DSX 18 (2024) 103003

Contents lists available at ScienceDirect

Diabetes & Metabolic Syndrome: Clinical Research & Reviews

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

Artificial intelligence facial recognition system for diagnosis of endocrine and metabolic syndromes based on a facial image database

Danning Wu [a,b,1], Jiaqi Qiang [a,1], Weixin Hong [c,d,1], Hanze Du [a], Hongbo Yang [a], Huijuan Zhu [a,e], Hui Pan [a,e,2,***], Zhen Shen [c,2,**], Shi Chen [a,2,*]

[a] Key Laboratory of Endocrinology of National Health Commission, Department of Endocrinology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, China
[b] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing, 100142, China
[c] State Key Laboratory of Multimodal Artificial Intelligence Systems, Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China
[d] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, 100049, China
[e] State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 100730, China

ARTICLE INFO

Keywords:
Endocrine and metabolic syndrome
Facial recognition
Diagnosis
Artificial intelligence
Facial image database

ABSTRACT

Aim: To build a facial image database and to explore the diagnostic efficacy and influencing factors of the artificial intelligence-based facial recognition (AI-FR) system for multiple endocrine and metabolic syndromes.
Methods: Individuals with multiple endocrine and metabolic syndromes and healthy controls were included from public literature and databases. In this facial image database, facial images and clinical data were collected for each participant and dFRI (disease facial recognition intensity) was calculated to quantify facial complexity of each syndrome. AI-FR diagnosis models were trained for each disease using three algorithms: support vector machine (SVM), principal component analysis k-nearest neighbor (PCA-KNN), and adaptive boosting (AdaBoost). Diagnostic performance was evaluated. Optimal efficacy was achieved as the best index among the three models. Effect factors of AI-FR diagnosis were explored with regression analysis.
Results: 462 cases of 10 endocrine and metabolic syndromes and 2310 controls were included into the facial image database. The AI-FR diagnostic models showed diagnostic accuracies of 0.827–0.920 with SVM, 0.766–0.890 with PCA-KNN, and 0.818–0.935 with AdaBoost. Higher dFRI was associated with higher optimal area under the curve (AUC) ($P = 0.035$). No significant correlation was observed between the sample size of the training set and diagnostic performance.
Conclusions: A multi-ethnic, multi-regional, and multi-disease facial database for 10 endocrine and metabolic syndromes was built. AI-FR models displayed ideal diagnostic performance. dFRI proved associated with the diagnostic performance, suggesting inherent facial features might contribute to the performance of AI-FR models.

* Corresponding author. Key Laboratory of Endocrinology of National Health Commission, Department of Endocrinology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No.1 Shuaifuyuan, Wangfujing, Dongcheng District, Beijing, 100730, China.
** Corresponding author. State Key Laboratory of Multimodal Artificial Intelligence Systems, Beijing Engineering Research Center of Intelligent Systems and Technology, Institute of Automation, Chinese Academy of Sciences, No.95 Zhongguancun E. Road, Beijing 100190, China.
*** Corresponding author. Key Laboratory of Endocrinology of National Health Commission, Department of Endocrinology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No.1 Shuaifuyuan, Wangfujing, Dongcheng District, Beijing, 100730, China.
E-mail addresses: panhui20111111@163.com (H. Pan), zhen.shen@ia.ac.cn (Z. Shen), cs0083@126.com (S. Chen).
[1] Danning Wu, Jiaqi Qiang, and Weixin Hong contributed equally to this work.
[2] Shi Chen, Zhen Shen, and Hui Pan shared the co-corresponding authorship.

https://doi.org/10.1016/j.dsx.2024.103003
Received 16 August 2023; Received in revised form 25 March 2024; Accepted 2 April 2024
Available online 3 April 2024

1. Introduction

Many endocrine and metabolic syndromes are associated with specific facial features, including wide ocular hypertelorism, triangular face, saddle nose, low-set ears, microtia, dental dysplasia, cleft lip, cleft palate, exophthalmos, and dermatologic manifestations [1,2]. The diagnosis of endocrine and metabolic syndromes requires a set of biochemical and even genetic tests, leading to a complex and time-consuming diagnosis process [3,4]. Additionally, the spectrum of endocrine and metabolic syndromes is diverse thus easily confused, delayed and error diagnoses are common. Endocrinologists, particularly junior or primary care physicians, face challenges in providing early identification and diagnosis to these patients, not to mention early intervention to improve prognosis. Therefore, there is an urgent need to investigate auxiliary diagnostic methods to help these doctors improve their diagnostic efficacy.

Artificial intelligence (AI) has been widely applied in the analysis and identification of medical images, such as lung nodules [5], colon polyps [6], breast nodules [7], and ocular fundus [8]. In recent years, artificial intelligence-based facial recognition (AI-FR) technology also displayed ideal performance in the automatic detection of endocrine and metabolic syndromes and genetic disorders with distinct facial features [9], such as Cushing's syndrome [10], acromegaly [11], Turner syndrome [12,13], and De Lange's syndrome [14]. Based on our previous meta-analysis on diagnostic accuracy of AI-FR system for endocrine and metabolic syndromes in 20 studies, we proposed facial recognition intensity (FRI), an index to describe facial complexity of a target disease. We also validated that FRI has an impact on the diagnostic performance of AI-FR. The adjustment of parameters in artificial intelligence technology might help improve the diagnostic accuracy for diseases with lower FRI [15].

Though AI-FR-assisted systems have shown potentials in improving the diagnostic process of endocrine and metabolic syndromes, there is still a lack in exploring effect factors of AI-FR's diagnostic performance with a real-world database [16,17]. Moreover, current diagnostic systems mostly focus on a single disease in relatively limited population, to build a medical facial image database comprising of various endocrine and metabolic syndromes is in need. Therefore, the purpose of this study is to build such a facial image repository covering multiple diseases and participants from diverse races and regions. We also aim to establish an auxiliary AI-FR diagnostic system based on this database, to compare the performance of different machine learning algorithms, and to explore possible influencing factors of diagnostic efficacy.

