Ensemble Learning for the Detection of Facial Dysmorphology

Qian Zhao, Naoufel Werghi, Kazunori Okada, Kenneth Rosenbaum, Marshall Summar,
and Marius George Linguraru

Abstract—**Down syndrome is the most common chromosomal condition that presents characteristic facial morphology and texture patterns. The early detection of Down syndrome through an automatic, non-invasive and simple way is desirable and critical to provide the best health management to newborns. In this study, we propose such a computer-aided diagnosis system for Down syndrome from photography based on facial analysis with ensemble learning. First, geometric and texture facial features are extracted based on automatically located facial landmarks, followed by feature fusion and selection. Then multiple classifiers (i.e. support vector machines, random forests and linear discriminant analysis) are adopted to identify patients with Down syndrome. An accurate and reliable decision is finally achieved by optimally combining the outputs of these individual classifiers via ensemble learning that captures both the shared and complementary information from different classifiers. The best performance was achieved by using the median ensemble rule with 0.967 accuracy, 0.977 precision and 0.933 recall.**

I. Introduction

Down syndrome (DS), caused by trisomy of chromosome 21, is the most common chromosomal condition. The incidence rate varies from 1:319 to 1:1000 worldwide [1]; in the United States, one out of 691 infants is born with DS and over 400,000 people are living with it, while the rate is as high as 1:350 in UAE [2, 3]. Patients with DS have a high incidence of serious medical complications (e.g. cardiac, respiratory and hearing problems) and intellectual disability that require treatment. Thus, the early detection of DS is fundamental for providing patients with lifelong medical care that may involve specialists in many fields.

Prenatal screening for DS based on ultrasound has an accuracy rate as low as 79% [4]. If a screening test is positive, a more invasive test may be used to confirm the diagnosis. More recently, the non-invasive prenatal test (NIPT) which requires samples of maternal blood has been introduced with very high accuracy and no miscarriage risk [5]. However, access to specialized genetic clinics is limited especially for non-academic centers in rural settings. Moreover, the genetic tests are expensive and associated medical costs are high.

After birth, the diagnosis of DS is often based on a number of physical variations and dysmorphology [6]. These differences may be subtle and influenced by the length of gestation, the effects of labor and delivery and the ethnical backgrounds of the family, making the diagnostic rate as low as 50% - 60% for pediatricians prior to cytogenetic tests [7]. Therefore, the development of an automated, non-invasive and reliable assessment system to detect DS in newborns could increase the diagnostic accuracy and reduce the cost and time associated with genetic tests.

The symptoms of DS mainly present as facial morphology (or geometry) and appearance (or texture) patterns, which pave a way for developing a computer-aided diagnosis system for DS based on photogrammetry. For DS identification, different facial geometric and texture features have been investigated. Gabor wavelet transformation was applied to manually cropped facial image and manually labeled landmarks to discriminate DS from other disorders in [8, 9]. Burçin et al. investigated local pattern patterns (LBP) on non-overlapped blocks covering the entire face [10], which also required manual cropping and pre-processing. In our prior work, we proposed to combine the geometric and local texture features based on facial landmarks to identify DS from a non-syndrome group [11-13], which outperformed methods using geometric or texture features alone. For classification, support vector machines (SVM), *k*-nearest neighbor (*k*-NN), random forest (RF), linear discriminant analysis (LDA) and simple template matching have been investigated individually and separately [8-13].

In this study, we propose an automated and boosted classifier for the non-invasive and accurate detection of DS from facial photography based on ensemble learning. After locating the facial landmarks based on a constrained local model, geometric and texture features are extracted and selected following the method in [11]. Then SVM, RF and LDA are performed separately to discriminate between DS and non-syndrome groups. Finally, we boost the performance of these individual classifiers via ensemble learning, which integrates the shared and complementary information of individual classifiers. Ensemble learning is a machine learning paradigm that combines multiple hypotheses and it has widely applied to classification [14], segmentation [15] and registration [16]. In this study, multiple decision fusion methods are compared and evaluated in terms of accuracy, precision and recall.

II. Methods

The DS dataset consists of 130 frontal facial images (one image per subject) including 50 DS patients and 80

Qian Zhao is with the Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, Washington DC 20010 USA (phone: 202-476-1285; e-mail: qzhao@ cnmc.org).

Naoufel Werghi is with the Electrical and Computer Engineering Department, Khalifa University of Science, Technology & Research, Sharja, UAE.

Kazunori Okada is with the Computer Science Department, San Francisco University, San Francisco, CA, USA.

