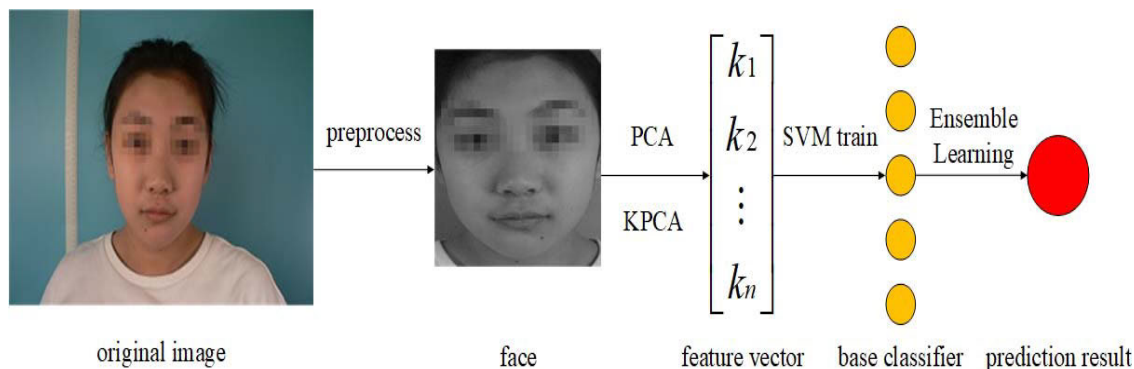


FIGURE 1. Framework of our work includes: (1) Face images preprocessed; (2) Feature extraction by PCA and KPCA; (3) Basic learning models construction; (4) Ensemble of multiple base learning models.

wavelet transformation and used the leave-one-pass verification method to classify them. The recognition rates of experimental and control groups were 71.9% and 91.5%, slightly higher than medical experts (63.2% and 80.8%, respectively) and significantly higher than those of general practitioners (42.1% and 87%) [8]. Ralph *et al.* used the support vector machine to classify the photos of the patients with acromegaly and normal populations. The accuracy of this method reached up to 86% [7]. In 2013, Kosilek *et al.* applied Gabor wavelet filtering to analyze the facial features by comparing differences of texture and geometry within a grid of nodes between photos of the patients and normal populations, which successfully detected 91.7% of the patients with Cushings Syndrome [9]. In 2014, researchers used several different classifiers (including support vector machines, K-nearest neighbours, random forests, and linear discriminant analysis) to identify patients with Down Syndrome and achieved the highest accuracy rate of 96.7% [21]. In 2015, Chen *et al.* proposed a method to extract and identify multiple facial features to diagnose chronic fatigue syndrome (CFS) in patients. This method extracted several facial features by Gabor wavelet filtering and classified them as CFS patients and healthy volunteers. The accuracy achieved by this method was 88.32% [10]. In 2016, Lina *et al.* analyzed facial images of the patients with CdLS (Cornelia de Lange Syndrome) using the FDNA (Facial Dysmorphology Novel Analysis) technology. The correctly classified patients were 87% of the total samples, which was slightly higher than the detection rate by a human expert (77%) [11]. In 2017, Song *et al.* found 68 facial feature points as a basis of Turner Syndrome automatic diagnosis, he used SVM and AdaBoost algorithm for the detection of TS patients and achieved an accuracy score of 84.6% [13]. Schuring *et al.* discuss issues and prevention measures associated with Turner Syndrome [22]. Hyun Yoo *et al.* discusses short stature issue of a 10 years old girl due to lack of 45,X/47,XXX chromosome. Furthermore, almost of presented work lacks in having sufficient dataset and also lacks in the implementation of machine learning and deep learning schemes [23]–[26]. Therefore, this research is almost novel in the sense that it used machine learning schemes to TS prognosis.

Although face recognition technology has extensively been used in the diagnoses process of various syndromes, with the use of facial diagnostic software, the accuracy rate did not achieve the desired score. Whereas, immense development in the face recognition technology has led to continuous improvement in the diagnosis process and proved best in the identification of numerous diseases. Moreover, the research to identify Turner syndrome using facial features is at its infancy stage; and the reported accuracy is very low. Therefore, to improve the automatic diagnosis results of the TS, we need to help clinicians in improving the prognosis process.

The primary objective of an ensemble learning is to build a multi-basic learning model that combines and to solve the same problem with expedited accuracy score [15], [27] [28]. If each of the basic learning models is considered as an expert, multiple experts may be better than any single expert, who provided their judgment is appropriate. Since the integration idea has a great potential for reducing the learning bias of the underlying learning model, it shows better performance in many classification tasks than a single basic model [27]. In this research, we propose an ensemble learning framework to automate TS diagnosis. It combines multiple basic learning models to enhance the prediction accuracy of the TS.

III. DIAGNOSIS AND PREDICTION OF TURNER SYNDROME USING ENSEMBLE LEARNING

Ensemble learning combines multiple weak supervised models to obtain a more comprehensive strong supervised model, it often yields significantly better generalization results than a single learner [29]. In this work, we propose an automatic diagnosis method for the diagnosis of TS using ensemble learning techniques. The framework of this research is shown in Figure 1. First, we preprocess the original face image by Multi-task Cascaded Convolutional Networks (MTCNN) [30] and turn it into a gray-scale image. In the feature selection and extraction process, Principle Component Analysis (PCA) and Kernel PCA (KPCA) are used to extract the feature vector of each image. Second, we trained different basic classifiers with SVM using different kernel functions and parameter settings. SVM is elected because it is proved to be the most popular classifier in recent years, which

TABLE 1. Presents additional information related to possible viewpoints based on possible areas where forensic investigation could be performed.

Symbol	Meaning
D	Dwarf system database.
D_{temp}	Temporary file to hold the dataset.
$\varphi(D_{temp})$	Function to observe key points in human face using MTCNN.
D_{mtcnn}	Holds result of $\varphi(D_{temp})$.
$\phi(D_{mtcnn}, 2960 * 1920)$	Scales the obtained result to a 128*128 resolution image format.
D_{scaled}	Hold results of $\phi(D_{mtcnn}, 2960 * 1920)$.
$\Phi(D_{scaled}, 128 * 128)$	Fucniton to scale image to 128*128 resolution.
D_{rgb}	Hold results of $\Phi(D_{scaled}, 128 * 128)$.
$\Psi(D_{rgb}, 0.2989*R + 0.5870*G + 0.1140*B)$	Function to convert the image to Gray Scale format.
D_{gs}	$\Psi(D_{rgb}, 0.2989*R + 0.5870*G + 0.1140*B)$.

