Predictive analytics

for a facial emotion recognition task

Group 8

Jiancong Shen, Vikki Sui, Jinxu Xiang, Ruiqi Xie, Wenjie Xie

Presenter: Jinxu Xiang

CONTENTS

Data preprocessing

Including rotation, zoom and translation.

Data preprocessing - 1

The area surrounded by the green square is in the range of 100 and 300 of x and y. This is used to help confirm the location of the image pixels.

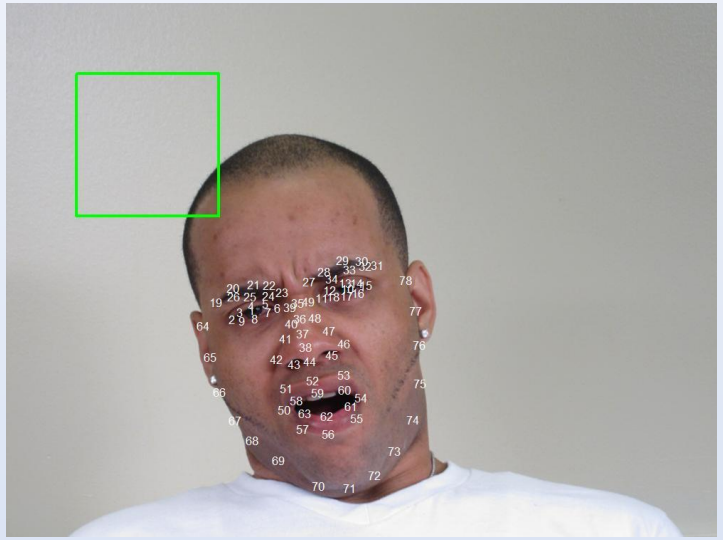


Image 2375

The person's head in this picture is severely tilted to the right and the face area is small.

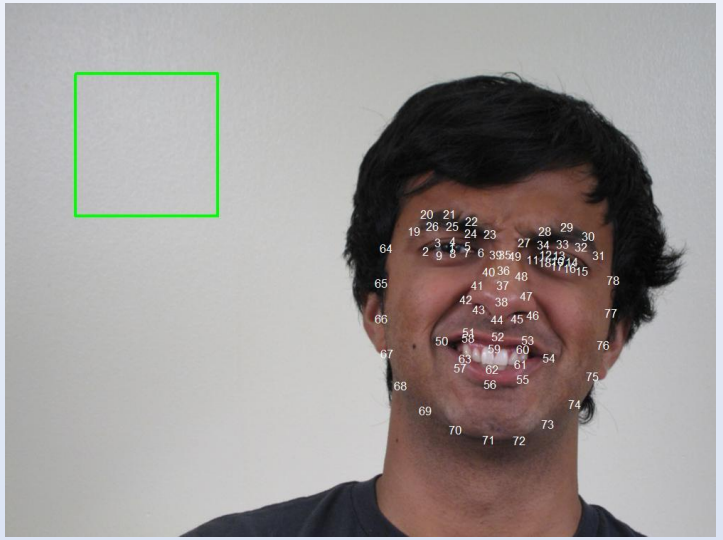


Image 0944

The person's head in this picture is slightly tilted to the left, the face area is small, and most of them are on the right side of the image.

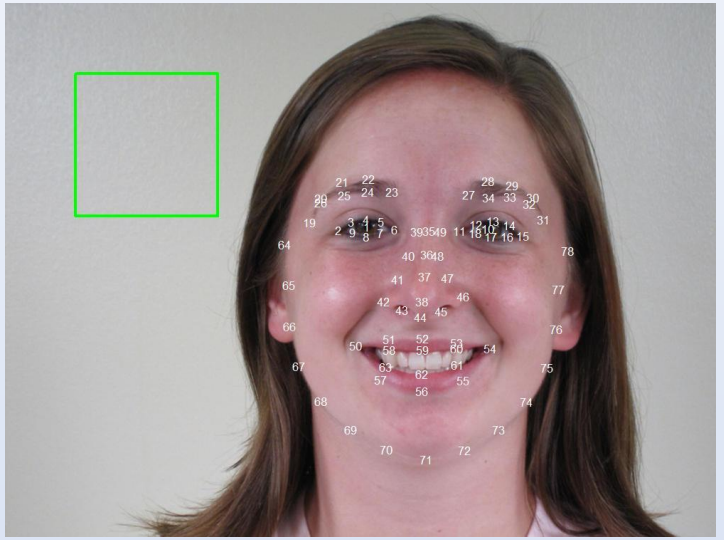


Image 0113

The person's face area is slightly off-center.

Data preprocessing - 2

The blue points are the midpoint of the symmetrical points of the face and the single points of the face.
The red line is the regression line of the blue points.

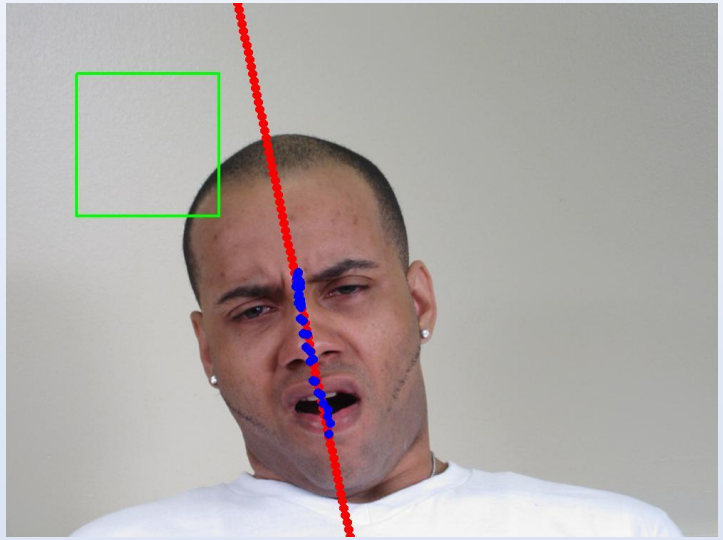


Image 2375

The person's head in this picture is severely tilted to the right and the face area is small.

