Yingyan Shi

yyshi17@fudan.edu.cn

Institute of Brain-Inspired Circuits and Systems

Fudan University

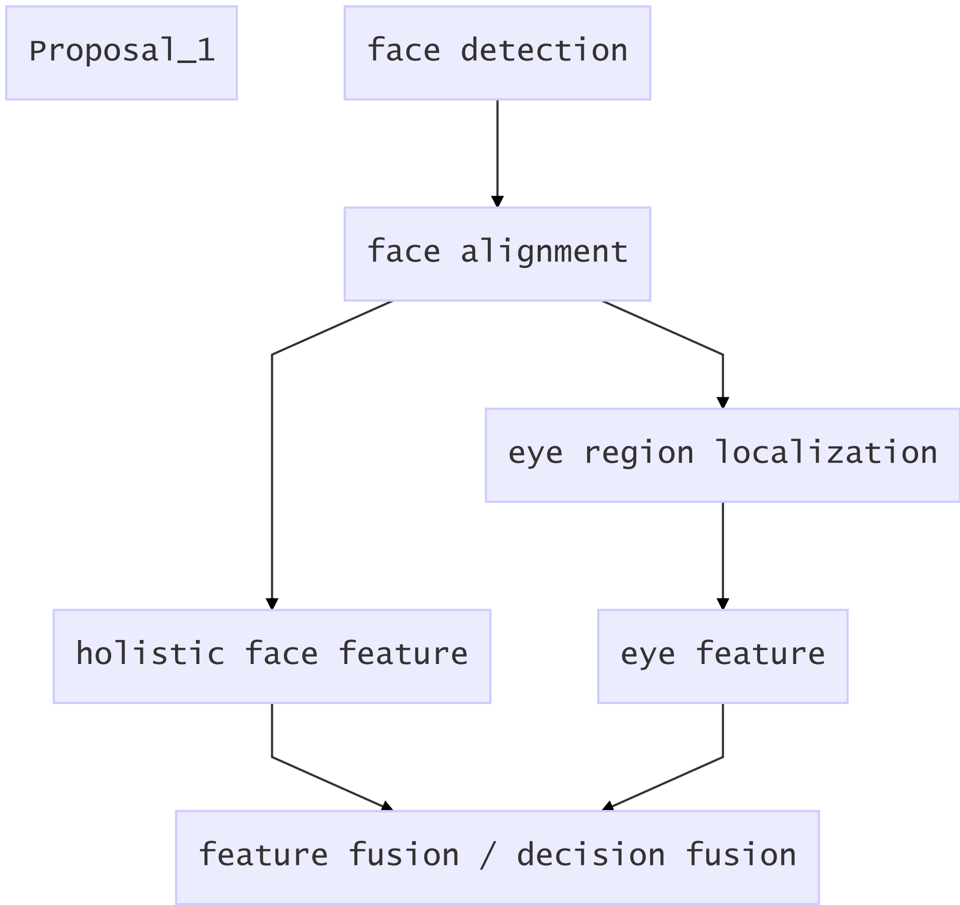
Automatic Facial Symptom Diagnosis

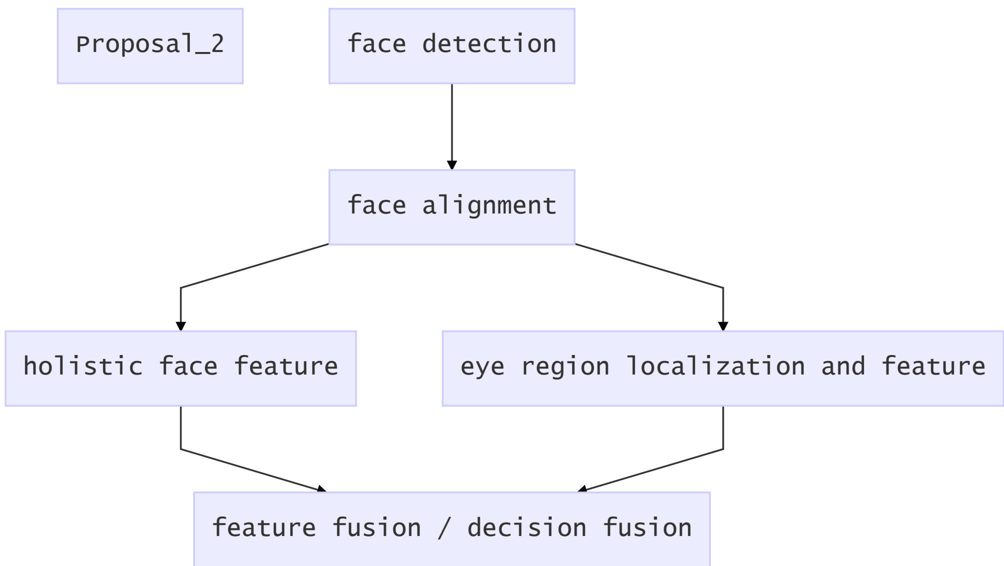
---

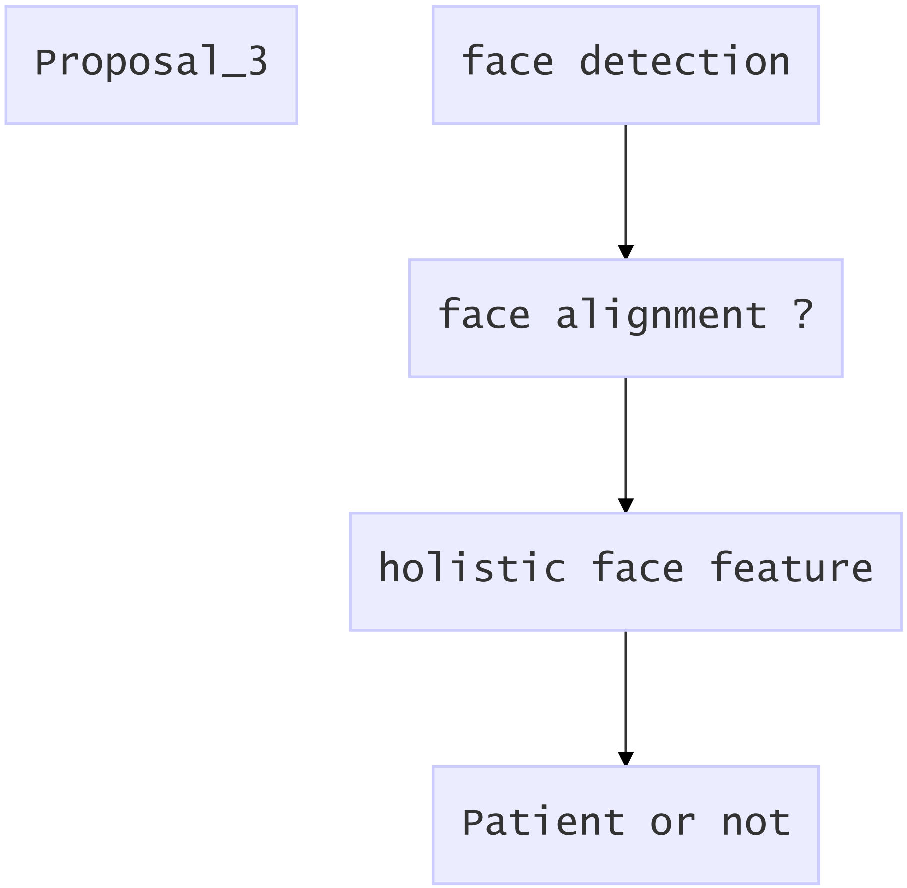
1. **Identifying facial phenotypes of genetic disorders using deep learning**
2. **A deep learning system for differential diagnosis of skin diseases**
3. **The Application of Deep Learning in the Risk Grading of Skin Tumors for Patients Using Clinical Images**
4. **Studies on Different CNN Algorithms for Face Skin Disease Classification Based on Clinical Images**
5. **Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders**