TABLE 2. Kernel functions used in this article.

Name	Expression	Parameter
Gaussian (RBF) kernel function	$k(x_i, x_j) = \exp(-\frac{\ x_i - x_j\ ^2}{2 * \sigma^2})$	$\sigma > 0$
Laplace kernel function	$k(x_i, x_j) = \exp(-\frac{\ x_i - x_j\ }{\sigma})$	$\sigma > 0$
Polynomial and PolyPlus	Polynomial : $k(x_i, x_j) = (x_i * x_j)^d$ Poly-Plus : $k(x_i, x_j) = (x_i * x_j + r)^d$	$d \in Z^+ r = 1$

has a solid mathematical and fundamental theory. It uses the maximum hyperplane margin concept to promote a reduction in the error rate. It is generally analyzed for linearly separable cases. Moreover, various kernels were also used as a mean of diversity as presented in Table. 2 and kernel with best performance results was elected. Finally, random forest and stacking techniques are utilized to predict the labels of testing samples. Algorithm. 1 presents the comprehensive detail of the feature extraction process.

Algorithm 1 Identification of Facial Features From Turner Syndrome Dataset

```

Input: DWARF System Database ( $D$ ).
Output: Features for Classification

1 LOAD  $D_{temp} \leftarrow D$ 
2 For every record in  $D_{temp}$ 
3    $D_{mtcnn} \leftarrow \varphi(D_{temp})$ 
4    $D_{scaled} \leftarrow \phi(D_{mtcnn}, 2560 * 1920)$ 
5    $D_{rgb} \leftarrow \Phi(D_{scaled}, 128 * 128)$ 
6    $D_{gs} \leftarrow \Psi(D_{rgb}, 0.2989 * R + 0.5870 * G + 0.1140 * B)$ 
7 End For;
8  $PCA_{features\_res} \leftarrow \psi_{pca}(D_{gs})$ ;
9  $KPCA_{features\_res} \leftarrow \psi_{kpca}(D_{gs})$ ;
10:  $mRMR_{features\_res} \leftarrow \psi_{mrmr}(D_{gs})$ ;
11:  $ANN_{features\_res} \leftarrow \psi_{ann}(D_{gs})$ ;
12: END
    
```

The symbols detail are presented in Table. 1

A. FACIAL IMAGE PREPROCESSING

1) DATA ACQUISITION

The dataset is collected from the Dwarf System Database of the Endocrinology Department of Beijing Union Medical

College Hospital [12]. During the data acquisition process, Panasonic brand DMC-FZ5 camera is used to capture images and one room is designated for the said task. The images are stored in the JPEG format, and the resolution is set to 2560*1920 with a bit variation in a few of the images. The background of the image is set to blue using a plastic paper with a calliper for measuring height, which is used on the left side of the blue plastic paper. The person stands upright against the wall, looking at the camera with a neutral expression, whose forehead and ears are exposed. Moreover, due to the indoor environment, the light conditions are complex, and light authorization is uneven. The position of the camera and the person is set consistent to ensure a similar distance. The dataset used in the experimentations contains 767 facial images of the children under the age of 14. It includes images of 113 TS and 654 non-TS patients. Figure. 2 represents positive and control images used in the experiment: (a) is the sample TS, and (b) is that of Non-TS.



FIGURE 2. An example of facial images for Control and experimental groups (a) TS (b) Non-TS.

2) IMAGE PREPROCESSING

Once the data is obtained from the Dwarf System Database of the Endocrinology Department of Beijing Union Medical

College Hospital. Endocrinologists manually label each patient's image as a TS or non-TS. As discussed earlier, the resolution of the data collected by the hospital is almost 2560*1920, with minor variations in specific image resolutions, which is unfavourable for the later experimentations. As per the needs of the experimental process, we preprocessed the original data and converted all of the images to a 128*128 grayscale image in experiments. The Illumination, posture, or occlusion of the facial image may also influence the collection process. Therefore, in the preprocessing step, the facial images are normalized. The following steps are used to preprocess the obtained facial images.

- **Rename:** First, we renamed all of the facial images and removed patients with personal information to protect their personal privacy. This is an extremely desired step for later experimentations.
- **Geometric normalization:** Face detection and alignment are indispensable parts of many face-based applications. We used Multi-task Cascaded Convolutional Networks (MTCNN) to detect and roughly locate key points in a human face [31]. Since this study does not require the information of the five feature points. Therefore, the parts related to finding feature points are not taken into consideration.
- **Gray-scale normalization:** In order to weaken the influence of light intensity and to improve the recognition rate, we performed gray-scale normalization on the facial image. We converted RGB values to gray-scale values by using: $\text{Gray} = 0.2989 * R + 0.5870 * G + 0.1140 * B$. Here, the coefficient for calculating the gray-scale values are taken similar to the coefficient of calculating luminance.

B. FEATURE EXTRACTION

To represent information of the original data, facial features were extracted, we used principal component analysis (PCA) and kernel-based principal component analysis (KPCA) for feature extraction. Principal component analysis (PCA) [32] is an algebraic feature analysis method. The primary objective of PCA is to project a high-dimensional vector through a special feature vector matrix into a low-dimensional vector space to represent it as a low-dimensional vector [33], [34]. In this work, 95% (contribution rate) of the original data was retained during the feature extraction process.