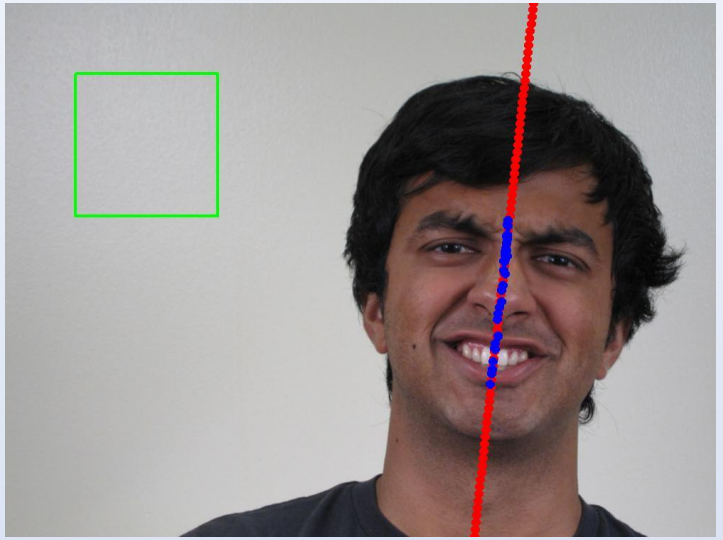


Image 0944

The person's head in this picture is slightly tilted to the left, the face area is small, and most of the area is on the right side of the image.

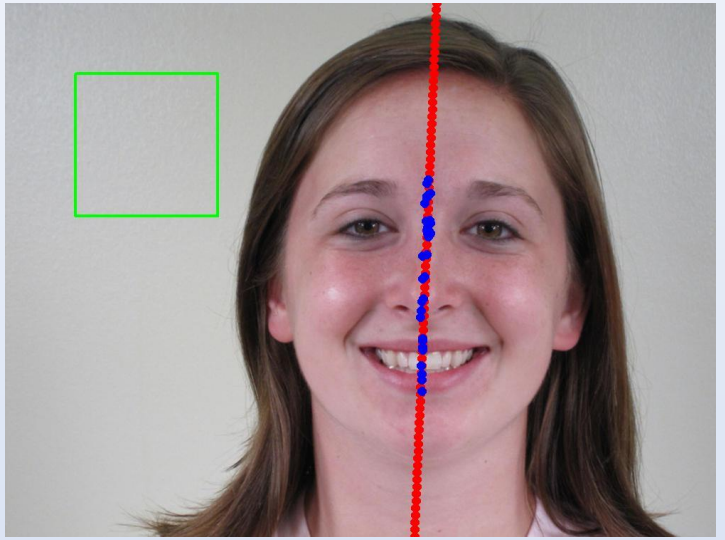


Image 0113

The person's face area is slightly off-center.

Data preprocessing - 3

The black area is created by the rotation function of EBImage. We can see that everyone's face are now vertical.

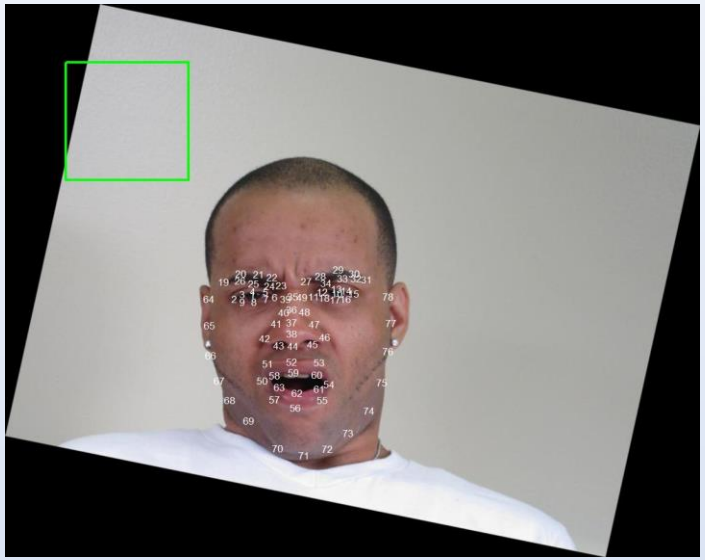


Image 2375

The person's face area is small and his head in not exactly at the center of the image.

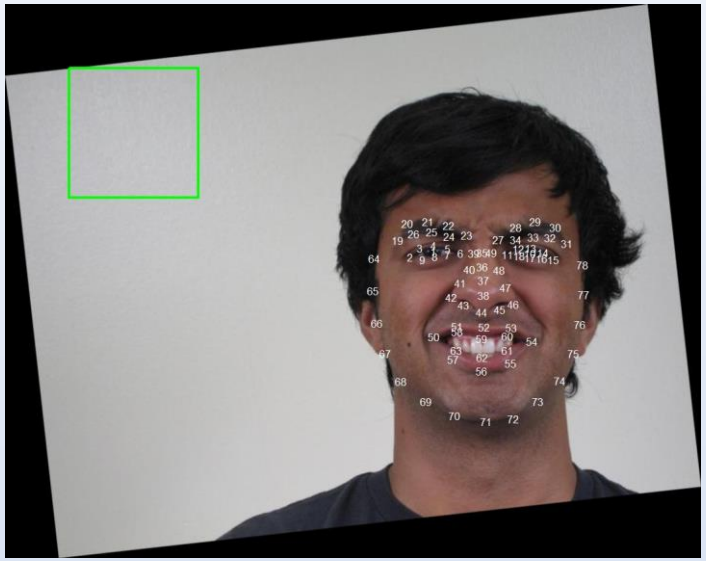


Image 0944

The person's face area is small and most of the area is on the right side of the image.