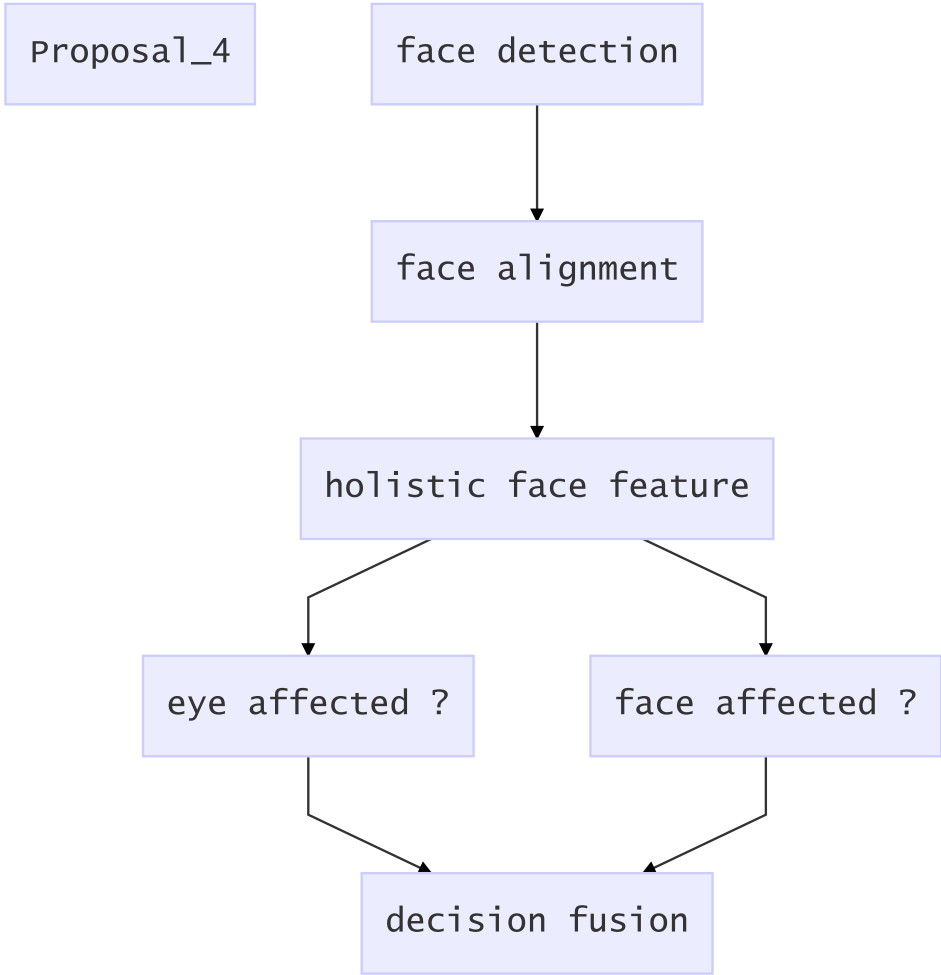
Consideration:

1. directly classify the input image into normal or not
2. compare the input image with the normal one via a threshold
3. since video, many frames can be assembled and averaged
4. insufficient training data and overfitting in deep learning
5. Face alignment? 将public datasets图片进行处理(CycleGAN)，使之类似于病脸，解决domain adaption problem，关键是landmark严重影响分类的效果
6. Data augmentation
7. 脸色青紫(whole face)和眼球浑浊(local characteristics)可以一起预测， holistic vs. local attributes
8. 有没有其他的facial attribute可以用来表征，discover attribute relationships adaptively
9. how to make the best of the relationships among attributes?
10. how to make networks focus more on the locations of attributes?
11. class-imbalanced learning: data re-sampling or cost-sensitive learning
12. different data sources, a. k. a. the domain adaption problem
13. the combination of different face-related tasks, considering the inherent dependencies, 姿态估计
14. some methods with good accuracies might have tremendous computation or memory costs, 考虑部署在边缘端
15. every attribute has an individual classifier or a single multi-task classifier?
16. end-to-end optimization, individual loss + ultimate loss
17. t-SNE visualization of faces
18. 同一个人正常与病态的paired images， contrastive loss
19. nominal label ｛-1, +1｝ or ｛0, 1｝？corresponding loss: MSE or BCE?
20. training and testing sets splitting matters
21. 利用弱监督学习来让神经网络自动查找和属性相关的区域，可以来辅助医学知识发现，参考A Deep Cascade Network for Unaligned Face Attribute Classification
22. 输出值与阈值比较，得出有病或无病
23. 分类性能，速度，模型大小，参数数量
24. 人脸检测和对齐是自己训练还是依赖他人的算法
25. 要不要搭建一个系统出来，还是只是算法原型?
26.


Proposal_1
face detection
face alignment
eye region localization
holistic face feature
eye feature
feature fusion / decision fusion


Proposal_2
face detection
face alignment
holistic face feature
eye region localization and feature
feature fusion / decision fusion


Proposal_3
face detection
face alignment ?
holistic face feature
Patient or not



1. how to properly assign shared information and attribute-specific information at different layers of networks?
2. how to explore relationships among distinct attributes for learning more discriminative features?

1. Identifying facial phenotypes of genetic disorders using deep learning

Nature Medicine | VOL 60 25 | JANUARY **2019** | 60–64 |

Israel, Germany, UCSD

Average the training images illustrating the general appearance of each cohort (genetic mutation)

Recognizable facial features are highly informative to clinician.