During the face recognition step, a one-dimensional vector is obtained by cascading the lines of a two-dimensional human face gray-scale matrix. There are M (767) long vectors of length N (128*128=16384) in the sample set, which means the following (p_j is the gray value of the pixel): $x_i = [p_1, p_2, \dots, p_N]^T$ ($i = 1, 2, \dots, M$). Let m be the mean of $\{x_i | i = 1, \dots, M\}$, which is $m = \frac{1}{M} \sum_{i=1}^M x_i$; ω_i is the deviation of each vector from the mean, and $\omega_i = x_i - m$. There, $W = [\omega_1, \omega_2, \dots, \omega_M]$ calculates the covariance matrix $C = WW^T$. The orthogonal unit basis formed by each feature vector of the matrix C corresponds to a set of projection coordinates for ω_i . Since the size of the matrix, C is

$N * N$ (16384*16384), solving eigenvalues and eigenvectors will consume a considerable amount of resources, so we used the singular value decomposition (SVD) method for the calculation. For any of a matrix $A \in R^{m \times n}$, which can be written as $A = UDV^T$, where $U \in R^{m \times n}$, $V \in R^{n^2}$, $D \in R^{n^2}$, and D is a diagonal matrix. The column of U is the feature vector of AA^T , and the column of V is the feature vector of $A^T A$. The largest K columns in V are the largest K eigenvectors of the $C = WW^T$ that we require, which reflects the maximum dissimilarity in the facial images.

We also use the kernel principal component analysis (KPCA) method as a feature extraction method in this work. The primary objective of using kernel function is to project the input space to a high-dimensional space through a nonlinear function that performs data processing in the feature space [35], [36]. KPCA is a nonlinear extended algorithm of linear PCA [37], which uses nonlinear methods to extract principal components, i.e., the feature in our work. The complexity of the kernel-based principal component analysis method depends on samples. The time and memory required for its calculations are not related to the dimensions of input space, but it is closely related to the number of samples. The contribution of each sample point to the dimension reduction is not similar. Therefore, the number of samples is reduced by the method of filtering the samples with feature vector $V = \{v_1, \dots, v_m\}$ that correspond to the first principal element (where m is the number of samples). The implementation is described as follows:

- 1) The matrix corresponding to the face image is expanded into column vectors $X = \{x_1, x_2, \dots, x_n\}$;
- 2) Appropriate kernel functions with parameters are selected to train the samples, kernel matrix K from (1) is obtained using the kernel matrix \bar{K} by Equation (2).
- 3) Finally, the eigenvalues and eigenvectors of the kernel matrix \bar{K} are obtained. The eigenvalues are sorted in ascending order to find the eigenvectors corresponding to the eigenvalues that contained 95% of the information.

$$K_{ij} = K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \quad i, j = 1, 2, \dots, N \quad (1)$$

$$\bar{K}_{ij} = K_{ij} - \frac{\sum_{m=1}^M l_{im} k_{mj}}{M} - \frac{\sum_{n=1}^M k_{in} l_{nj}}{M} + \frac{1}{M^2} l_{im} k_{mn} l_{nj} \quad (2)$$

Furthermore, to signify the importance of the proposed scheme, we introduced a widely used feature selection scheme named as Minimum Redundancy - Maximum Relevance (mRMR). mRMR selects features that are highly correlated with the class and having a low correlation between themselves. Besides, Artificial neural network (ANN) with multiple hidden layers and neurons is also exploited for the comparative perspective.

C. BASIC LEARNING MODEL CONSTRUCTION

After obtaining feature vectors from the facial image, the extracted features are given to the classifier to classify them as TS or non-TS. It is a critical step to select an appropriate classifier for obtaining fruitful classification results. SVM is proved to be the most popular classifier in recent years, which has a solid mathematical and fundamental theory. It uses the maximum hyperplane margin concept to promote a reduction in the error rate. It is generally analyzed for linearly separable cases.

For the classification task, we used the Support vector machine (SVM) as a single basic learning model. To construct a nonlinear classifier, we used various kernel techniques that help in the development of the largest hyperplane. To implement nonlinearity, SVM first completes the computation in the low-dimensional space. Second, it projects the input space to the high-dimensional feature space using a kernel function. Finally, an optimal separating hyperplane is constructed in the high-dimensional feature space, to separate the nonlinear data that are not well-divided in the hyperplane. The key to solving the linear inseparability problem at this time is to select an appropriate kernel function and its parameter. The kernel functions in Table 1 are mainly used in this work.

D. ENSEMBLE STRATEGY OF MULTIPLE BASE LEARNING MODELS

It has previously been discerned in various studies that a combination of classifiers always performs better than a single classifier [19]. As the assumption space for learning tasks is often huge, there may be multiple hypotheses that achieve the same performance on the training set. Using a single learner may result in poor generalization performance, and a combination of multiple learners may reduce this risk. Moreover, learning algorithms tend to fall into a local minimum, where, the localization performance of a few of the local minimum points can be inferior. The combination of multiple runs can reduce the risk of getting into a local minimum. The actual assumptions of some learning tasks may not be in the hypothesis space considered by the current learning algorithm. Therefore, this research proposes two standard ensemble methods - majority voting and stacking methods for facial image classification tasks with the integration of multiple learning models.

1) MAJORITY VOTING

The ensemble method means that the meta-classifier has better generalization performance than a single classifier with a combination of different classifiers in a single meta-classifier. Majority voting is one of the most commonly used ensemble methods. The majority voting principle refers to the results of most classifier predictions as the final prediction class [15], [19]. The theoretical basis of Majority voting is the well-known Condorcet's theorem. The theorem states that when the probability of each voter making a correct decision is higher than 50%, adding more voters will increase the

likelihood that the majority decision is right. Because of its simplicity and excellent performance, majority voting has become a trendy combination scheme. In this paper, we adopt it as one of the integrated solutions for the facial image classification.

2) STACKING

The stacking is the abbreviation of Stacked Generalization [38]. Stacking is one of the most influential methods during the integration process. The two-level structure is applied to the stack: Meta-level classifier generates the final decision by using the output of the base-level classifier as an input. The basic idea of stacking is to combine different classifiers from different classification algorithms, such as decision tree, multi-layer propagation, and naive Bayes to generate a higher level classification system. The diversity of base-level classifiers is significant for generating an ensemble. The algorithm of generating classifiers applies to different assumptions. Therefore, their errors and bias differ from each other.