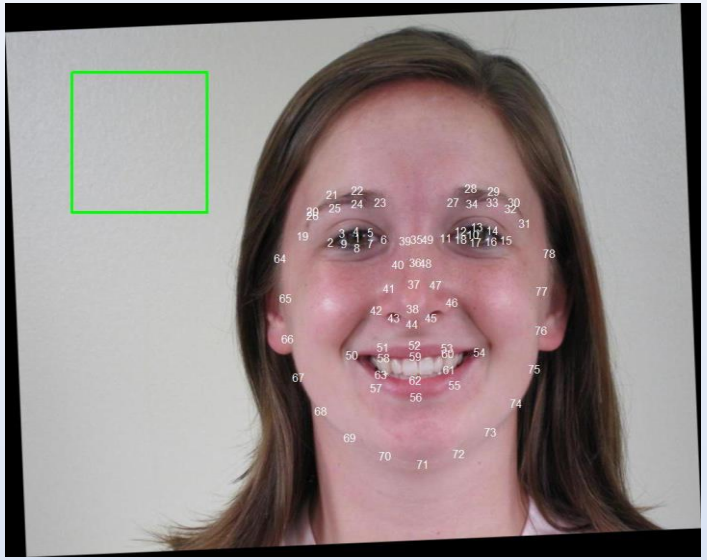


Image 0113

The person's face area is slightly off-center.

Data preprocessing - 4

We choose the distance between the midpoint of two pupils and the bottom of the nose as a fix distance to zoom every images. Looking at the area of the green square, it still represents 100-300 for x and y.

Image 2375

The person's face area is slightly off-center.

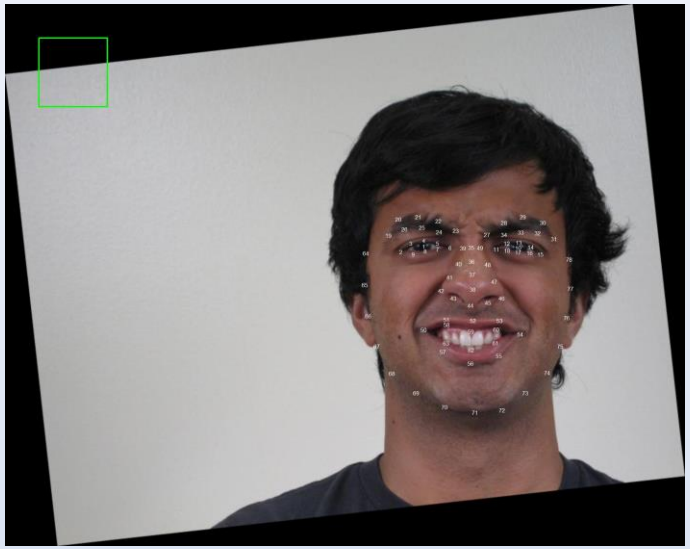


Image 0944

The person's face area is off-center.

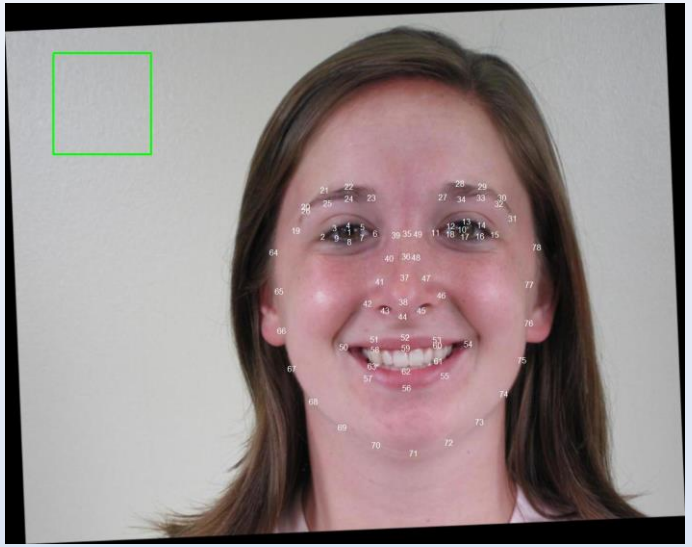


Image 0113

The person's face area is slightly off-center.

Data preprocessing - 5

Although the face area has become smaller visually, it contains more pixels.
Then we translate image's number 37 point to the (500, 375). And put the other points to the appropriate position.

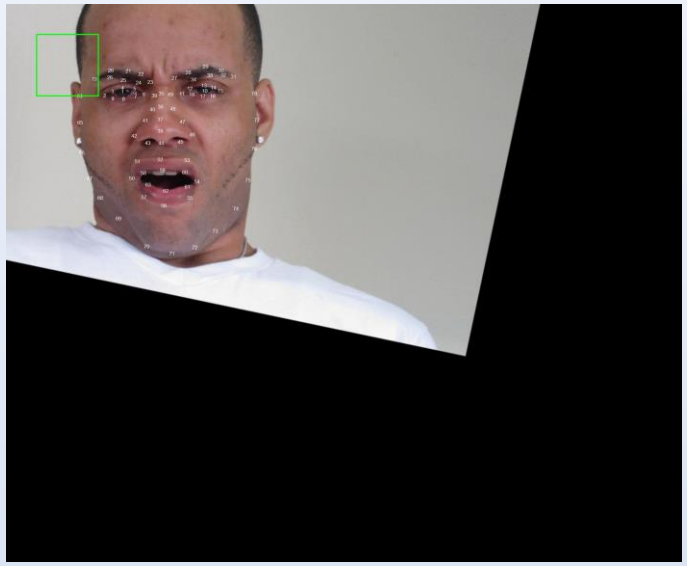


Image 2375

☆☆☆☆☆

All good!