2. Methods

2.1. Study design

This study first included patients of endocrine and metabolic syndromes and healthy controls from public sources. A database was built to collect facial images and clinical data of all included participants. Based on the dataset in the database, AI-FR diagnosis models for endocrine and metabolic syndromes were trained using three classical AI algorithms. Then diagnostic performance of the AI-FR models was evaluated and possible influencing factors were explored through regression analysis.

2.2. Study population

Participants of the disease group and the control group from diverse regions, races, and ages were included from data sources with public use permissions. The disease group was collected from peer-reviewed literature, published books, and two medical image databases [*Face2Gene Library* (www.face2gene.com/lmd-library-london-medical-database-dysmorphology, accessed on February 25th, 2023) and *Atlas of Human Malformations in Diverse Populations of the National Human Genome Research Institute* [18] (research.nhgri.nih.gov/atlas, accessed on February 27th 2023)]. The control group was collected from the large-scale face dataset UTKFace (susanqq.github.io/UTKFace, accessed on March 1st 2023).

Participants were included according to predefined criteria. For the disease group, patients were diagnosed according to the gold standard in the guideline or consensus of the endocrine and metabolic syndrome. For the control group, participants were from healthy population and have not been diagnosed with any endocrine or metabolic syndrome.

The exclusion criteria were (1) Unreliable or unpermitted data sources; (2) Lack of diagnosis confirmed by gold standard methods; (3) History of major facial surgery, trauma, facial injection (e.g., hyaluronic acid), or orthodontic treatment; (4) Combination of diseases resulted in a significant change in facial appearance (e.g., scleroderma, systemic lupus erythematosus, dermatomyositis).

2.3. Data collection

Facial images, demographic information and disease diagnosis were collected for each participant from the data source. Identity information was removed to ensure privacy protection before further processing. Then data of each participant underwent a cleaning process to eliminate inconsistencies or errors. All participant data was formatted according to a standard criterion to ensure uniformity and ease of analysis.

2.4. Facial image database building

The facial image database was built to manage the data collected in the above steps. The database will support the storage, organization, retrieval and update of the multi-modal data, including full resolution color images, numerical information, text information, etc. The facial image database was organized and stored with a website only available in the intranet cloud server.

For each endocrine and metabolic syndrome involved in the database, dFRI (disease FRI) was calculated to represent the facial feature complexity of the target disease. As previously described, dFRI was defined as the product of the number of independent facial phenotypes (N_f) of the disease and the maximum penetrance (P_{max}) of the facial features [15]. After calculation, the dFRI is stored in the database as one of the entries of each patient.

2.5. Model training of AI-FR for diagnosis of endocrine and metabolic syndromes

The disease group and the control group were randomly selected at the ratio of 5:1 matched by sex, age, and race. Facial images of both groups were first extracted and preprocessed. The face area was first recognized and intercepted from the entire image. To minimize the impact of orientation and lighting, the image was then rectified, normalized, and converted into grayscale. Afterwards, facial images from both the disease group and the control group were merged and randomly split into a training set and a test set at a ratio of 3:7. AI-FR automatic classification models were trained for each disease. To leverage the strengths of various methods, three classical machine learning algorithms were applied: support vector machine (SVM), k-nearest neighbor (KNN), and adaptive boosting (AdaBoost). SVM models maximize the interval in the feature space to achieve binary classification. KNN models measure the distance between different feature points and use the principal component analysis (PCA) method to reduce the numerous features before classification. AdaBoost models combine multiple different classifiers trained by one dataset and formed a weighted classification model with optimal performance. Following the initial training, parameters were adjusted to achieve optimal results. The test set was then fed into each model to determine the accuracy of automatic classification. This process was iterated five times randomly to enhance robustness.

Model training was performed with Python packages dlib (version 19.22.1), scipy (version 1.5.2), sklearn (version 0.23.2), and matplotlib (version 3.3.2).

2.6. Evaluation of the diagnostic performance and statistical analysis

The diagnostic performance of each AI-FR models in each disease were evaluated with four indices: accuracy, sensitivity, specificity, and area under receiver operating characteristic (AUC). Then possible influencing factors of diagnostic performance were explored. Linear regression analysis was performed separately for each endocrine and metabolic syndrome to investigate the relationship between diagnostic performance indices and two factors: sample size of the training set and dFRI. The statistical analysis was performed with SPSS (version 26.0). P-value below 0.05 was considered statistically significant.

3. Results

3.1. Composition of the facial image database

The facial image database organized the following entries for each included patient or healthy individual: the facial image, disease diagnosis or health status, and dFRI of the disease. The study population included 462 patients with 10 endocrine and metabolic syndromes into the disease group and 2310 individuals from healthy population into the control group. In the disease group, there were 68 cases of Down syndrome, 58 cases of Cornelia de Lange syndrome, 193 cases of Noonan syndrome, 16 cases of Turner syndrome, 35 cases of Prader-Willi syndrome, 22 cases of Angelman syndrome, 17 cases of fragile-X syndrome, 20 cases of Aymé-Gripp syndrome, 13 cases of achondroplasia, and 20 cases of Laron syndrome.

Of each included endocrine syndrome, N_f, P_{max}, and dFRI were calculated as shown in Table 1. Both Down syndrome and fragile X syndrome have 9 facial features and a maximum penetrance of 100%, and the calculated dFRI is 9 [19,20]. Noonan syndrome has 8 facial features and a maximum penetrance of 100%, and the calculated dFRI is 8 [21]. Cornelia de Lange syndrome has 9 facial features and a maximum penetrance of 82.7%, and the calculated dFRI is 7.443 [22]. Aymé-Gripp syndrome [23], Angelman syndrome, and Prader-Willi syndrome [24] have 6 facial features and a maximum penetrance of 100%, and the calculated dFRI of them were all 6. Laron syndrome has 5 facial phenotypes with a maximum penetrance of 100%, and the dFRI is 5 [25]. There are 4 facial phenotypes of achondroplasia and the maximum penetrance is 100%, so the dFRI is 4 [26]. Turner syndrome has 6 facial phenotypes and the maximum penetrance is 56%, so dFRI is 3.36 [13].