Kenneth Rosenbaum and Marshall Summar are with the Division of Genetics and Metabolism, Children's National Medical Center, Washington DC 20010 USA.

Marius George Linguraru is with Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Medical Center, and with the Departments of Radiology and Pediatrics at the School of Medicine and Health Sciences, George Washington University, Washington DC, USA.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

non-syndrome subjects. Photographic data acquisition and processing with a variety of cameras and under variable illumination, expression and poses was approved by the Institutional Review Board (IRB) and Children's National Medical Center. The subjects are from multiple ethnicities including 98 Caucasian, 20 African American and 12 Asian and both genders (86 males and 44 females). The age of patients varies from 0 to 36 month.

Forty-four anatomical landmarks covering the eyes (10), nose (14), mouth (9) and along the contour of the face (11) are first located automatically by using a constrained local model (CLM) with independent component analysis (ICA) described in [11].

A. Feature Extraction and Selection

To characterize the characteristic facial geometry and texture of DS patients, we extract and combine geometric and local texture features based on the facial landmarks after aligning the patient image with a reference image. The alignment is performed using Procrustes analysis to remove the translation, in-plane rotation and scaling [17].

The geometric features are defined via interrelationships among the facial landmarks to incorporate the clinical criteria of DS diagnosis. We extract 27 geometric features including 13 corner angles and 4 horizontal and 10 vertical Euclidean distances, shown in Fig. 1 (a). All angles are acute (<90°) according to their definitions, therefore monotonic in our application. The horizontal and vertical Euclidean distances are normalized by their baselines. The horizontal baseline is the distance between the left corner of left eye and the right corner of right eye (the width of the face), and the vertical baseline is the vertical distance between the eyes and the lower lip (the height of the face). The normalized geometric features are invariant to scale, translation and rotation.

The local texture features are extracted from the square blocks centered at each inner facial landmark (Fig. 1 (b)) based on LBP histogram statistics. For each square block, a uniform LBP histogram [18] is computed in which the number of neighboring sample points is not limited

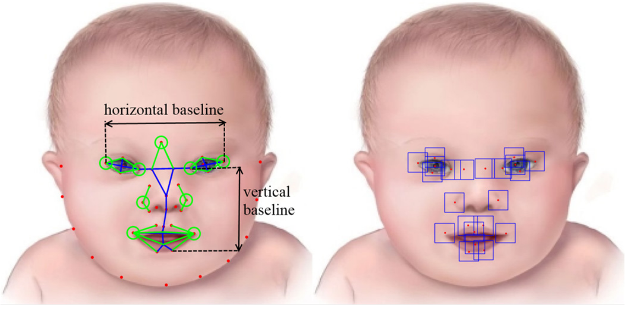
$$LBP_{P,R}^{riu2}\left(x,y\right)=\begin{cases}\sum_{p=0}^{P-1}s\left(g_p-g_c\right), \text{ if } U\left(LBP_{P,R}\right)\le 2\\ P+1, \text{ otherwise}\end{cases}, \quad (1)$$

$$U\left(LBP_{P,R}\right)=\left|s\left(g_{P-1}-g_c\right)-s\left(g_0-g_c\right)\right| + \sum_{p=1}^{P}\left|s\left(g_p-g_c\right)-s\left(g_{p-1}-g_c\right)\right|, \quad (2)$$

where $s(\cdot)$ is a sign function, g_p corresponds to the grey values of P equally spaced pixels on a circle of radius R, and g_c is the grey value of the central pixel. In this study, we set $P=8$ and $R=1$ based on experimental results from testing performance with varying settings of P and R.

The LBP histogram represents the distribution of the local micro-patterns, such as lines, ridges, spots and flat regions, which is suitable to delineate facial texture patterns (e.g. epicanthic folds and flattened philtrum for DS). To obtain a compact expression of texture information, six first-order statistical measurements of the LBP histogram are computed including the mean, variance, skewness, kurtosis, energy and entropy. Finally, the feature vectors in all square blocks are concatenated to form the LBP-based local texture features for the facial image. Therefore, the 132 LBP-based local texture features also contain the spatial information of the texture.

To obtain a more comprehensive representation of facial features, geometric and texture features are concatenated to 159 combined features. Feature selection is performed using the method in [19], which is based on manifold learning and $\mathcal{L}1$ regularized models for subset selection. The selected features preserve the multi-cluster structure of the data. Specifically, the feature selection method measures the correlations among different features by using spectral analysis techniques [20]. The corresponding optimization problem only involves a sparse eigen-problem and a $\mathcal{L}1$ -regularized least square problem, thus can be solved easily. The optimal dimension for feature space is found based on maximizing the area under the receiver operating characteristic (AUROC) curves by empirical exhaustive search.