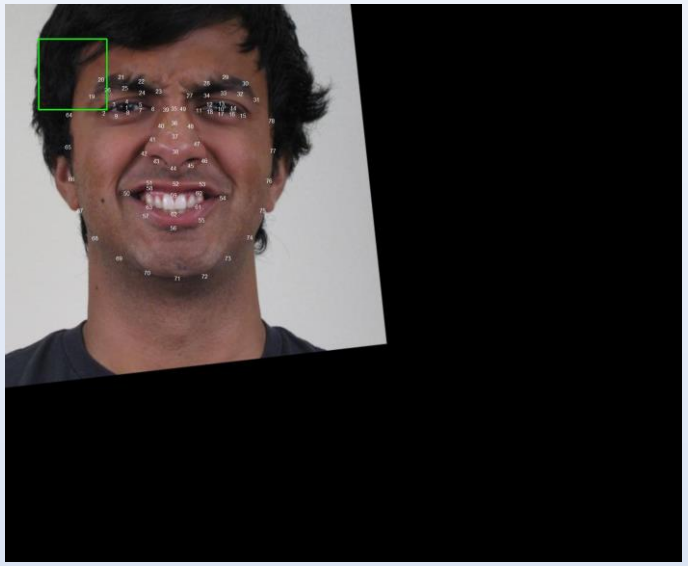


Image 0944

☆☆☆☆☆

All good!

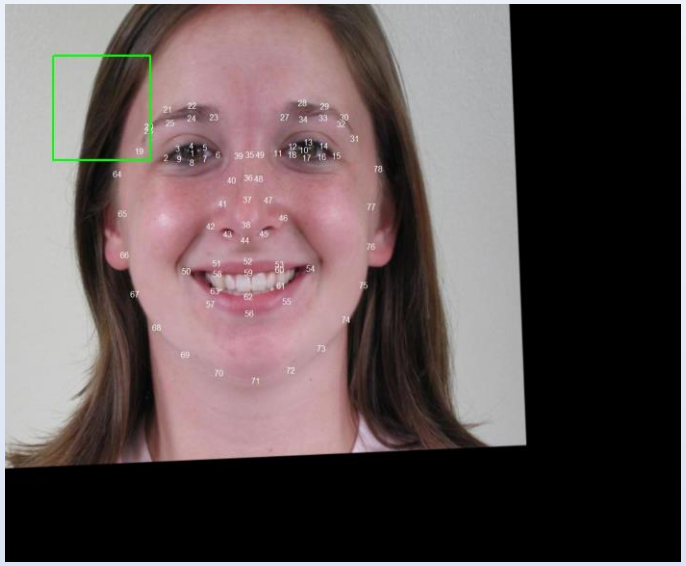


Image 0113

☆☆☆☆☆

All good!

Data preprocessing - 6

Choose part of image to observe possible features.
The blue points is new feature points after rotation, zoom and translation.
The red points is the center (500,375)

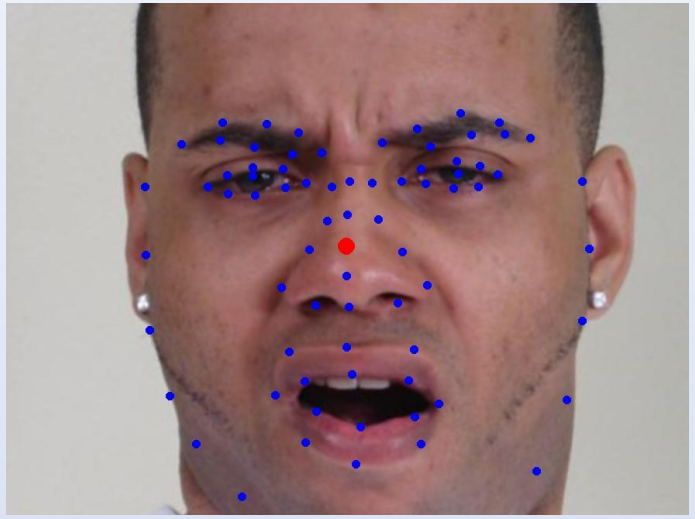


Image 2375

Perfect!

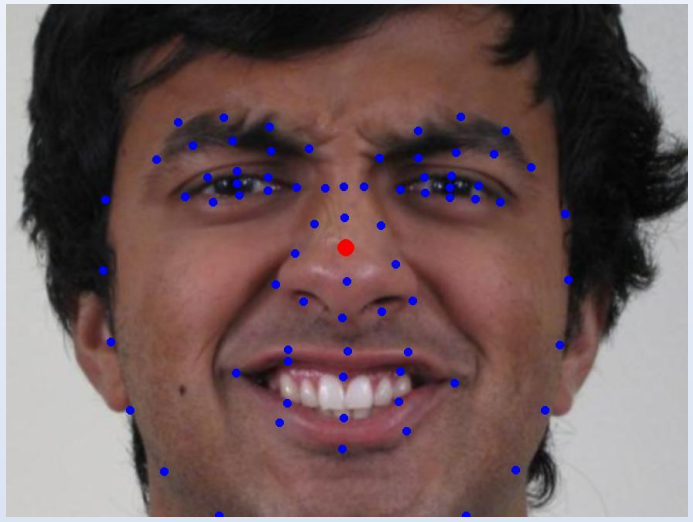


Image 0944

Perfect!