Table 1
DFRI of endocrine and metabolic syndromes included in the facial image database.

Disease	Facial features	N_f	P_{max}	dFRI
Down syndrome	Short face, slanting eyes, epicanthic fold, brushfield spots, low ear, small ear, low bridge of nose, small mouth, tongue extension	9	100%	9
Fragile-X syndrome	Narrow face, long face, prominent forehead, high palatal arch, prominent chin, large jaw, large ears, large mouth, thick lips	9	100%	9
Noonan syndrome	Large forehead, high palatal arch, wide eye distance, ptosis, short nose, wide nasal base, low ear, full upper lip	8	100%	8
Cornelia de Lange syndrome	Short face, small jaw, arched eyebrows, connected eyebrows, short nose, forward nostrils, long philtrum, thin upper lip, upturned corners of mouth	9	82.7%	7.443
Aymé-Gripp syndrome	Short face, ptosis, short nose tip, medium length, small mouth, low ear position	6	100%	6
Angelman syndrome	Narrow forehead, wide jaw, almond eyes, narrow nasal bridge, thin upper lip, tongue extension	6	100%	6
Prader-Willi syndrome	Narrow face, small jaw, almond eyes, narrow nose bridge, thin upper lip, drooping corners of the mouth	6	100%	6
Laron syndrome	Protruding forehead, low bridge of nose, short face, small jaw, blue sclera	5	100%	5
Achondroplasia	Large head, protruding forehead, low bridge of nose, broad jaw	4	100%	4
Turner syndrome	Small jaw, epicanthic fold, ptosis, wide eye distance, low ear, multiple nevi	6	56%	3.36

Abbreviations: dFRI, disease facial recognition intensity; N_f, number of independent facial phenotypes; P_{max}, the maximum penetrance of the facial features.

3.2. Diagnostic performance of AI-FR models for endocrine and metabolic syndromes

For each endocrine syndrome, the diagnostic performance (accuracy, sensitivity, and specificity) of each AI-FR model (SVM, PCA-KNN, and AdaBoost) was shown in Table 2. The AUC and the diagnostic receiver operating characteristic (ROC) curves of SVM and AdaBoost models were shown in Table 2, Fig. 1, and Fig. 2. The AUC and ROC curve of PCA-KNN algorithm was not shown because this algorithm does not produce probability scores for classifications [27].

In Down syndrome, the best accuracy is 0.935 obtained by the AdaBoost model, the best sensitivity is 0.698 obtained by the AdaBoost model, the best specificity is 1.000 obtained by the SVM model, and the best AUC value is 0.984 obtained by the SVM model. The optimal accuracy and sensitivity for Cornelia de Lange syndrome are 0.930 and 0.739 in the AdaBoost model, specificity of 1.000 in the SVM model, and AUC of 0.951 in the AdaBoost model. In Noonan syndrome, the best accuracy and sensitivity were 0.923 and 0.681 in the AdaBoost model, while the best specificity and AUC values were 0.995 and 0.983 in the SVM model. The best diagnostic accuracy of Turner syndrome is 0.843 obtained by the SVM model, the best sensitivity of 0.367 in the PCA-KNN model, the best specificity of 1.000 in the SVM model, and the best AUC value of 0.797 in the SVM model. The highest accuracy and sensitivity of Prader-Willi Syndrome were 0.923 and 0.626 in the AdaBoost model, while the highest specificity and AUC were 1.000 and 0.957 in the SVM model. The best accuracy of Angelman syndrome was achieved in the AdaBoost model at 0.916, the optimal sensitivity was 0.597 in the PCA-KNN model, and the optimal specificity and AUC were achieved in the SVM model at 1.000 and 0.965. The optimal accuracy for Aymé-Gripp Syndrome is 0.903 in the AdaBoost model, 0.710 in the PCA-KNN model, and 1.000 and 0.985 for specificity and AUC in the SVM model. The optimal accuracy for Fragile-X syndrome is 0.880 in the AdaBoost model, 0.528 in the PCA-KNN model, and 1.000 and 0.981 in the SVM model for specificity and AUC. In Laron syndrome, the best accuracy and sensitivity were 0.921 and 0.725 in the AdaBoost model, while the best specificity and AUC values were 1.000 and 0.990 in the SVM model. Among achondroplasia, the SVM model showed the best accuracy, specificity, and AUC performance, with values of 0.827, 1.000, and 0.917, respectively. The optimal sensitivity was 0.599 of the PCA-KNN model.

3.3. Optimal diagnostic performance and influencing factors of AI-FR models

To obtain the best diagnostic performance in each endocrine and metabolic syndrome, the highest accuracy, sensitivity, specificity, and AUC were selected across the three models. Possible influencing factors

Table 2
Diagnostic efficacy of different AI-FR models in the 10 endocrine and metabolic syndromes.

Disease	SVM				PCA-KNN			AdaBoost			
	Acc.	Sen.	Spe.	AUC	Acc.	Sen.	Spe.	Acc.	Sen.	Spe.	AUC
Down syndrome	0.880	0.194	1.000	0.984	0.839	0.576	0.904	0.935	0.698	0.984	0.974
Cornelia de Lange syndrome	0.911	0.520	1.000	0.937	0.861	0.497	0.923	0.930	0.739	0.973	0.951
Noonan syndrome	0.920	0.566	0.995	0.983	0.879	0.636	0.930	0.923	0.681	0.968	0.948
Turner syndrome	0.843	0.050	1.000	0.797	0.793	0.367	0.873	0.836	0.283	0.957	0.705
Prader-Willi syndrome	0.855	0.045	1.000	0.947	0.890	0.531	0.954	0.923	0.626	0.981	0.957
Angelman syndrome	0.863	0.000	1.000	0.965	0.884	0.597	0.962	0.916	0.576	0.982	0.914
Fragile-X syndrome	0.846	0.000	1.000	0.981	0.766	0.528	0.821	0.880	0.348	0.980	0.812
Aymé-Gripp syndrome	0.836	0.000	1.000	0.985	0.879	0.710	0.910	0.903	0.706	0.971	0.972
Achondroplasia	0.827	0.000	1.000	0.917	0.791	0.599	0.843	0.818	0.423	0.910	0.855
Laron syndrome	0.848	0.240	1.000	0.990	0.855	0.664	0.919	0.921	0.725	0.969	0.972

Abbreviations: SVM, support vector machine; PCA-KNN, principal component analysis-k-nearest neighbor; AdaBoost, adaptive boosting; Acc., accuracy; Sen., sensitivity; Spe., specificity; AUC, area under receiver operating characteristic.