The stacking uses a meta-level classifier to map the output of the base-level classifier to the final decision. The output of each base-level classifier's training instance will be treated as a stand-alone attribute, and the actual class label of the instance will be treated as a dependent attribute. For all training instances, a new training set is generated to train the meta-level classifier. When all the training processes of the base-level classifier and the meta-level classifier are completed, a stacking assembly is obtained. To classify new instances, the meta-level classifier takes the prediction of the base-level classifier as its input and uses its prediction as to the decision. During the Stacking process, most of the time it spends on selecting meta-level data or algorithms to generate a meta-level classifier [39], [40]. Fig. 3 illustrates the two-layer structure of a stack-based integration method for the face based image classification. By constructing a new data set with PCA and SVM. Firstly, 40 base classifiers are used to output the classification results SVM is used to construct a new model for the new data set. Finally, predictions to generate the final classification results are made.

IV. EXPERIMENTAL SETUP

A. EVALUATION METRICS

Based on previously published researches, we selected Accuracy, Sensitivity, and Specificity as the performance evaluation metrics. The details of their formations based on confusion matrix are as follows.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (3)$$

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (4)$$

and,

$$Specificity = \frac{TN}{(TN + FP)} \quad (5)$$

where TP, TN, FP, and FN are true positives, true negative, false positive, and false negative respectively.

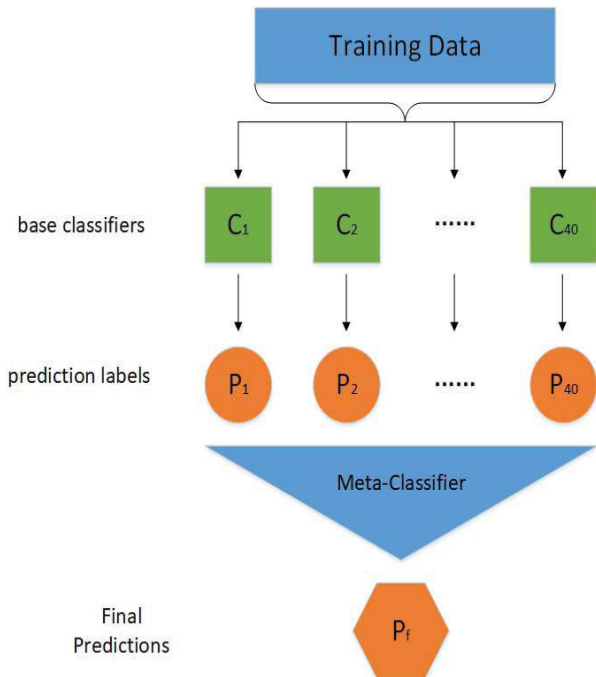


FIGURE 3. Two-layer structure of stacking for face image classification. ($C_1 \dots C_{40}$) express fixed number of samples and features are randomly selected to construct 40 base classifiers, and the prediction labels ($P_1 \dots P_{40}$) of 40 base classifiers are obtained. The tag vector is used as a new sample to construct a new training data set, and the classifier is trained to obtain the final classification model (P_f).

B. EXPERIMENT DESIGN

The Intel i5-6500 3.2GHz quad-core CPU with 16 gigabytes of memory, a graphics card with NVIDIA GTX1070 8GB, and Windows 10 x64 Education Edition is used for the experimentations. The preprocessing step of the facial images is mainly normalized by the MTCNN network and implemented using Python language. Feature extraction and classification is implemented using the MATLAB tool. Later, written our definitions for the functions of PCA, KPCA, SVM, and a variety of kernels without the interface provided by the pattern recognition toolbox of MATLAB. In this paper, ensemble forests and stacking are used to implement ensemble learning methods for the diagnosis of TS. Random Forests grows many classification trees. To classify a new object from an input vector, put the input vector down each of the trees in the forest. Each tree gives a classification, which we call “votes” of the tree for a particular class. The forest chooses the classification result, considering the majority votes obtained (over all the trees in the forest) [41]. Random forest is a subclass of ensemble learning. It depends on the voting choice of the decision tree to determine the final classification result. In this paper, the implementation of a random forest is as follows: A total of 767 samples are included in the experiment (experimental group: 113, control group: 654), and then each pre-processed grayscale image (128*128) was expanded into a 16384-dimensional vector. Next, a total of 30 sub-training sets are constructed. The construction process of each sub-training set is as follows:

- 1) From the 767 training set, 100 positive samples and 100 negative samples (this sampling method is called the bootstrap sample method) are randomly selected as the sub-training samples, and the remaining data sets are used as sub-test samples;
- 2) Principal component analysis (PCA) and Kernel Principal Component Analysis (KPCA) are used to reduce the sample dimension, and we used the $Wy + \mu$ method to refactor it and selected only those that induced at least 95% of the amount of information by reserving 485 and 200 dimensions. At this stage, KPCA's kernel functions include Gaussian, Laplacian, and polynomial kernels. From the feature dimension after dimensionality reduction, 100-dimensional features are randomly selected for training. The primary motivation behind using PCA is to reduce high dimension data into low dimension data. It is a commonly used technique in face recognition and image compression tasks [42].
- 3) SVM models are trained for 30 times. selected the optimal parameters of each model by adjusting the parameters of the kernel function. Final prediction results are obtained using prediction results of the 30 classifiers based on majority votes.

Stacking (sometimes called stacked generalization) refers to train a model with the combination of other models [43]. We train several different models; used the output of each previously trained model as input to train a new model to obtain the final output. The main idea of stacking is as follows: First, all trained base models are predicted for the entire training set. The predicted value of the $j - th$ base model for the $i - th$ training sample is taken as the $j - th$ of the new training set for the $i - th$ sample. The eigenvalues are finally trained based on the new training set. In the same way, the forecasting process also needs to generate a new test set through all the base model predictions, and finally predict the test set. In this study, firstly, for the 767 samples in the experiments (experimental group: 113, control group: 654), each pre-processed grayscale image (128*128) was expanded into a 16384-dimensional vector form. The model building process is divided into the following steps.