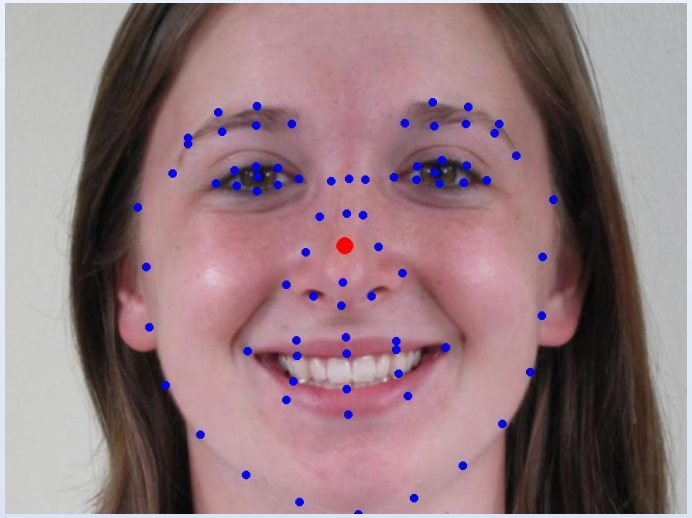


Image 0113

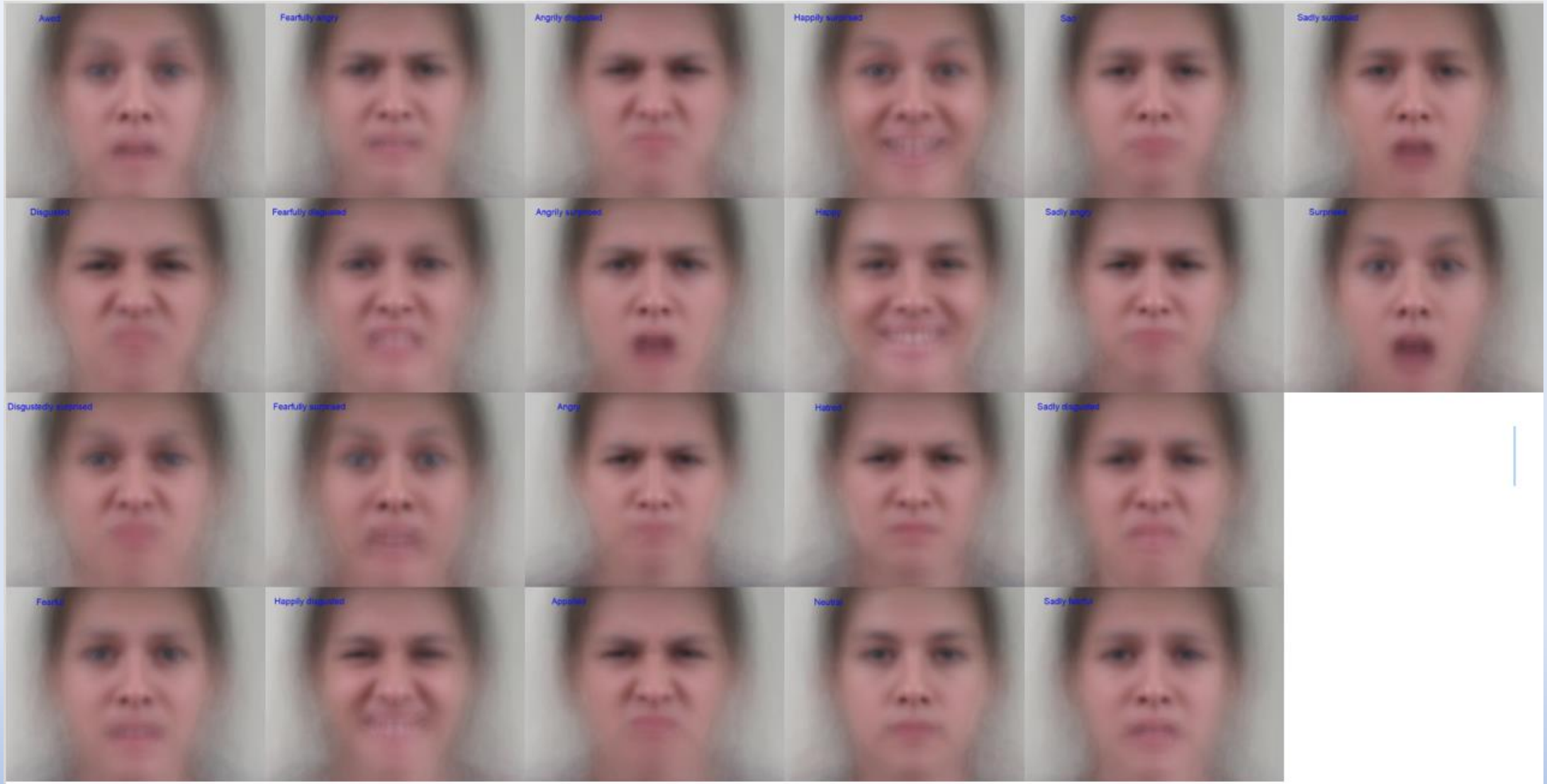
Perfect!

What can we do base on these images?

Data preprocessing - 7

HD images are saved in the ‘fig’ folder of our project.
It looks like there are more female images in our training set.

The average image of each emotion!



Feature extraction

Including points features and image features

Feature extraction

The color is green means we used this feature.
The color is grey means we didn't choose it.

What kind of features can we use?

Point coordinates

Since we have just finished the data preprocessing, the positions of all points should be standardized. Their coordinates are valid features. We have 78 feature points. Except for the center point with the same location, the feature dimension is 154.

Geometric features

In addition to distance, the slope between two points, the angle between three points, the area within the point set, and the regression coefficient of several points are all features that can summarize the expression. We took about 40 such features.

Distance features

In the baseline, 78 points pairwise distance are used, but we don't need so much. In fact, we can select some efficient distances to express nearly all the information. We extracted about 90 distance features.