(a) (b)

Figure 1. Feature extraction: (a) The graphic definition of geometric features; the blue lines are normalized by the vertical or horizontal baseline and the green circles illustrate the location of corners; (b) shows the 22 square blocks of inner face landmarks for local texture feature extraction.

B. DS Detection via Individual Classifiers

After feature extraction and selection, we identify DS from a non-syndrome group by using four individual classifiers separately, including SVM with radial basis function kernel (SVM-RBF), SVM with linear kernel (SVM-linear), RF and LDA.

SVM is a robust and powerful classifier [21], which uses a kernel function to map the data into a high-dimensional feature space. The kernel functions, such as the linear first-order polynomial and the non-linear RBF, have an impact on the classifier performance depending on the distribution of the analyzed data. The random forest is a state-of-the-art robust ensemble non-linear classifier that consists of multitude decision trees [22]. Each tree is trained independently and the forest output is the mode of the classes output by individual trees. LDA is another commonly used classifier, this time a linear classifier, that maximizes the ratio of between-class variance to within-class variance in the data, thereby maximizing separability [23]. LDA is suitable for cases where the within-class frequencies are unequal. The parameters for the SVM (C, σ) are found optimally by grid search [24], and the number of trees in RF is set experimentally to 150.

The performance of these individual classifiers dependents on both the distribution of the data and the classifier properties. To make a more accurate and reliable diagnosis decision for DS, we adopt an ensemble learning system to boost the individual decisions of multiple classifiers via information fusion in the boosted decision level.

C. DS Detection via Ensemble Learning

The outputs of the individual classifiers are regarded as conditionally-independent belief vectors that can be either identity vectors or probabilities. Given multiple classifiers $S_m, m = 1,2,\ldots,M$, we suppose each of them classifies an unknown object into one of K classes c_k $k = 1,2,\ldots,K$ ($K = 2$ here, DS or non-syndrome) with a probability $p(C = c_k \mid S_m, I)$, where C is the classifier output and I the image information. In a general Bayesian framework, the *a posteriori* class probability of the ensemble classifier can be written as

$$p(\mathrm{C} = \mathrm{c}_k \mid \mathrm{S}_1,\mathrm{S}_2,\ldots,\mathrm{S}_M,\mathrm{I}) = \prod_{m=1}^{M} p(\mathrm{C} = \mathrm{c}_k \mid \mathrm{S}_m,\mathrm{I}), \tag{3}$$

which can be approximated by using the following rules [14]

$$p(\mathrm{C} = \mathrm{c}_k \mid \mathrm{S}_1,\mathrm{S}_2,\ldots,S_M,\mathrm{I}) \sim \begin{cases} \operatorname{mean}_m\left(p(\mathrm{C} = \mathrm{c}_k \mid \mathrm{S}_m,\mathrm{I})\right) \\ \operatorname{median}_m\left(p(\mathrm{C} = \mathrm{c}_k \mid \mathrm{S}_m,\mathrm{I})\right), \\ \max_m\left(p(\mathrm{C} = \mathrm{c}_k \mid \mathrm{S}_m,\mathrm{I})\right) \end{cases} \tag{4}$$

which are less sensitive to noise and outliers.

For the binary identity belief vectors, the ensemble learning is achieved by majority vote (MV) or pairwise fusion matrix (PFM) [14]. Majority vote outputs the mode of classes output by individual classifiers. In the pairwise fusion matrix we divide the M classifiers into an ensemble $\boldsymbol{P} = (\mathrm{P}_1,\mathrm{P}_2,\ldots,\mathrm{P}_N)^T$ of all possible pairs of classifiers, where $P_1 = (\mathrm{S}_1,\mathrm{S}_2)$, $P_2 = (\mathrm{S}_1,\mathrm{S}_3),\ldots,$ $P_N = (\mathrm{S}_{M-1},\mathrm{S}_M)$ and $N = M\left(M-1\right)/2$. For each pair of classifiers, we record the occurrence of the K^2 class label combinations $\omega_l, l \in \{1,2,\ldots,\mathrm{K}^2\}$ for the training data. Suppose the unknown object belongs to a certain combination ω_l classified by the paired classifiers of P_i , the output of the pair classifier P_i, c^*, is computed as

$$c^* = \arg\max_k\left(n\left(k,l\right)\right), \tag{5}$$

where $n\left(k,l\right)$ is the number of training samples that belong to class $C = c_k$ and that have the l^{th} combination of class labels ω_l. Finally the set of N classifications are fused together by using majority vote.