- 1) We randomly selected 2/3 of the facial images from the positive samples (113) and negative samples (654) in the data set. A total of 511 images are used as the training set, and the remaining samples (256 samples) are used as the test set.
- 2) Principal component analysis (PCA) and kernel principal component analysis (KPCA) were used to reduce the dimension of the sample to 485 and 120 dimensions respectively. At this time, KPCA's kernel functions include Gaussian kernel, Laplacian kernel, and polynomial kernel.
- 3) We build 40 sub-training sets. The construction process of each sub-training set are as follows: a) From the total training set, we randomly selected P positive samples, N negative samples ($P=100, N=100; P=50, N=150$) as the sub-training samples, and the remaining data

- sets as the sub-test samples; b) From (485 and 120) dimension, randomly selected F features (F=100, 200, 50; F=100, 200, 50);
- 4) We trained 40 SVM models (including 10 Gaussian, 10 Laplace, 10 Polynomial, and 10 Polyplus), and selected the optimal parameters for each model.
 - 5) Constructed a new data set: 40 SVM sub-models are used to predict the corresponding sub-test samples. Each sub-test sample obtained 40 prediction tags corresponding to the new data corresponding to the sample. To build a new total training set on 511 images, the same reason, to build a new test set of 256 images;
 - 6) New model training: SVM training is performed on the new training set of 511 images to obtain the model and make predictions. Finally, the optimal results for each of the SVM are obtained.

TABLE 3. Shows the experimental results of the TS classification task.

Measures Classifiers	TS(%) sensitivity	Non-TS(%) specificity	accuracy(%) accuracy
PCA	74.6	73.5	74.2
KPCA	73.2	69.7	70.5
SVM	80.4	83.5	81.9
KCPCA_SVM	79.8	83.2	82.3
PCA_SVM	83.1	82.4	83.2
ANN	52.9	72.4	65.2
mRMR_SVM	48.8	70.0	59.4

V. DISCUSSION AND ANALYSIS

A. THE PERFORMANCE OF BASE CLASSIFIERS

There are only 76 samples tag, the 3-fold cross-validation method is used to evaluate the performance of the classifier. All facial images are equally divided into 3 subsets, each of them contains 255 images. In each experiment, 2 subsets were selected as the training set, and the remaining subset was used as the test set to evaluate the basic classifier. In this study, seven independent classifiers are established through the combination of feature extraction methods i.e., PCA, KPCA features, and mRMR with base classifier SVM and Artificial Neural Network (ANN). The best experimental results of the seven classifiers for the classification task of TS are listed in Table 3. Based on Table 3 results, PCA_SVM performed best while detecting TS, and achieved an average accuracy score of 83.2%. Besides, KPCA performed worst in the detection of TS. Furthermore, mRMR with SVM and mRMR with ANN did not perform well compared to the proposed scheme. The reason for the low performance of mRMR with SVM might be because of insufficient numbers of features selected for the classification task, or the performance might increase if further ranked features are added during the experimentation process. But, due to high dimensional data of TS, it would be difficult to elect further features using the mRMR feature selection scheme which

already took more than 10 days to select 100 features from the TS dataset. Whereas, the reason for ANN with mRMR being worst might be because of inappropriate hidden layers and neurons selections keeping in view lesser features selected with mRMR feature selection scheme. ANN can perform better but it might need an increased number of features and one to infinite possibilities of hidden layers and neurons selection which makes in complex and computationally infeasible. Comparing with the deep learning method (i.e., ANN), our proposed diagnosis framework can solve the problem with less computational complexity because the complexity of kernel-based method depends on the number of samples.

B. THE PERFORMANCE OF ENSEMBLE CLASSIFIERS

The 3-fold cross-validation method is used for the performance evaluation of the two integrated methods. Table 3 shows the experimental results of the TS classification task. By comparing the majority voting and stacking methods as the base classifiers; the performance of the stacking method outperformed the majority voting and achieved an accuracy score of 88.1%. The experimental results show that the proposed method is superior to the traditional method in classification accuracy, especially in sensitivity performance metrics. Comparing the performance of the base classifier (shown in Table 3) and its collection classifier (shown in Table 4), the Random Forest and Stacking methods outperform the classifier based on the best single learning model in the TS detection tasks.

TABLE 4. Performance results of Stacking and Majority voting schemes.

Measures Classifiers	TS(%) sensitivity	Non-TS(%) specificity	accuracy(%) accuracy
Majority voting	82.6	83.5	83.2
Stacking	89.7	87.5	88.1

We discuss the effects of PCA (KPCA) feature extraction and the dimensionality of randomly selected features on the final integrated model classification results during the implementation of random forest and Stacking, which can be seen in the Fig: 4.

Because of the current diagnostic research of Turner syndrome compared with the limitations of traditional clinical diagnosis. This paper proposes an ensemble learning method based on facial features recognition to help Turner's clinical diagnosis using the random forest algorithm and Stacking method. As per Fig V-B, when using the PCA (KPCA) feature extraction to retain 95% of the information reduced to 500 (200) The Random Forest and Stacking Constructed training set is randomly selected when the 200 (110) dimension feature, the classifier's can obtain the best performance.

Since the experimental data of this research contains a large difference between the number of training samples of positive and negative samples. It will cause a low sensitivity rate. Therefore, a random forest improvement algorithm based on putting back random re-sampling and ensemble learning is