Color features

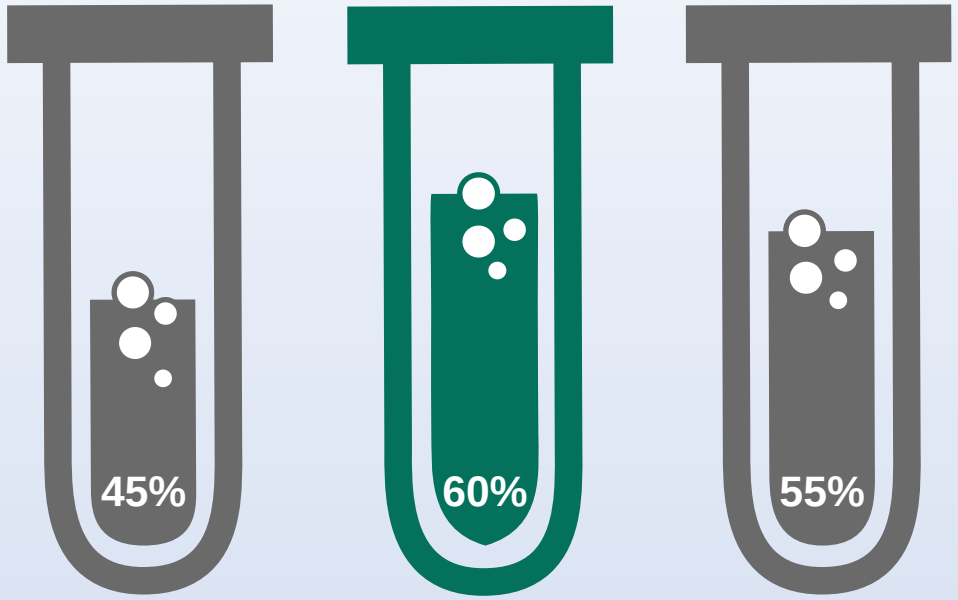
In addition to the features that can be obtained from 78 points, information can also be obtained from the image, such as forehead wrinkles and nasolabial folds. We use a proportion of pixels much darker than skin tones in an area to express these features. The dimension of these features is 10.

Classifier application

Including GBM, SVM and XGBoost.

Classifier application

When randomly selecting a 20% test set, we are not randomly selecting 20% images, but randomly selecting all images of 20% individuals.



Gradient Boosting Machine(GBM)

Classifier parameters:

n.trees = 200,
bag.fraction = 0.8,
shrinkage = 0.1,
cv.folds = 3.

Features and dimension:

Baseline (distance), dim = 6006.
Improved (distance + geometric), dim = 129.
Colored (distance + geometric + color), dim = 139.

Support Vector Machine(SVM)

Classifier parameters:

kernel = 'poly', degree = 1, gamma = 0.008

Features and dimension:

Improved (distance + geometric), dim = 129.
Colored (distance + geometric + color), dim = 139.

eXtreme Gradient Boosting(XGB)

Classifier parameters:

eta = 0.05,
gamma = 0.1,
colsample_bytree = 0.6,
nrounds = 500,

Features and dimension:

Improved (distance + geometric), dim = 129.
Colored (distance + geometric + color), dim = 139.

We extracted color features and used the SVM classifier as our improved model.