(sample size of training set and dFRI) were also presented (Fig. 3, Table 3). The linear regression analysis showed that diseases with higher dFRI had higher optimal AUC ($P = 0.035$), while there was no significant association between dFRI and diagnostic accuracy ($P = 0.059$), sensitivity ($P = 0.0416$) or specificity ($P = 0.234$). There was no significant correlation between the sample size of the training set and the diagnostic performance.

In terms of AI model selection, AdaBoost models had the best diagnostic accuracy; PCA-KNN models and AdaBoost models had higher sensitivity; SVM models had the highest specificity and AUC. For optimal diagnostic performance obtained with the SVM model, the higher the dFRI of the disease, the better the diagnostic AUC ($P = 0.040$ in the linear regression). However, there was no significant association between dFRI and diagnostic performance in other AI models.

4. Discussion

This study has profound clinical relevance focusing on the automatic AI facial recognition system for diagnosis of endocrine and metabolic syndromes. Research in automated identification applied in clinical medicine is exploding. Artificial-intelligence-based facial recognition has been found to have superior performance in diagnosis of diseases. This interdisciplinary field is promising for the optimization of the screening and diagnosis process and assisting in clinical evaluation and decision-making [28]. In this study, we generated main findings as follows: (1) an extensive facial image database encompassing diverse ethnicities, regions, and individuals with endocrine and metabolic syndromes was developed for AI-FR diagnosis; (2) the AI-FR models designed for diagnosing endocrine and metabolic syndromes demonstrated exceptional efficacy within this facial image repository; (3) the concept of dFRI, which indicates the complexity of facial features, could potentially impact the diagnostic performance of AI-FR algorithms.

In recent years, the AI-FR technique has been widely used in the medical diagnosis of endocrine and metabolic syndromes with typical facial features, showing ideal diagnostic efficiency and practical prospects [16,17]. Nevertheless, most AI-FR methods focused on a single type of disease with participants of limited race and regions. Since there are multiple types of endocrine syndromes and the standardized diagnostic procedures are complex, it is necessary to develop diagnostic methods that could be easily and widely applied. This study has built the first medical facial image database of endocrine and metabolic syndromes with multi-diseases, multi-ethnicities, and multi-regions. By combining data collected from various public sources, 462 patients with 10 types of endocrine and metabolic syndromes were included. The database was centered around facial images and also included clinical information. It not only provided data for the need of AI-FR model training, but also supported the application of medical teaching and knowledge popularization of rare diseases.

Based on this facial image database, the AI-FR diagnostic system exhibited varying performance levels when trained with three different algorithms. Across the 10 endocrine and metabolic syndromes, the SVM showed diagnostic accuracy of 0.827–0.920; the PCA-KNN showed accuracy of 0.766–0.890; and the AdaBoost showed accuracy of 0.818–0.935. The AdaBoost model had the best diagnostic accuracy; the PCA-KNN and the AdaBoost models had better sensitivity; and the SVM model had the highest specificity and AUC. Various algorithms have different approaches and identify distinct patterns within the same dataset. These patterns have different impacts on the same disease. Consequently, we selected the most effective algorithm for each disease through comparative analysis, aiming to enhance the model's performance.

Various diseases exhibit distinct diagnostic performance even when employing the same model for training. For the optimal diagnostic index across the 10 diseases, the accuracy was 0.827–0.935; the sensitivity was 0.367–0.739; the specificity was 0.995–1.000; and the AUC was 0.797–0.990. Down syndrome had the highest accuracy of 0.935 by AdaBoost. Cornelia de Lange syndrome had the highest sensitivity of 0.739 by AdaBoost. All the 10 diseases had a specificity of over 99% by SVM. Laron syndrome had the highest AUC of 0.990 by SVM. These results suggested that to enhance performance in clinical practice, AI models should be tailored according to the specific disease type. Additionally, it prompted us to explore deeper into the factors determining variations in diagnostic performance across different diseases.

Previously, there is few study on influencing factors of diagnostic efficacy of AI-FR. As far as we know, our team is one of the first researchers to pay attention to this issue. In a meta-analysis of all researches on AI-FR models of disease diagnosis, we described the complexity of facial features of diseases using a novel index called dFRI and suggested it may influence the accuracy of AI-FR [15]. In this research, we validated that dFRI could potentially enhance AI diagnostic performance across various diseases and machine learning algorithms. Interestingly, we found no notable correlation between the sample size of the training set and accuracy performance. These results aligned with prior research on breast cancer and colon polyps, suggesting that distinct disease image features play a crucial role in the accurate performance of AI recognition [29,30].

Further, we have also introduced a hypothesis suggesting that the complexity of the research object determines the complexity of the AI processing process and potentially impacts the accuracy of AI diagnosis. Adjusting AI parameter could potentially boost diagnostic accuracy for subjects with lower complexity. This theory, referred to as the Object Complexity Theory (OCT), was initially derived from assumptions based on literature [15]. Notably, this study is consistent with the finding that FRI is a possible influencing factor for diagnostic performance of facial recognition, further validating OCT within practical settings [31]. These discoveries might serve as a theoretical foundation for future applications of AI in clinical and research domains, such as AI image recognition for pulmonary nodules [5] and optic neuropathy [8].