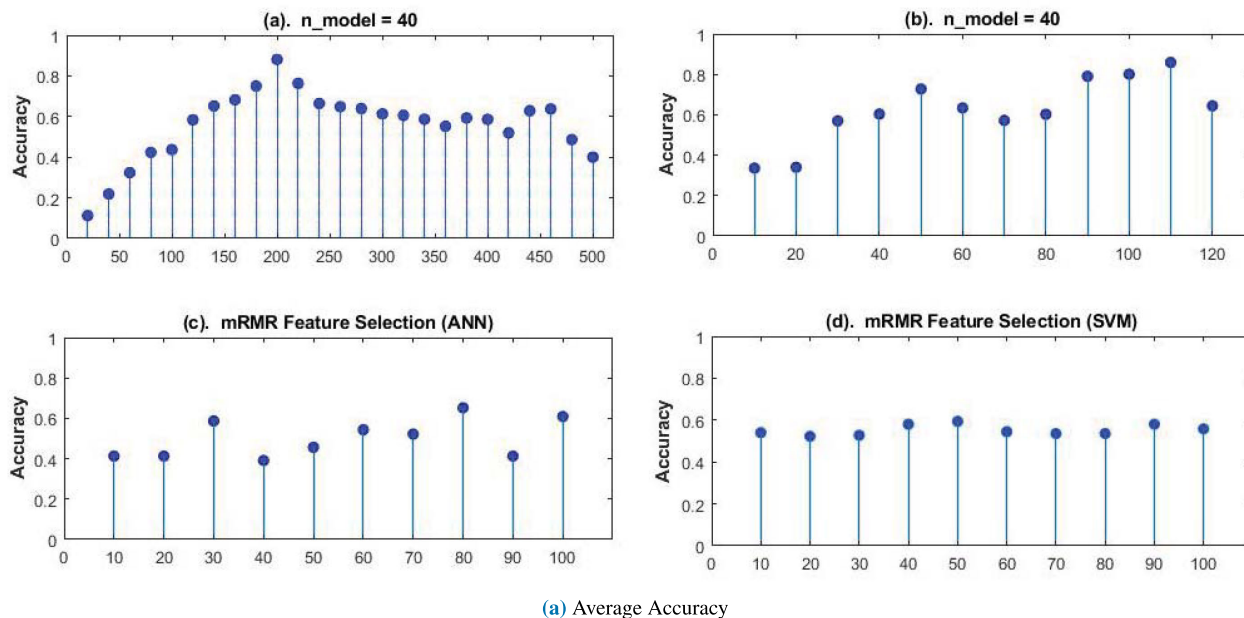


FIGURE 4. Using PCA feature extraction to retain 95% of the information to reduce the dimension to 485, the random forest and Stacking constructor training set are randomly selected 200-dimensional features, the classifier performance is the best. (b): Using KPCA feature extraction to retain 95% of the information to reduce the dimension to 120, the random forest and Stacking constructor training set is randomly selected 110-dimensional features, the classifier performance is the best. (c): Using Artificial neural network with various hidden layers and neurons with 100 ranked features using the mRMR feature selection scheme. (d): Using Using SVM with various kernels with 100 ranked features using the mRMR feature selection scheme.

applied. Random forest classification models trained on each data set, by constructing multiple categories of balanced data sets with random re-sampling. Multiple classification models determine the category of pending samples by majority votes. In the process of training random forest-based classifiers, the decision tree introduces a random feature variable selection technique when nodes are divided. Not all the features participate in node splitting, but certain randomly selected features. By this method, the correlation is reduced between each decision tree in the forest, and the classification accuracy of each decision tree is improved; ultimately, it improved the performance of the entire forest. Performing random re-sampling with reversal can enhance the diversity of the base classifier in the “forest”. The ensemble of the base classifier can compensate for the information loss and “over-fitting” caused by random sampling.

C. COMPUTATIONAL COMPLEXITY

Regarding the computational complexity, since the ensemble learning method is the integration of multiple basic learners and is constructed during the execution period, the computational cost of most voting methods should be the sum of the basic learners involved. In the stacking method, we should also add the time for class 1 classifier’s training. During the execution time, the cost of calculation of most of the voting methods (stacking methods) is the decision time of the basic learning model involved plus the time of the majority voting process (classification of first decision time).

VI. CONCLUSION AND FUTURE WORK

In this paper, we propose to use an ensemble learning classification method to diagnose and predict Turner syndrome. It is also obvious from the results of Table 3 that the ensemble learning method proposed in this research has a good effect on the automatic diagnosis of the Turner syndrome, namely random forest and stacking. Stacking is proved to be the best choice. It is a typical combination of classifiers. The experimental results show that the proposed method can obtain better classification results compare to a base classifier. However, the final classification results are not ideal at all. Moreover, after observing the experimental samples, it is also found that there is a close relationship in-between the positive and negative characteristics of the sample. We hope that in our future research, it will be a very meaningful research direction to extract the most discriminative TS characteristics to assist the clinician in the prognosis process.

Furthermore, this research proposes an ensemble learning method that combines multiple heterogeneous learning models to accurately predict TS from facial images. We used 95% of the information retained by PCA (KPCA) to extract features, SVM is used as the basic learning model. Whereas, mRMR feature selection scheme and Artificial deep neural network (ANN) is also exploited for the comparative perspective. Multiple voting (random forest) and Stacking integration methods were also used; where multiple sets of classifiers were combined to improve Turner syndrome identification. In this study, the accuracy of our integration method is elevated up to 88.1%. It is also seen that the automated Turner

Syndrome diagnosis method based on facial images has a significance in the early diagnosis of the special syndrome. It can also be used for early prognosis of other chromosomes or genetic diseases. We will hopefully investigate these subjects in our future research.

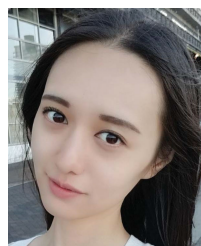
ACKNOWLEDGMENT

The authors declare that they have no conflicts of interest. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants involved in the study. The study protocol and consent have been reviewed by the School of Software, Beijing University of Technology.