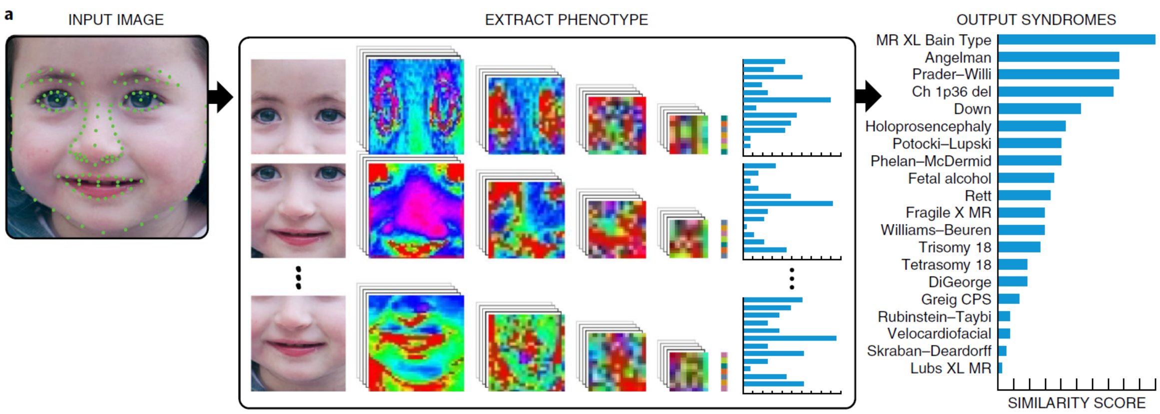
Binary classification problem: distinguishing a specific syndrome from a set of other syndromes.

differentiating unaffected from affected individuals

Methodology:

Given an input image,

- face detection
- facial alignment
- cropped into multiple regions whose results are averaged
  - Face, upper half
  - Middle face (ear to ear)
  - Face, lower half
  - Full face
- scaled to a fixed size (100x100 pixels)
- converted to grayscale
- initially trained on CASIA-WebFace for face identification (transfer learning)
- fine-tuned to the syndrome domain