04

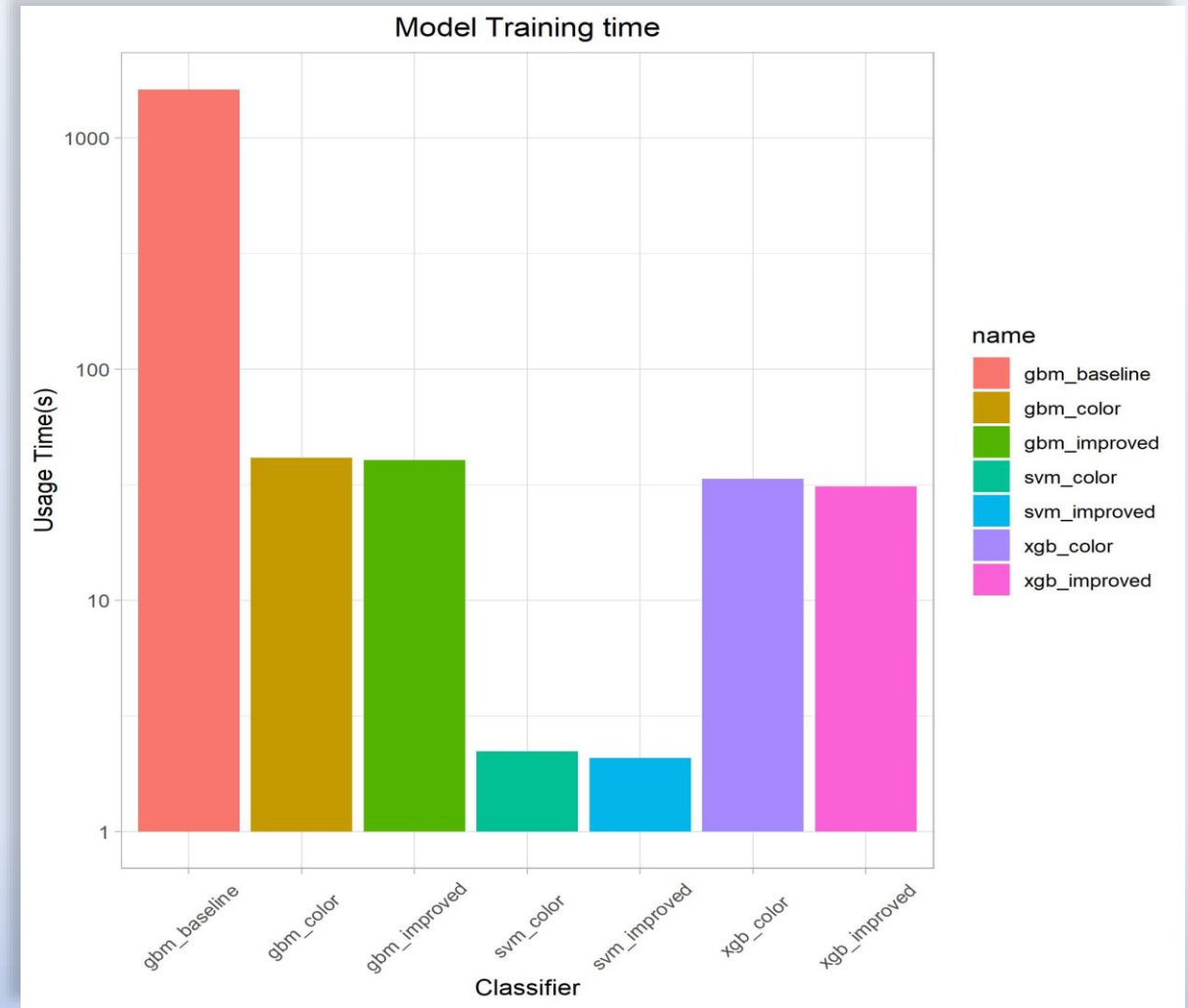
Conclusion

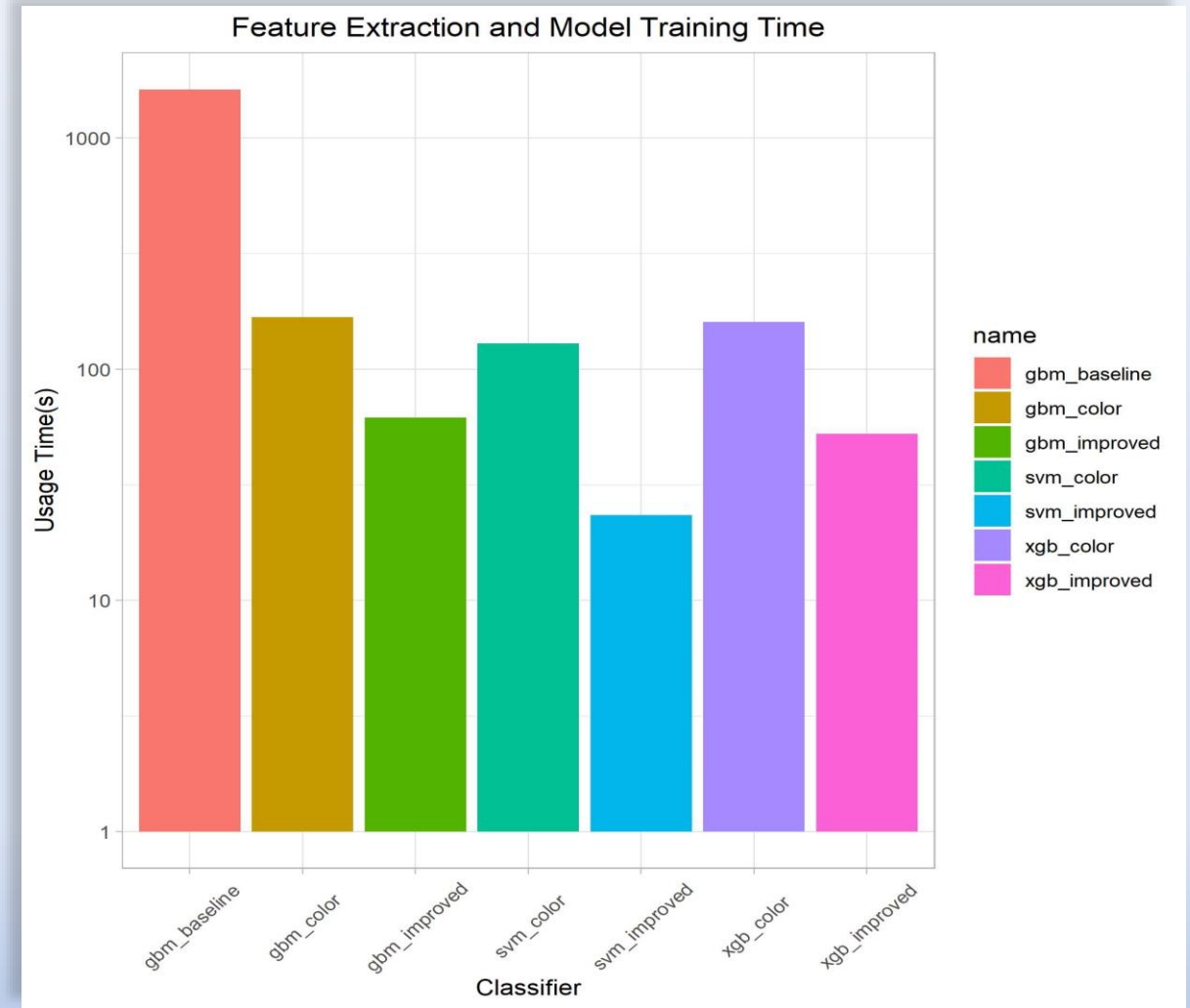
Including accuracy analysis and time analysis

Conclusion - 1

All training time =
Data preprocessing time + Feature extraction time + Model training time