REFERENCES

- [1] C. H. Gravholt, N. H. Andersen, G. S. Conway, O. M. Dekkers, M. E. Geffner, K. O. Klein, A. E. Lin, N. Mauras, C. A. Quigley, K. Rubin, D. E. Sandberg, T. C. J. Sas, M. Silberbach, V. Söderström-Anttila, K. Stochholm, J. A. van Alfen-van der Velden, J. Woelfle, P. F. Backeljauw, and International Turner Syndrome Consensus Group, "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, Sep. 2017.
- [2] N. Frezel, F. Cassim, P. Derambure, and P. Bocquillon, "Parsonage-turner syndrome: Topographic diagnosis," *Pratique Neurologique-FMC*, vol. 6, no. 3, pp. 212–217, 2015.
- [3] S. Carpini, A. B. Carvalho, G. Guerra-Junior, M. T. Baptista, S. H. Lemos-Marini, and A. T. Maciel-Guerra, "Spontaneous puberty in girls with early diagnosis of turner syndrome," *Arquivos Brasileiros De Endocrinologia E Metabologia*, vol. 56, no. 9, pp. 653–661, 2012.
- [4] R. E. Miller, E. G. Learned-Miller, P. Trainer, A. Paisley, and V. Blanz, "Early diagnosis of acromegaly: Computers vs clinicians," *Clin. Endocrinology*, vol. 75, no. 2, pp. 226–231, Aug. 2011.
- [5] J. Long, S. Zhang, and C. Li, "Evolving deep echo state networks for intelligent fault diagnosis," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4928–4937, Jul. 2020.
- [6] J. Long, Z. Sun, C. Li, Y. Hong, Y. Bai, and S. Zhang, "A novel sparse echo autoencoder network for data-driven fault diagnosis of delta 3-D printers," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 3, pp. 683–692, Mar. 2020.
- [7] H. J. Schneider, R. P. Kosilek, M. Günther, J. Roemmler, G. K. Stalla, C. Sievers, M. Reincke, J. Schopohl, and R. P. Würtz, "A novel approach to the detection of acromegaly: Accuracy of diagnosis by automatic face classification," *J. Clin. Endocrinol. Metabolism*, vol. 96, no. 7, pp. 2074–2080, Jul. 2011.
- [8] R. Kosilek, J. Schopohl, M. Grunke, M. Reincke, C. Dimopoulou, G. Stalla, R. Würtz, A. Lammert, M. Günther, and H. Schneider, "Automatic face classification of Cushing's syndrome in women—A novel screening approach," *Experim. Clin. Endocrinol. Diabetes*, vol. 121, no. 9, pp. 561–564, Jul. 2013.
- [9] Y. Chen, W. Liu, L. Zhang, M. Yan, and Y. Zeng, "Hybrid facial image feature extraction and recognition for non-invasive chronic fatigue syndrome diagnosis," *Comput. Biol. Med.*, vol. 64, pp. 30–39, Sep. 2015.
- [10] L. Basel-Vanagaite, L. Wolf, M. Orin, L. Larizza, C. Gervasini, I. D. Krantz, and M. A. Deardoff, "Recognition of the cornelia de lange syndrome phenotype with facial dysmorphology novel analysis," *Clin. Genet.*, vol. 89, no. 5, pp. 557–563, May 2016.
- [11] R. P. Kosilek, R. Frohner, R. P. Würtz, C. M. Berr, J. Schopohl, M. Reincke, and H. J. Schneider, "Diagnostic use of facial image analysis software in endocrine and genetic disorders: Review, current results and future perspectives," *Eur. J. Endocrinol.*, vol. 173, no. 4, pp. M39–M44, Oct. 2015.
- [12] 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.
- [13] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 3rd Quart., 2006.
- [14] Y. P. F. Akhtar, J. Li, and M. Azeem, "A semi-supervised technique for LGA prognosis," in *Proc. Int. Workshop Future Technol. (FUTECH)*, Taichung City, Taiwan: Korean Institute Information Technology, 2018, pp. 36–37. [Online]. Available: <http://www.dpbia.co.kr/Article/NODE07608966>
- [15] L. Rokach, "Ensemble-based classifiers," *Artif. Intell. Rev.*, vol. 33, nos. 1–2, pp. 1–39, 2010.
- [16] F. Akhtar, J. Li, Y. Guan, A. Imran, and M. Azeem, "Monitoring biochemical indicators using machine learning techniques for an effective large for gestational age prediction model with reduced computational overhead," in *Frontier Computing*, J. C. Hung, N. Y. Yen, and L. Hui, Eds. Singapore: Springer, 2019, pp. 130–137.
- [17] Z. Hu, H. Qiu, Z. Su, M. Shen, and Z. Chen, "A stacking ensemble model to predict daily number of hospital admissions for cardiovascular diseases," *IEEE Access*, vol. 8, pp. 138719–138729, 2020.
- [18] F. Akhtar, J. Li, Y. Pei, A. Imran, A. Rajput, M. Azeem, and Q. Wang, "Diagnosis and prediction of large-for-gestational-age fetus using the stacked generalization method," *Appl. Sci.*, vol. 9, no. 20, p. 4317, 2019.
- [19] F. E. Keskin, D. O. Yetkin, H. M. Ozkaya, O. Haliloglu, S. Sadri, N. Gazioglu, N. Tanriover, H. Ak, E. Hatipoglu, and P. Kadioglu, "The problem of unrecognized acromegaly: Surgeries patients undergo prior to diagnosis of acromegaly," *J. Endocrinol. Invest.*, vol. 38, no. 6, pp. 695–700, Jun. 2015.
- [20] Q. Zhao, K. Okada, K. Rosenbaum, L. Kehoe, D. J. Zand, R. Sze, M. Summar, and M. G. Linguraru, "Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ICA," *Med. Image Anal.*, vol. 18, no. 5, pp. 699–710, Jul. 2014.
- [21] J. Li, C. Liu, B. Liu, R. Mao, Y. Wang, S. Chen, J.-J. Yang, H. Pan, and Q. Wang, "Diversity-aware retrieval of medical records," *Comput. Ind.*, vol. 69, pp. 81–91, May 2015.
- [22] A. Schüring, F. Nawroth, and M. von Wolff, "Turner syndrome," in *Fertility Preservation in Oncological and Non-Oncological Diseases*. Cham, Switzerland: Springer, 2020, pp. 145–156. [Online]. Available: <https://www.springer.com/gp/book/9783030341480>
- [23] F. S. Oberhoffer, H. Abdul-Khalig, A.-M. Jung, T. R. Rohrer, and M. A. El Rahman, "Left ventricular remodelling among turner syndrome patients: Insights from non-invasive 3D echocardiography-derived pressure-volume loop analysis," *Clin. Res. Cardiol.*, vol. 109, no. 7, pp. 892–903, Jul. 2020.
- [24] F. P. Pires, E. L. Monteiro, N. Reais, F. Oliveira, P. A. Carvalho, and A. Miranda, "Parsonage-turner syndrome—report of two clinical cases and literature review," *Int. J. Orthopaedics*, vol. 7, no. 3, pp. 1302–1304, 2020.
- [25] F. Kulakli and S. Duman, "Parsonage-turner syndrome presenting with left shoulder pain and weakness: A case report," *Eur. Res. J.*, vol. 6, no. 4, pp. 359–361, Jan. 2020.
- [26] E. Dabrowski, E. K. Johnson, V. Patel, Y. Hsu, S. Davis, A. L. Goetsch, R. Habiby, W. J. Brickman, and C. Finlayson, "Turner syndrome with y chromosome: Spontaneous thelarche, menarche, and risk of malignancy," *J. Pediatric Adolescent Gynecol.*, vol. 33, no. 1, pp. 10–14, Feb. 2020.
- [27] Z. H. Zhou, "Ensemble learning," in *Encyclopedia Biometrics*. New York, NY, USA: Springer-Verlag, 2015, pp. 270–273. [Online]. Available: <https://www.springer.com/gp/book/9781441993250>
- [28] F. Akhtar, J. Li, M. Azeem, S. Chen, H. Pan, Q. Wang, and J.-J. Yang, "Effective large for gestational age prediction using machine learning techniques with monitoring biochemical indicators," *J. Supercomput.*, vol. 76, no. 8, pp. 6219–6237, Jan. 2019, doi: [10.1007/s11227-018-02738-w](https://doi.org/10.1007/s11227-018-02738-w).
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [30] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
- [31] J. Xiang and G. Zhu, "Joint face detection and facial expression recognition with MTCNN," in *Proc. 4th Int. Conf. Inf. Sci. Control Eng. (ICISCE)*, Jul. 2017, pp. 424–427.
- [32] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Amer. A, Opt. Image Sci. Vis.*, vol. 4, no. 3, pp. 519–524, 1987.
- [33] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 103–108, Jan. 1990.