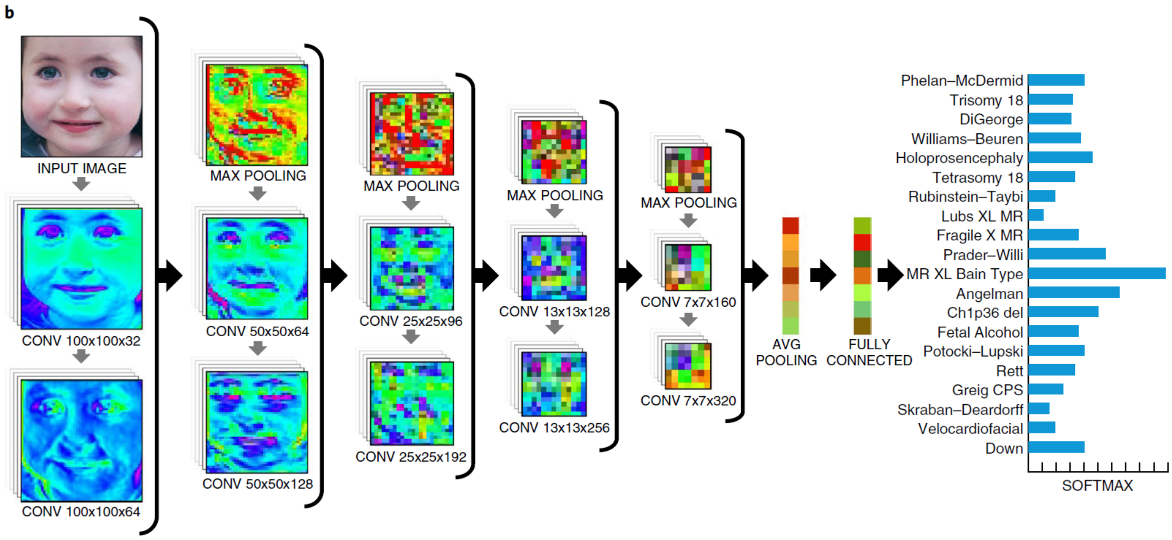


Fig. 1 | DeepGestalt: high-level flow and network architecture. **a**, A new input image is first preprocessed to achieve face detection, landmarks detection and alignment. After preprocessing, the input image is cropped into facial regions. Each region is fed into a DCNN to obtain a softmax vector indicating its correspondence to each syndrome in the model. The output vectors of all regional DCNNs are then aggregated and sorted to obtain the final ranked list of genetic syndromes. The histogram on the right-hand side represents DeepGestalt's output syndromes, sorted by the aggregated similarity score. **b**, The DCNN architecture of DeepGestalt. A snapshot of an image passing through the network. The network consists of ten convolutional layers, and all but the last are followed by batch normalization and a rectified linear unit (ReLU). After each pair of convolutional (CONV) layers, a pooling layer is applied (maximum pooling after the first four pairs and average pooling after the fifth pair). This is then followed by a fully connected layer with dropout (0.5) and a softmax layer. A sample feature map is shown after each pooling layer. It is interesting to compare the low-level features of the first layers with respect to the high-level features of the final layers; the latter identify more complex features in the input image, and distinctive facial traits tend to emerge while identity-related features disappear. The photograph is published with parental consent.

DeepGestalt, like many artificial intelligence systems, cannot explicitly explain its predictions and provides no information about which facial features drove the classification.

Pre-Training:

- Keras with TensorFlow as the backend
- CASIA WebFace dataset, 494,414 images from 10,575 different subjects
- He Normal Initializer
- ADAM for 40 epochs
    - initial learning rate 1x10^-3
- cross-entropy loss
- SGD for 10 epochs
    - learning rate 1x10^-4
- momentum 0..9
- dropout (0.5)
- batch normalization

Fine-Tuning:

- Xaiver Normal Initializer with a scale of 0.3
- SGD for 500 epochs
- learning rate 5x10^-3
- momentum 0.9
- No weight decay or kernel regularization

Data Augmentation: (each region)

- rotation with a range of 5 degrees
- small vertical and horizontal shifts with a shift range of 0.05
- shear transformation with a shear range of 5π/180
- random zoom with a range of 0.05

- horizontal flip

Backbone: Learning Face Representation from Scratch, CASIA (arXiv: 2014)

Name	Type	Filter Size /Stride	Output size	Depth	#Params
Conv11	convolution	3×3 / 1	100×100×32	1	0.28K
Conv12	convolution	3×3 / 1	100×100×64	1	18K
Pool1	max pooling	2×2 / 2	50×50×64	0	
Conv21	convolution	3×3 / 1	50×50×64	1	36K
Conv22	convolution	3×3 / 1	50×50×128	1	72K
Pool2	max pooling	2×2 / 2	25×25×128		
Conv31	convolution	3×3 / 1	25×25×96	1	108K
Conv32	convolution	3×3 / 1	25×25×192	1	162K
Pool3	max pooling	2×2 / 2	13×13×192	0	
Conv41	convolution	3×3 / 1	13×13×128	1	216K
Conv42	convolution	3×3 / 1	13×13×256	1	288K
Pool4	max pooling	2×2 / 2	7×7×256	0	
Conv51	convolution	3×3 / 1	7×7×160	1	360K
Conv52	convolution	3×3 / 1	7×7×320	1	450K
Pool5	avg pooling	7×7 / 1	1×1×320		
Dropout	dropout (40%)		1×1×320	0	
Fc6	fully connection		10575	1	3305K
Cost1	softmax		10575	0	
Cost2	~~contrastive~~		1	0	
Total				11	5015K

Table 2. The architecture of the proposed baseline convolutional network.