III. Experiments

Leave-one-subject-out cross validation was performed and evaluated in terms of accuracy, precision and recall. Accuracy is the overall correctness of the method that is the proportion of true results (both true positives and true negatives) in the population. Precision measures the proportion of the true positives against all the positive results, while recall measures the proportion of true positives which are correctly identified as such.

A. Feature Selection

For combined features, there were 24, 97, 78, and 32 features selected for SVM-RBF, linear SVM, RF and LDA, respectively. We show the top ten selected geometric features for linear SVM as examples in Table I; linear SVM methods (both linear and RBF bases) had the best performance of the individual classifiers (Table II). The top ranked geometric features included the orientation of the eyes, palpebral fissures and length of nose, which are consistent with the clinical findings of DS (upward slating eyes, narrow palpebral fissure and small nose). The top ranked texture features mainly lay in the eyes and mouth corners, which had more discriminative powers and mirrored the clinical criteria for DS diagnosis. In particular, the texture around inner eye corners describes the epicanthic folds that are clinically relevant for the diagnosis of Down syndrome.

TABLE I. THE SELECTED TOP RANKED GEOMETRIC FEATURES AND THEIR CLINICAL RELEVANCEA BY USING [19]

Feature Ranking	Geometric Features
1	Orientation of right eye
2	Length of upper part of nose
3	Length of Right palpebral fissure
4	Length of lower nose
5	Thickness of upper lip
6	Orientation of left eye
7	Length of philtrum
8	Distance between inner corners of eyes
9	Angle of outer left corner of mouth
10	Angle of left corner of right eye

B. Down Syndrome Detection

For DS detection, we compared the performance of individual classifiers and different ensemble learning method including majority vote, PFM, mean rule, median rule and maximum rule. The experimental results are shown in Table II. The best performance of individual classifiers was obtained by SVM with either RBF or linear kernel, as shown by accuracy and AUROC. The ensemble classifiers PFM, Mean and Median outperformed the best individual classifier by decreasing the misclassification rate by 19.5%. This is equivalent to 0.08 increase in accuracy in the tight space allowed for improvement. The largest AUROC 0.996 was achieved by the ensemble classifier with median rule, shown in Fig.2. As MV and PFM generate binary identity output instead of probabilities, their AUROC values are not computable.

TABLE II. THE PERFORMANCE OF DS DETECTION USING DIFFERENT INDIVIDUAL AND ENSEMBLE CLASSIFIERS.

	Accuracy	Precision	Recall	AUROC
SVM-RBF	0.959	0.932	0.953	0.994
linear SVM	0.959	0.932	0.953	0.994
RF	0.909	0.900	0.837	0.966

LDA	0.959	0.896	1.000	0.990
MV	0.959	0.954	0.932	-
PFM	**0.967**	0.977	0.933	-
Mean	**0.967**	0.977	0.933	0.995
Median	**0.967**	0.977	0.933	**0.996**
Max	0.926	1.000	0.827	0.990
Prod (3)	0.956	0.954	0.932	-

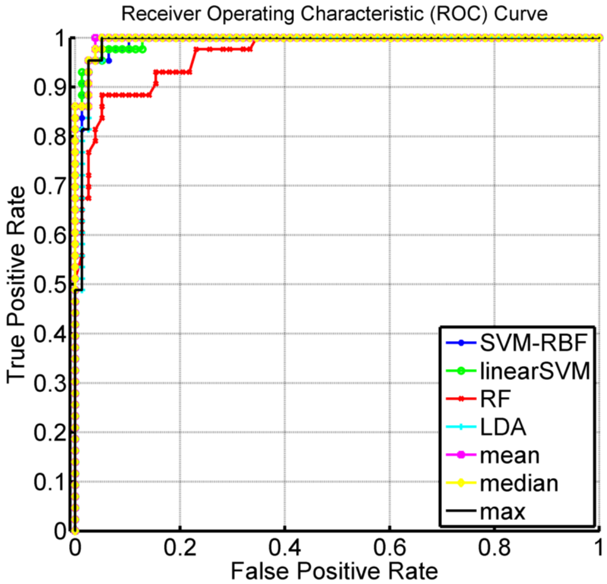


Fig. 2 Comparative ROC curves for Down syndrome detection.