- [34] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
- [35] B. Scholkopf, S. Mika, C. J. C. Burges, P. Knirsch, K.-R. Müller, G. Ratsch, and A. J. Smola, "Input space versus feature space in kernel-based methods," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1000–1017, Jun. 1999.
- [36] K. I. Kim, K. Jung, and H. Joon Kim, "Face recognition using kernel principal component analysis," *IEEE Signal Process. Lett.*, vol. 9, no. 2, pp. 40–42, Feb. 2002.
- [37] 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.
- [38] A. A. Ghorbani and K. Owrangh, "Stacked generalization in neural networks: Generalization on statistically neutral problems," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2001, pp. 1715–1720.
- [39] S. Dzeroski and B. Zenko, "Stacking with multi-response model trees," in *Proc. Int. Workshop Multiple Classifier Syst.*, 2002, pp. 201–211.
- [40] A. K. Seewald, "How to make stacking better and faster while also taking care of an kernel eigenvalue weakness," in *Proc. 19th Int. Conf. Mach. Learn.*, 2002, pp. 554–561.
- [41] S. Kotsiantis, "Combining bagging, boosting, rotation forest and random subspace methods," *Artif. Intell. Rev.*, vol. 35, no. 3, pp. 223–240, Mar. 2011.
- [42] S. Karamizadeh, S. M. Abdullah, A. A. Manaf, M. Zamani, and A. Hooman, "An overview of principal component analysis," *J. Signal Inf. Process.*, vol. 4, no. 3, p. 173, 2013.
- [43] C. Hernández-Espinosa, J. Torres-Sospedra, and M. Fernández-Redondo, "Researching on multi-net systems based on stacked generalization," in *Proc. IAPR Workshop Artif. Neural Netw. Pattern Recognit.*, Paris, France, 2008, pp. 193–204.



QING ZHAO received the B.S. degree from the Communication University of China, Nanguang College, Nanjing, China, in 2013, and the M.S. degree in software engineering from the Beijing University of Technology, Beijing, China, in 2016, where she is currently pursuing the Ph.D. degree in computer science and technology. Her research interests include semantic analysis and big data mining.



GUOHONG YAO received the M.S. degree in software engineering from the Beijing University of Technology, Beijing, China. She has also published national patents and international conferences. Her research interests include data mining, machine learning, neural networks, and big data.



FAHEEM AKHTAR received the M.S. degree in computer science from the National University of Computing and Emerging Science (NUCES FAST), Karachi, Pakistan, in 2011, and the Ph.D. degree from the Beijing University of Technology, China, in 2020. He is currently working as an Assistant Professor with the Department of Computer Science, Sukkur IBA University, Pakistan. He is the author of various SCI, EI, and Scopus indexed journals and international conferences.

He is part of various indexed international conference at different positions and a Reviewer of various SCI, EI, and Scopus indexed journal. His research interests include data mining, machine learning, deep learning, information retrieval, privacy protection, Internet security, the Internet of Things, and big data.



JIANQIANG LI (Senior Member, IEEE) received the B.S. degree in mechatronics from the Beijing Institute of Technology, Beijing, China, in 1996, and the M.S. and Ph.D. degrees in control science and engineering from Tsinghua University, Beijing, in 2001 and 2004, respectively. From 2004 to 2005, he worked as a Researcher with the Digital Enterprise Research Institute, National University of Ireland, Galway. From 2005 to 2013, he worked as a Researcher with NEC Labs China.

From 2009 to 2010, he was a Visiting Scholar with the Department of Computer Science, Stanford University. He joined the Beijing University of Technology, Beijing, as Beijing Distinguished Professor, in 2013. He has over 40 publications and 37 international patent applications (19 of them have been granted in China, U.S., or Japan). His research interests include Petri nets, enterprise information systems, business process, data mining, information retrieval, semantic web, privacy protection, and big data. He served as a PC members in multiple international conferences and organized the IEEE workshop on medical computing.



YAN PEI (Senior Member, IEEE) received the B.E. and M.E. degrees from Northeastern University, Shenyang, and the Ph.D. degree in engineering from Kyushu University, Fukuoka, Japan. He was a Software Engineer and a Software Project Manager of Chinese and German IT industries for several years before entering academic society. He is currently working as an Associate Professor with the University of Aizu. His research interests include evolutionary computation, machine learning, and software engineering.

He is a Senior Member of the IEEE SMC, IEEE CIS, and the Japanese Society for Evolutionary Computation. He received several awards, such as the Best Paper Awards from the ICGEC 2012, FC 2018, and FC 2019. He serves as the chair and an organizing committee member for many international conferences sponsored by the IEEE, ACM, and other international associations. He is also serving as editors for many international journals.

...