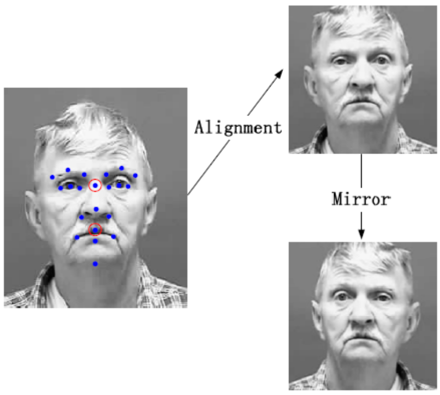
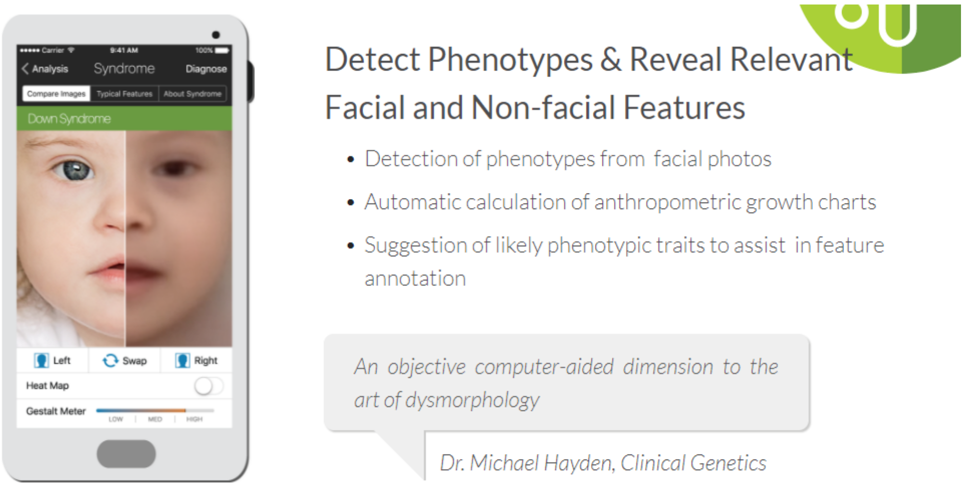


Figure 4. Face image alignment and augmentation. The read circles on the face are two selected landmarks for similarity transformation.

Compared to the most used eye centers, the distance between the selected two landmarks here is relative invariant to pose variations in yaw angle. After normalization, the distance between the two points is 25 pixels. Because face has nearly symmetric structure, we double the training set by mirror operation, which can result the representations more robust to pose variations.

Face2Gene (proprietary technology) https://www.face2gene.com/



灵敏度（**Sensitivity**，也称为真阳性率）是指实际为阳性的样本中，判断为阳性的比例（例如真正有生病的人中，被医院判断为有生病者的比例），计算方式是真阳性除以真阳性+假阴性（实际为阳性，但判断为阴性）的比值（能将实际患病的病例正确地判断为患病的能力，即患者被判为阳性的概率）。

特异度（**Specificity**，也称为真阴性率）是指实际为阴性的样本中，判断为阴性的比例（例如真正未生病的人中，被医院判断为未生病者的比例），计算方式是真阴性除以真阴性+假阳性（实际为阴性，但判断为阳性）的比值（能正确判断实际未患病的病例的能力，即试验结果为阴性的比例）。

阳性预测值是指真阳性人数占试验结果阳性人数的百分比，表示试验结果阳性者属于真病例的概率。

阴性预测值是指真阴性人数占试验结果阴性人数的百分比，表示试验结果阴性者属于非病例的概率。

准确度（**accuracy**）也称效率（efficiency），用真阳性与真阴性人数之和占受试人数的百分率表示。

2. A deep learning system for differential diagnosis of skin diseases

Google Health, UC San Francisco, MIT

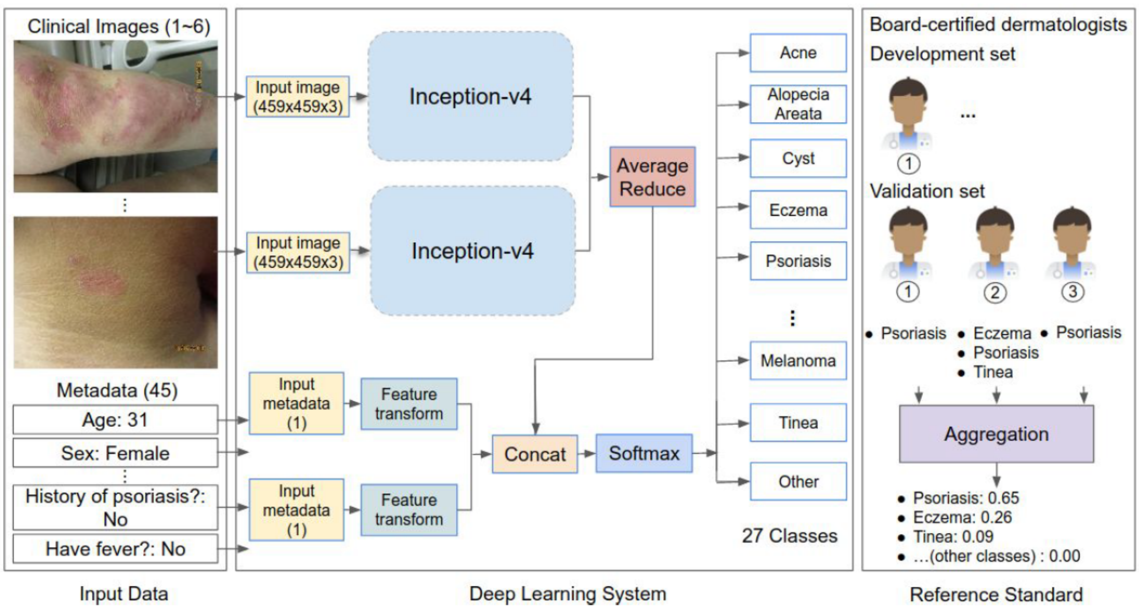


Fig. 1 | Overview of the development and validation of our deep learning system (DLS).