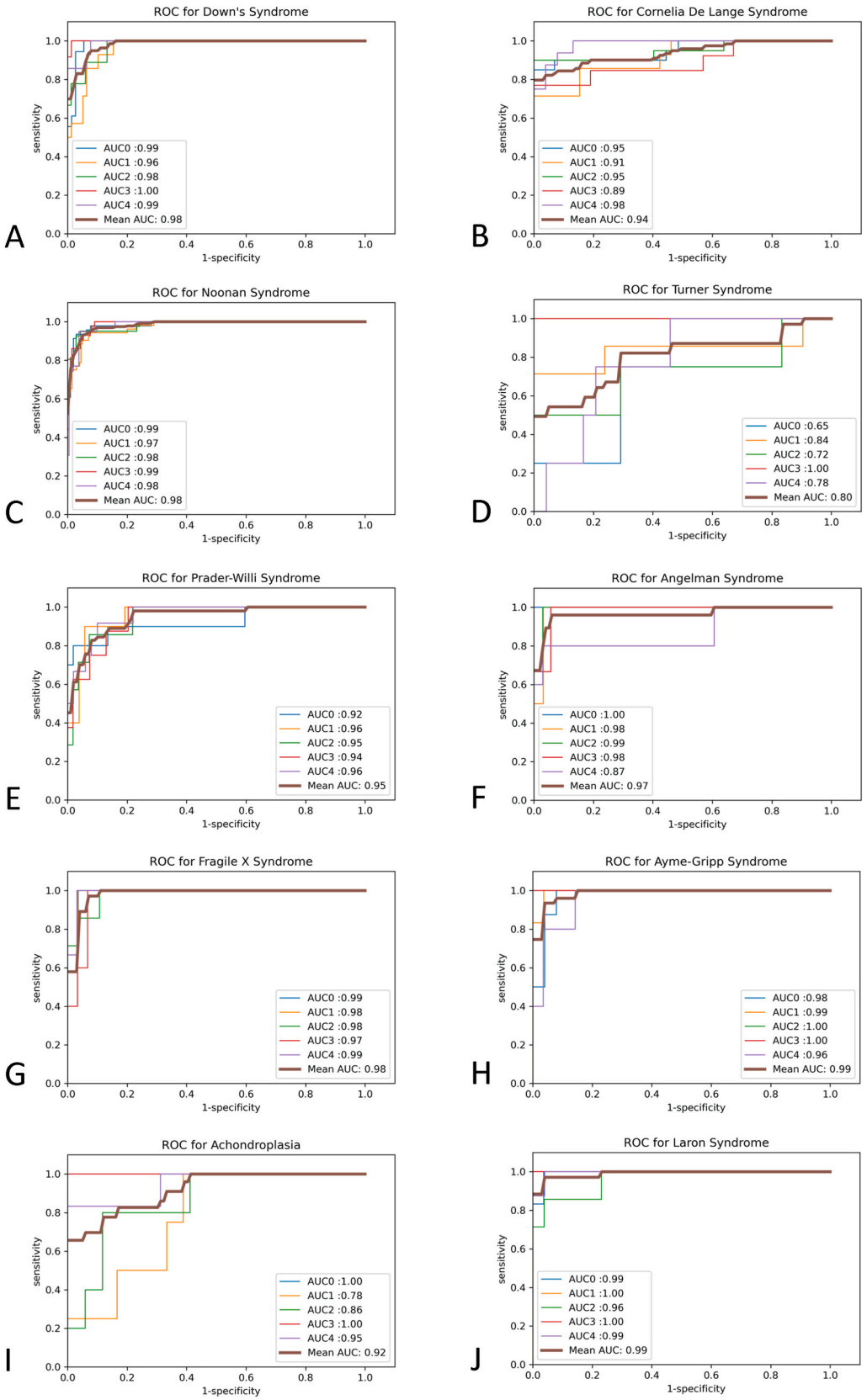

Fig. 1. Diagnostic receiver operating characteristic curve of the SVM model in the 10 endocrine and metabolic syndromes. A. Down syndrome. B. Cornelia de Lange syndrome. C. Noonan syndrome. D. Turner syndrome. E. Prader-Willi syndrome. F. Angelman syndrome. G. Fragile-X syndrome. H. Aymé-Gripp syndrome. I. Achondroplasia. J. Laron syndrome. Abbreviations: ROC, receiver operating characteristic; AUC, area under curve; SVM, support vector machine.