IV. Conclusion

We proposed an automated, non-invasive and accurate system for Down syndrome detection based on facial analysis and ensemble learning. Geometric and texture features were extracted based on the automatically located facial landmarks, followed by feature combination and selection. Then multiple linear and non-linear classifiers were trained to identify Down syndrome cases from a non-syndrome group. The final decision was achieved by fusing these individual classifiers via ensemble learning in the boosted decision level. The ensemble classifier outperformed the individual classifiers by decreasing the misclassification rate by 19.5%. The best performance was achieved by the median ensemble rule with 0.967 accuracy and 0.996 AUROC. These promising results encourage us to further investigate the detection of other types of genetic syndromes associated with facial dysmorphology. Future work will also include the investigation of more advanced ensemble learning methods for decision fusion.

Acknowledgment

This project was supported by a philanthropic gift from the Government of Abu Dhabi to Children's National Medical Center and a research grant KUIRF-Lev1-210055 from Khalifa University.

References

[1] G. de Graaf, *et al.*, "Changes in yearly birth prevalence rates of children with Down syndrome in the period 1986–2007 in the Netherlands," *Journal of Intellectual Disability Research,* vol. 55, pp. 462-473, 2011.

[2] S. K. Murthy, *et al.*, "Incidence of Down Syndrome in Dubai, UAE," *Medical Principles and Practice,* vol. 16, pp. 25-28, 2007.

[3] S. E. Parker, *et al.*, "Updated national birth prevalence estimates for selected birth defects in the United States, 2004–2006," *Birth Defects Research Part A: Clinical and Molecular Teratology,* vol. 88, pp. 1008-1016, 2010.

[4] P. A. Benn, "Advances in prenatal screening for Down syndrome: I. general principles and second trimester testing," *Clinica chimica acta; international journal of clinical chemistry,* vol. 323, pp. 1-16, 2002.

[5] R. W. K. Chiu, *et al.*, "Non-invasive prenatal assessment of trisomy 21 by multiplexed maternal plasma DNA sequencing: large scale validity study," *BMJ,* vol. 342, 2011-01-11 00:00:00 2011.

[6] F. K. Wiseman, *et al.*, "Down syndrome—recent progress and future prospects," *Human Molecular Genetics,* vol. 18, pp. R75-R83, 2009.

[7] B. G. Skotko, "First- and Second-Trimester Evaluation of Risk for Down Syndrome," *Obstetrics & Gynecology,* vol. 110, 2007.

[8] H. S. Loos, *et al.*, "Computer-based recognition of dysmorphic faces," *European Journal of Human Genetics,* vol. 11, pp. 555-560, 2003.

[9] Ş. Saraydemir, *et al.*, "Down Syndrome Diagnosis Based on Gabor Wavelet Transform," *Journal of Medical Systems,* vol. 36, pp. 3205-3213, 2012.

[10] K. Burçin and N. V. Vasif, "Down syndrome recognition using local binary patterns and statistical evaluation of the system," *Expert Systems with Applications,* vol. 38, pp. 8690-8695, 2011.

[11] Q. Zhao, *et al.*, "Hierarchical Constrained Local Model Using ICA and Its Application to Down Syndrome Detection," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. vol. 8150, 2013, pp. 222-229.

[12] Q. Zhao, *et al.*, "Automated Down Syndrome Detection using Facial Photographs," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Osaka, Japan, 2013.

[13] Q. Zhao, *et al.*, "Down syndrome detection from facial photographs using machine learning techniques," in *SPIE Medical Imaging*, 2013, pp. 867003-867003.

[14] H. B. Mitchell, "Ensemble Learning," in *Data Fusion: Concepts and Ideas*: Springer Berlin Heidelberg, 2012, pp. 295-321.

[15] J. Huo, *et al.*, "Ensemble segmentation for GBM brain tumors on MR images using confidence-based averaging," *Medical Physics,* vol. 40, pp. -, 2013.

[16] I. J. A. Simpson, *et al.*, "Ensemble Learning Incorporating Uncertain Registration," *IEEE Transactions on Medical Imaging,* vol. 32, pp. 748-756, 2013.

[17] J. C. Gower, "Generalized procrustes analysis," *Psychometrika,* vol. 40, pp. 33-51, 1975.

[18] T. Ojala, *et al.*, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 971-987, 2002.

[19] D. Cai, *et al.*, "Unsupervised feature selection for multi-cluster data," presented at the Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, DC, USA, 2010.

[20] A. Y. Ng, *et al.*, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*, 2001, pp. 849-856.

[21] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning,* vol. 20, pp. 273-297, 1995.

[22] L. Breiman, "Random Forests," *Machine Learning,* vol. 45, pp. 5-32, 2001.

[23] S. Mika, *et al.*, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX, 1999*, 1999, pp. 41-48.

[24] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *J. Mach. Learn. Res.,* vol. 13, pp. 281-305, 2012.