Backbone: **Inception-v4**, Inception-ResNet and the Impact of Residual Connections on Learning. Google (AAAI 2017)

3. The Application of Deep Learning in the Risk Grading of Skin Tumors for Patients Using Clinical Images

Journal of Medical Systems (2019) 43: 283

Central South University, Xiangya Hospital Central South University

Backbone: We fine-tuned the **Xception** pretrained on ImageNet to complete the task of soring degree.

Risk grading: low-disk, high-risk, dangerous (output)

4. Studies on Different CNN Algorithms for Face Skin Disease Classification Based on Clinical Images

SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH

Central South University, Xiangya Hospital Central South University

1. ResNet-50
2. Inception-v3
3. DenseNet121
4. Xception
5. Inception-ResNet-v2

All pretrained on ImageNet, random reverse and crop.

To address the problem of data imbalance, we used different weights in the cost function for different diseases.

Input size: 300x300

We replaced the first fully connected layer behind the last convolutional layer with global average pooling and a 1x1 convolution to reduce the number of parameters and maintain spatial information.

ResNet50	Inception V3	*DenseNet-121*	*Xception*	*Inception-Resnet V2*
Input layer 10*10*64, S=2, P=3*3, S=2, Out=75*75*64 ↓ **Conv1.x** [1*1*64, 3*3*64, 1*1*256]*3 Out=75*75*256 ↓ **Conv2.x** [1*1*128, 3*3*128, 1*1*512]*4 Out=38*38*512 ↓ **Conv3.x** [1*1*256, 3*3*256, 1*1*1024] *6 Out=19*19*1024 ↓ **Conv4.x** [1*1*512, 3*3*512, 1*1*2048] *3 Out=10*10*2048 ↓ **Global Average Pool** Out=1*1*2048 **Fully Connected** Out=1024 **Softmax** Out=6	**Input layer** 3*3*32, S=2, Out=150*150*32 ↓ **Conv1.x** 3*3*32, S=1, 3*3*64, S=1, P=3*3, S=2, Out=75*75*64 ↓ **Conv2.x** 3*3*80, S=1, 3*3*192, S=2, 3*3*288, S=1, Out=38*38**288 ↓ **3 * inception block** Out=19*19*768 **5 * inception block** Out=10*10*1280 **2 * inception block** Out=10*10*2048 ↓ **Global Average Pool** Out=1*1*2048 **Fully Connected** Out=1024 **Softmax** Out=6	**Input layer** 10*10*16, S=2, P=3*3, S=2, Out=75*75*16 ↓ **Dense_block1** [1*1, 3*3]*6 Out=38*38*160 **Transition layer1** 1*1, P=2*2, S=2, Out=38*38*160 ↓ **Dense-block2** [1*1, 3*3]*12 Out=38*38*304 **Transition layer2** 1*1, P=2*2, S=2, Out=19*19*304 ↓ **Dense-block3** [1*1, 3*3]*24 Out=19*19*448 **Transition layer3** [1*1, 3*3]*12 Out=10*10*448 ↓ **Dense-block4** [1*1, 3*3]*16 Out=10*10*595 ↓ **Global Average Pool** Out=1*1*1507 **Fully Connected** Out=1024 **Softmax** Out=6	**Input layer** 3*3*32, S=2, 3*3*64, Out=150*150*64 ↓ **Entry flow** [3*3, 3*3, P=3*3, S=2] *3 Concat each block by 1*1 Conv and S=2 Out=19*19*728 ↓ **Middle Flow** [3*3*728] *3 *8 ↓ **Exit flow** 3*3*728, *3*1024, P=3*3, S=2, 3*3*1536, 3*3*2048, Out=10*10*2048 ↓ **Global Average Pool** Out=1*1*2048 **Fully Connected** Out=1024 **Logistic Regression** Out=6	**Input layer** **(Stem)** 3*3*32, S=1, 3*3*32, 3*3*64, Out=147*147*64 ↓ **Mix** Mixed_3a Out=73*73*160 Mixed_4b Out=71*71*192 Mixed_5a Out=35*35*384 ↓ **5*inception-resnet-A** Out=35*35*256 **Reduction-A** Out=17*17*896 ↓ **10*inception-resnet-B** Out=17*17*896 **Reduction-B** Out=8*8*1792 ↓ **5*inception-resnet-C** Out=8*8*1792 ↓ **Global Average Pool** Out=1*1*1792 **Fully Connected** Out=1024 **Softmax** Out=6