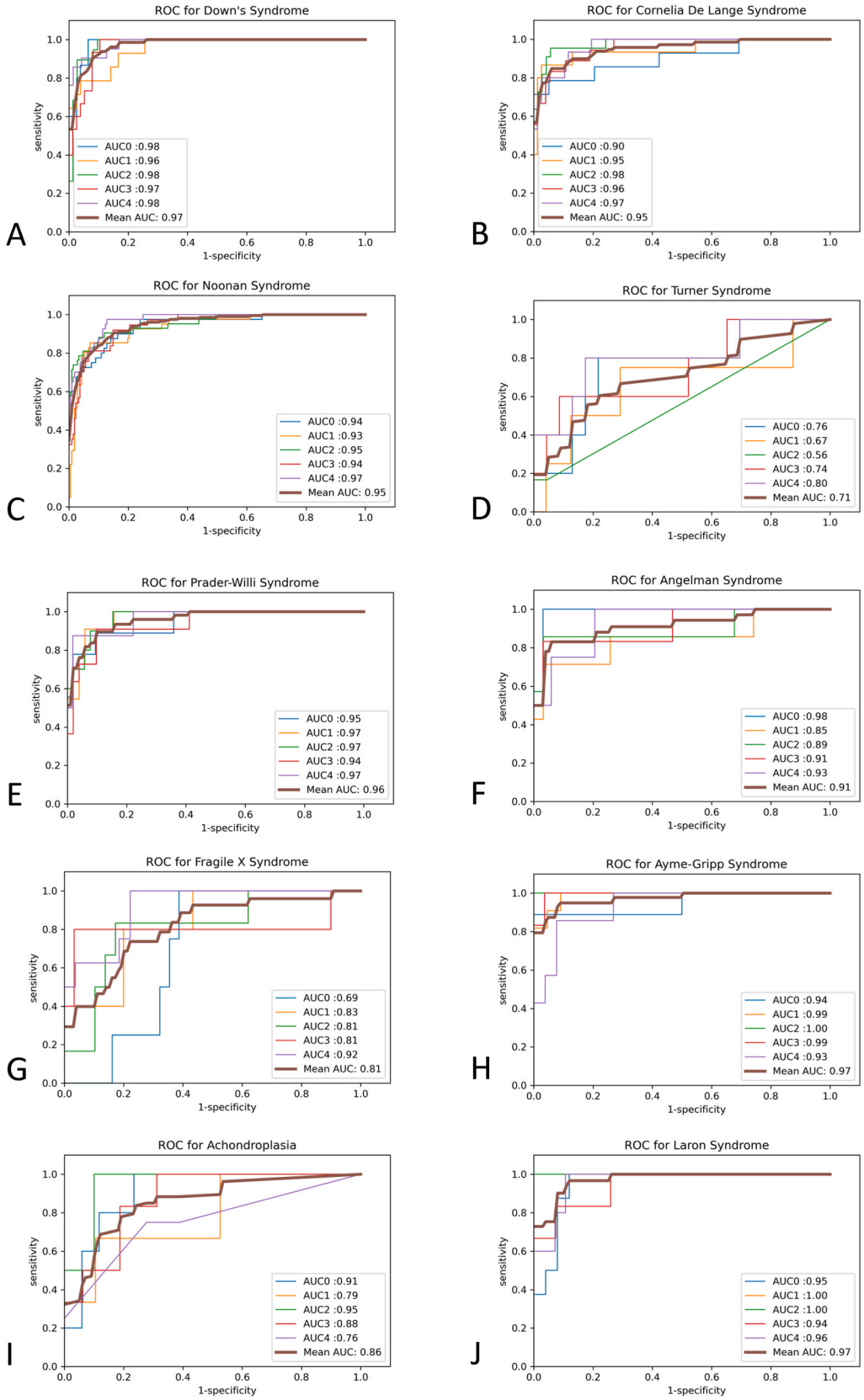

Fig. 2. Diagnostic receiver operating characteristic curve of the AdaBoost model in the 10 endocrine and metabolic syndromes. A. Down syndrome. B. Cornelia de Lange syndrome. C. Noonan syndrome. D. Turner syndrome. E. Prader-Willi syndrome. F. Angelman syndrome. G. Fragile-X syndrome. H. Aymé-Gripp syndrome. I. Achondroplasia. J. Laron syndrome. Abbreviations: ROC, receiver operating characteristic; AUC, area under curve; AdaBoost, adaptive boosting.

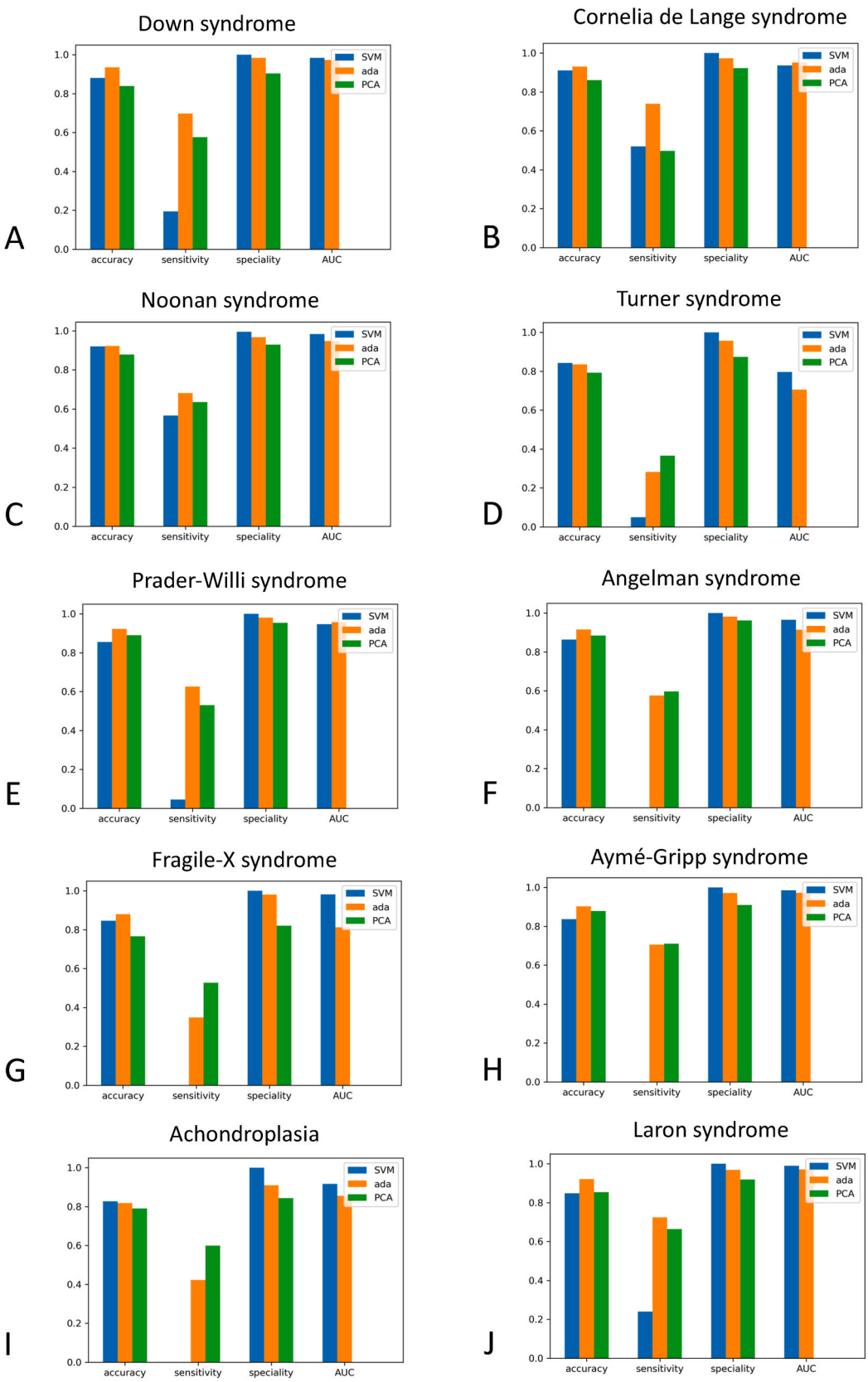

Fig. 3. Comparison of diagnostic efficacy of the three models in the 10 endocrine and metabolic syndromes. The optimal index was defined as the best performance among the three models. A. Down syndrome. B. Cornelia de Lange syndrome. C. Noonan syndrome. D. Turner syndrome. E. Prader-Willi syndrome. F. Angelman syndrome. G. Fragile-X syndrome. H. Aymé-Gripp syndrome. I. Achondroplasia. J. Laron syndrome. Abbreviations: SVM, support vector machine; ada, adaptive boosting; PCA, principal component analysis.

Table 3
Optimal diagnostic performance of artificial intelligence and influencing factors.

Disease	dFRI	Training size	Optimal accuracy	Optimal sensitivity	Optimal specificity	Optimal AUC
Down syndrome	9	68	0.935 (AdaBoost)	0.698 (AdaBoost)	1.000 (SVM)	0.984 (SVM)
Cornelia de Lange syndrome	7.443	58	0.930 (AdaBoost)	0.739 (AdaBoost)	1.000 (SVM)	0.951 (AdaBoost)
Noonan syndrome	8	193	0.923 (AdaBoost)	0.681 (AdaBoost)	0.995 (SVM)	0.983 (SVM)
Turner syndrome	3.36	16	0.843 (SVM)	0.367 (PCA-KNN)	1.000 (SVM)	0.797 (SVM)
Prader-Willi syndrome	6	35	0.923 (AdaBoost)	0.626 (AdaBoost)	1.000 (SVM)	0.957 (SVM)
Angelman syndrome	6	22	0.916 (AdaBoost)	0.597 (PCA-KNN)	1.000 (SVM)	0.965 (SVM)
Fragile-X syndrome	9	17	0.880 (AdaBoost)	0.528 (PCA-KNN)	1.000 (SVM)	0.981 (SVM)
Aymé-Gripp syndrome	6	20	0.903 (AdaBoost)	0.710 (PCA-KNN)	1.000 (SVM)	0.985 (SVM)
Achondroplasia	4	13	0.827 (SVM)	0.599 (PCA-KNN)	1.000 (SVM)	0.917 (SVM)
Laron syndrome	5	20	0.921 (AdaBoost)	0.725 (AdaBoost)	1.000 (SVM)	0.990 (SVM)

Abbreviations: FRI, facial recognition intensity; AUC, area under receiver operating characteristic; AdaBoost, adaptive boosting; SVM, support vector machine; PCA-KNN, principal component analysis-k-nearest neighbor.

This study also has some limitations. Firstly, this study included publicly available facial image data and there were differences in the quality of facial images from different sources and some missing clinical data. These discrepancies could potentially affect the development of diagnostic models due to variations in the number of images available for specific diseases. Nonetheless, the study demonstrates that dFRI exhibits effective evaluation capabilities across varying facial image qualities. Furthermore, certain diseases have limited facial image samples, posing challenges in constructing deep learning neural network models. Exploring alternative AI models might offer additional insights for clinical applications in such scenarios.

5. Conclusion

This research has built a facial image database designed for AI diagnosis of varied endocrine and metabolic syndrome with unique facial features. Employing three classical algorithms, AI diagnostic models demonstrated exceptional accuracy in identifying diseases through facial recognition. Moreover, a deeper exploration revealed that the complexity of disease-specific facial features was an important factor influencing the diagnostic precision performance of AI facial recognition models.

Funding

National High Level Hospital Clinical Research Funding (2022-PUMCH-A-015 to Shi Chen).

National Natural Science Foundation of China (92360307 and 92267103 to Zhen Shen).

Ethics committee approval

This study was approved by the Clinical Ethics Committee of PUMCH (approval ID: I-22PJ418). Written informed consent was obtained from each participant. Researchers were authorized to use the participants' facial images and clinical data for academic research and publication.

Conflict of interest

No financial or non-financial benefits have been received or will be received from any party related directly or indirectly to the subject of this article.

Contributors' statement

DW, SC, and HP conceptualized the study. DW and JQ designed the study protocol. DW recruited patients and collected data. WH and ZS performed the artificial intelligence computation and analysis. HP contributed to data collection. HZ, HD, and HY provided the clinical perspective and data interpretation. SC, JQ, and DW wrote the manuscript. SC, ZS, and HP reviewed and approved the manuscript. All authors agreed on this submission. All authors had access to the data. DW, JQ, and WH verified the data.