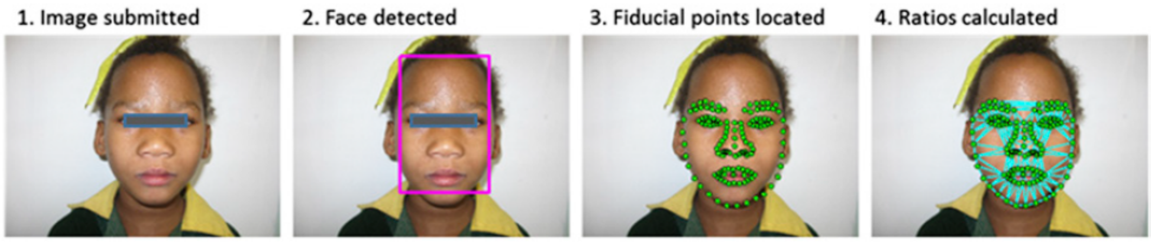
Net		BCC	LE	ROS	SK	AK	SCC	Average
Resnet50	Recall (%)	74.2	78.6	53.3	76.3	42.7	55.3	**63.4**
	Precision (%)	60.4	59.5	94.7	53.0	48.5	61.5	**62.9**
Incpetion V3	Recall (%)	79.2	87.6	62.6	64.5	47.6	58.2	**66.6**
	Precision (%)	45.5	62.1	92.9	74.5	53.3	55.6	**64.0**
Densenet-121	Recall (%)	76.9	84.8	58.2	76.5	47.6	65.2	**68.2**
	Precision (%)	57.5	60.1	93.6	61.9	62.4	77.1	**68.8**
Xception	Recall (%)	83.1	87.8	58.1	82.2	50.3	62.1	**70.6**
	Precision (%)	65.9	57.5	81.2	75.4	55.8	72.5	**68.1**
Inception-Resnet V2	Recall (%)	89.2	92.9	66.7	84.3	54.1	74.6	**77.0**
	Precision (%)	63.7	59.2	95.0	84.3	53.3	69.1	**70.8**

5. Computer-Aided Recognition of Facial Attributes for Fetal Alcohol Spectrum Disorders

PEDIATRICS Volume 140, number 6, December 2017

University of Arizona, The University of New Mexico, University of Mississippi Medical Center, University of Nebraska Medical Center

Face2Gene: robustness to individual differences in variables such as ethnicity, sex, and age, skewed image characteristics such as illumination, pose, and expression.



A description of the face under investigation is conducted as multiple lengths, angles. Ratios are computed for each face. These values are then used in combination with statistical mechanisms analyzed statistically to evaluate for the presence of dysmorphic features.

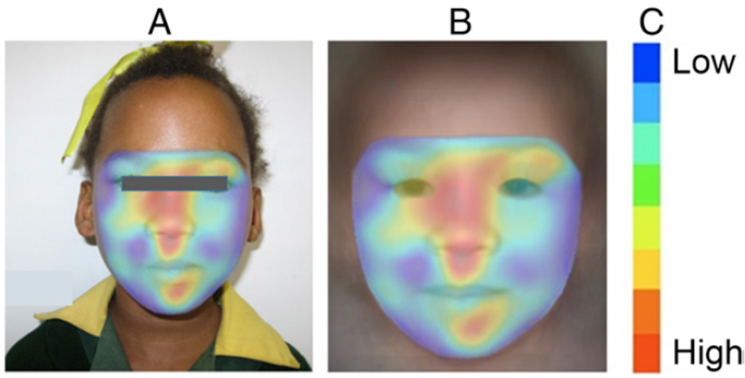


FIGURE 3

A, The FAS heatmap overlaid on the input image. B, The same heatmap overlaid on a typical FAS face. C, Heatmap index. A facial region most supportive of the FAS classification is marked in red, while regions that are less supportive are marked with cooler colors.