Declaration of competing interest

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

References

[1] Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nat Med 2019;25:60–4. https://doi.org/10.1038/s41591-018-0279-0.
[2] Lause M, Kamboj A, Fernandez Faith E. Dermatologic manifestations of endocrine disorders. Transl Pediatr 2017;6:300–12. https://doi.org/10.21037/tp.2017.09.08.
[3] Kosilek RP, Frohner R, Würtz RP, Berr CM, Schopohl J, Reincke M, et al. Diagnostic use of facial image analysis software in endocrine and genetic disorders: review, current results and future perspectives. Eur J Endocrinol 2015;173:M39–44. https://doi.org/10.1530/EJE-15-0429.
[4] Vinkšel M, Writzl K, Maver A, Peterlin B. Improving diagnostics of rare genetic diseases with NGS approaches. J Community Genet 2021;12:247–56. https://doi.org/10.1007/s12687-020-00500-5.
[5] Pei Q, Luo Y, Chen Y, Li J, Xie D, Ye T. Artificial intelligence in clinical applications for lung cancer: diagnosis, treatment and prognosis. Clin Chem Lab Med 2022;60:1974–83. https://doi.org/10.1515/cclm-2022-0291.
[6] Hassan C, Spadaccini M, Iannone A, Maselli R, Jovani M, Chandrasekar VT, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021;93:77–85.e6. https://doi.org/10.1016/j.gie.2020.06.059.
[7] Sechopoulos I, Teuwen J, Mann R. Artificial intelligence for breast cancer detection in mammography and digital breast tomosynthesis: state of the art. Semin Cancer Biol 2021;72:214–25. https://doi.org/10.1016/j.semcancer.2020.06.002.
[8] Wang S, Zhang Y, Lei S, Zhu H, Li J, Wang Q, et al. Performance of deep neural network-based artificial intelligence method in diabetic retinopathy screening: a systematic review and meta-analysis of diagnostic test accuracy. Eur J Endocrinol 2020;183:41–9. https://doi.org/10.1530/EJE-19-0968.
[9] Qiang J, Wu D, Du H, Zhu H, Chen S, Pan H. Review on facial-recognition-based applications in disease diagnosis. Bioengineering 2022;9:273. https://doi.org/10.3390/bioengineering9070273.
[10] Kosilek RP, Schopohl J, Grunke M, Reincke M, Dimopoulou C, Stalla GK, et al. Automatic face classification of Cushing's syndrome in women - a novel screening approach. Exp Clin Endocrinol Diabetes 2013;121:561–4. https://doi.org/10.1055/s-0033-1349124.
[11] Kong X, Gong S, Su L, Howard N, Kong Y. Automatic detection of acromegaly from facial photographs using machine learning methods. EBioMedicine 2018;27:94–102. https://doi.org/10.1016/j.ebiom.2017.12.015.
[12] Chen S, Pan Z-X, Zhu H-J, Wang Q, Yang J-J, Lei Y, et al. 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 2018;8:9317. https://doi.org/10.1038/s41598-018-27586-9.
[13] Pan Z, Shen Z, Zhu H, Bao Y, Liang S, Wang S, et al. Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome. Endocrine 2021;72:865–73. https://doi.org/10.1007/s12020-020-02539-3.
[14] Basel-Vanagaite L, Wolf L, Orin M, Larizza L, Gervasini C, Krantz I, et al. Recognition of the Cornelia de Lange syndrome phenotype with facial dysmorphology novel analysis. Clin Genet 2016;89. https://doi.org/10.1111/cge.12716.

[15] Wu D, Chen S, Zhang Y, Zhang H, Wang Q, Li J, et al. Facial recognition intensity in disease diagnosis using automatic facial recognition. J Personalized Med 2021;11:1172. https://doi.org/10.3390/jpm11111172.
[16] N H, H P, Y R. Machine learning applications in endocrinology and metabolism research: an overview. Endocrinology and Metabolism (Seoul, Korea) 2020;35. https://doi.org/10.3803/EnM.2020.35.1.71.
[17] Wei R, Jiang C, Gao J, Xu P, Zhang D, Sun Z, et al. Deep-learning approach to automatic identification of facial anomalies in endocrine disorders. Neuroendocrinology 2020;110:328–37. https://doi.org/10.1159/000502211.
[18] Muenke M, Adeyemo A, Kruszka P. An electronic atlas of human malformation syndromes in diverse populations. Genet Med 2016;18:1085–7. https://doi.org/10.1038/gim.2016.3.
[19] Zhao Q, Rosenbaum K, Okada K, Zand DJ, Sze R, Summar M, et al. Automated Down syndrome detection using facial photographs. Annu Int Conf IEEE Eng Med Biol Soc 2013;2013:3670–3. https://doi.org/10.1109/EMBC.2013.6610339.
[20] Peprah E. Fragile X syndrome: the FMR1 CGG repeat distribution among world populations. Ann Hum Genet 2012;76:178–91. https://doi.org/10.1111/j.1469-1809.2011.00694.x.
[21] Romano AA, Allanson JE, Dahlgren J, Gelb BD, Hall B, Pierpont ME, et al. Noonan syndrome: clinical features, diagnosis, and management guidelines. Pediatrics 2010;126:746–59. https://doi.org/10.1542/peds.2009-3207.
[22] Kline AD, Moss JF, Selicorni A, Bisgaard A-M, Deardorff MA, Gillett PM, et al. Diagnosis and management of Cornelia de Lange syndrome: first international consensus statement. Nat Rev Genet 2018;19:649–66. https://doi.org/10.1038/s41576-018-0031-0.
[23] Amudhavalli SM, Hanson R, Angle B, Bontempo K, Gripp KW. Further delineation of Aymé-Gripp syndrome and use of automated facial analysis tool. Am J Med Genet 2018;176:1648–56. https://doi.org/10.1002/ajmg.a.38832.
[24] Wang T-S, Tsai W-H, Tsai L-P, Wong S-B. Clinical characteristics and epilepsy in genomic imprinting disorders: Angelman syndrome and Prader-Willi syndrome. Ci Ji Yi Xue Za Zhi 2020;32:137–44. https://doi.org/10.4103/tcmj.tcmj_103_19.
[25] Janecka A, Kołodziej-Rzepa M, Biesaga B. Clinical and molecular features of Laron syndrome, A genetic disorder protecting from cancer. In Vivo 2016;30:375–81.
[26] Pineau M, Farrow E, Nicot R, Ferri J. Achondroplasia: orocraniofacial features and orthodontic-surgical management guidelines proposal. J Craniofac Surg 2018;29:2186–91. https://doi.org/10.1097/SCS.0000000000004819.
[27] Uddin S, Haque I, Lu H, Moni MA, Gide E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Sci Rep 2022;12:6256. https://doi.org/10.1038/s41598-022-10358-x.
[28] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44–56. https://doi.org/10.1038/s41591-018-0300-7.
[29] Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of radiomics in breast cancer diagnosis and prognostication. Breast 2020;49:74–80. https://doi.org/10.1016/j.breast.2019.10.018.
[30] Wang P, Berzin TM, Glissen Brown JR, Bharadwaj S, Becq A, Xiao X, et al. Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study. Gut 2019;68:1813–9. https://doi.org/10.1136/gutjnl-2018-317500.
[31] Wolpert DH, Macready WG. No free lunch theorems for optimization. IEEE Trans Evol Comput 1997;1:67–82. https://doi.org/10.1109/4235.585893.