Training time for different methods

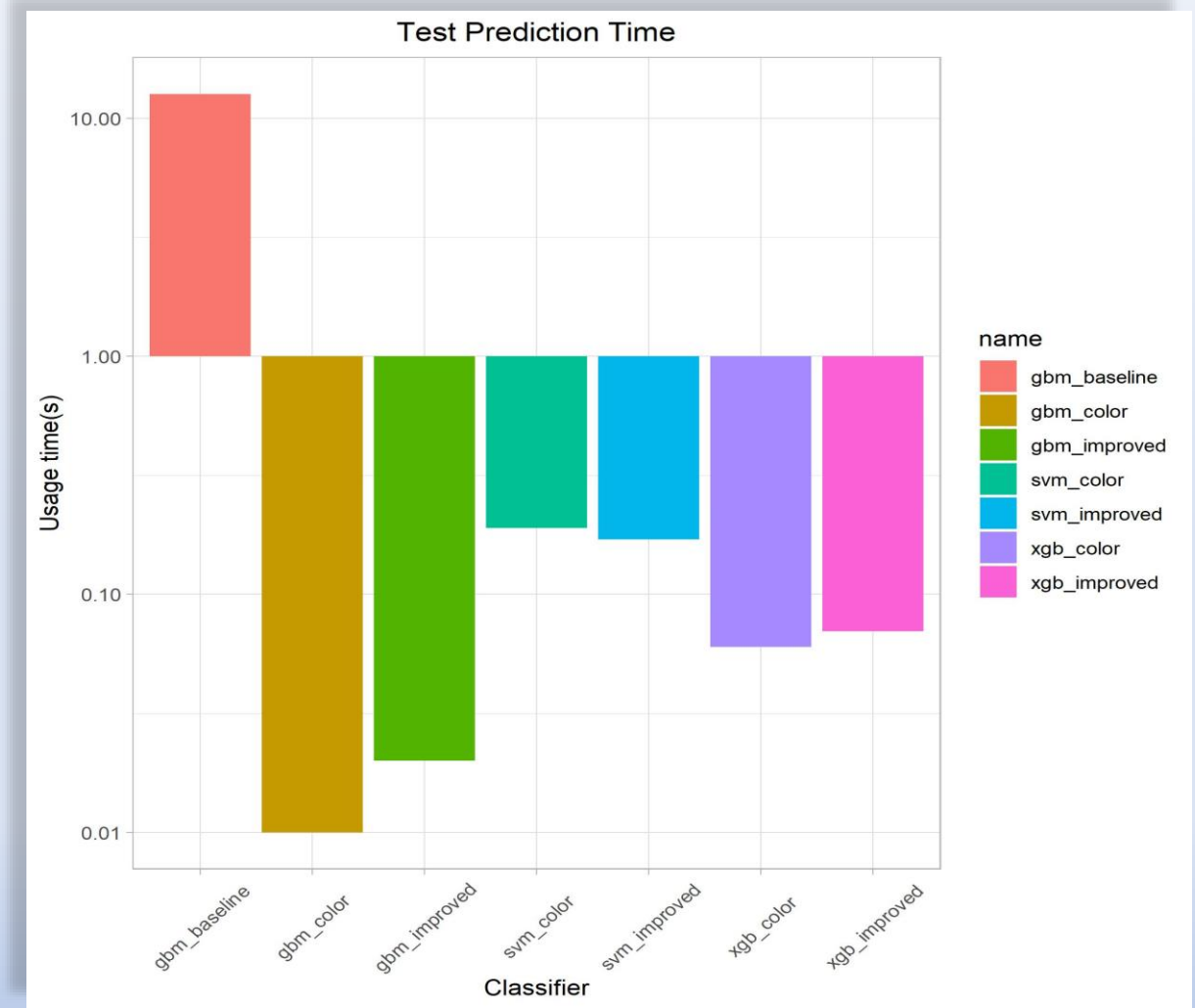


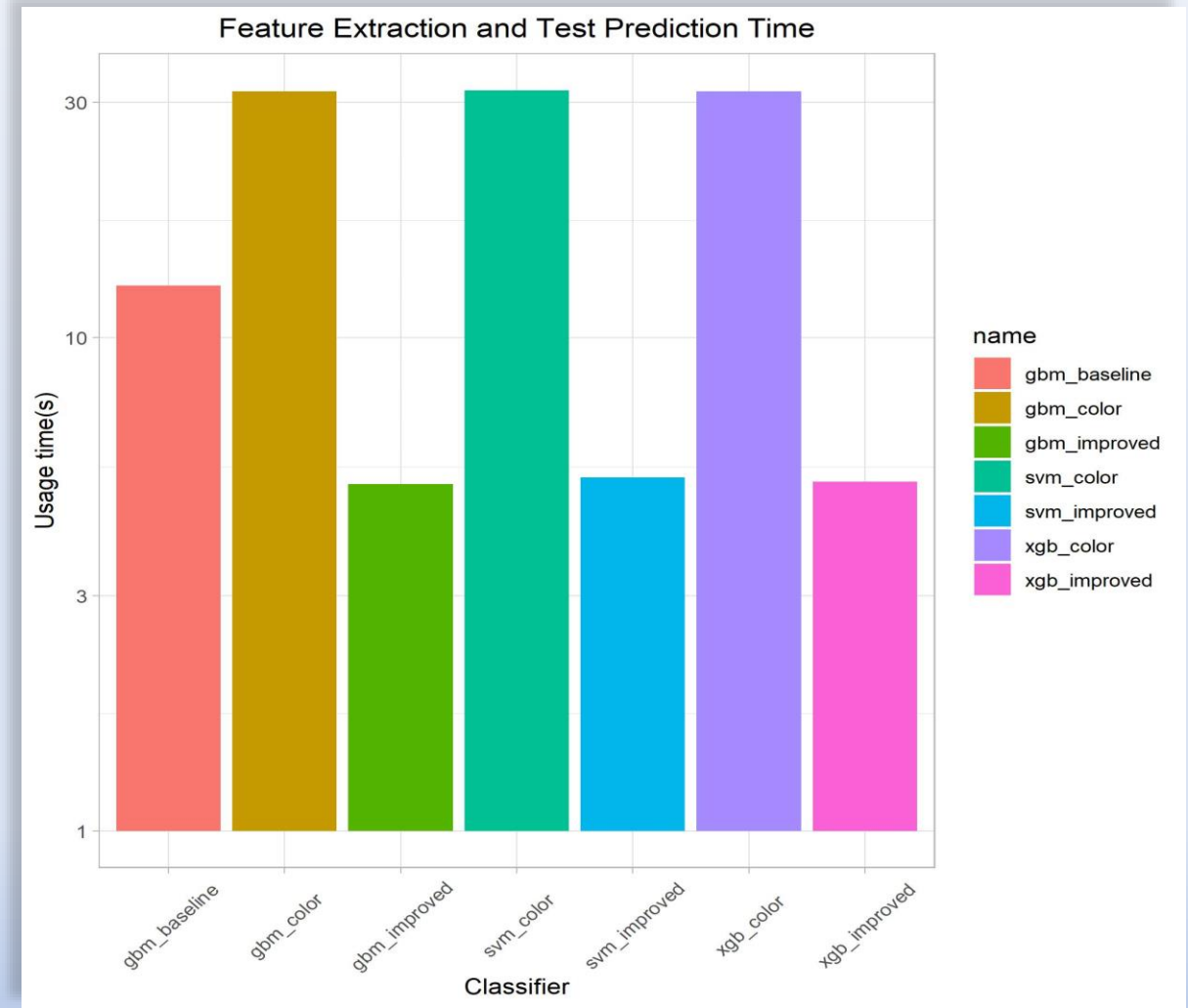


Conclusion - 2

All test time =
Data preprocessing time + Feature extraction time + Test prediction time

Test time for different methods

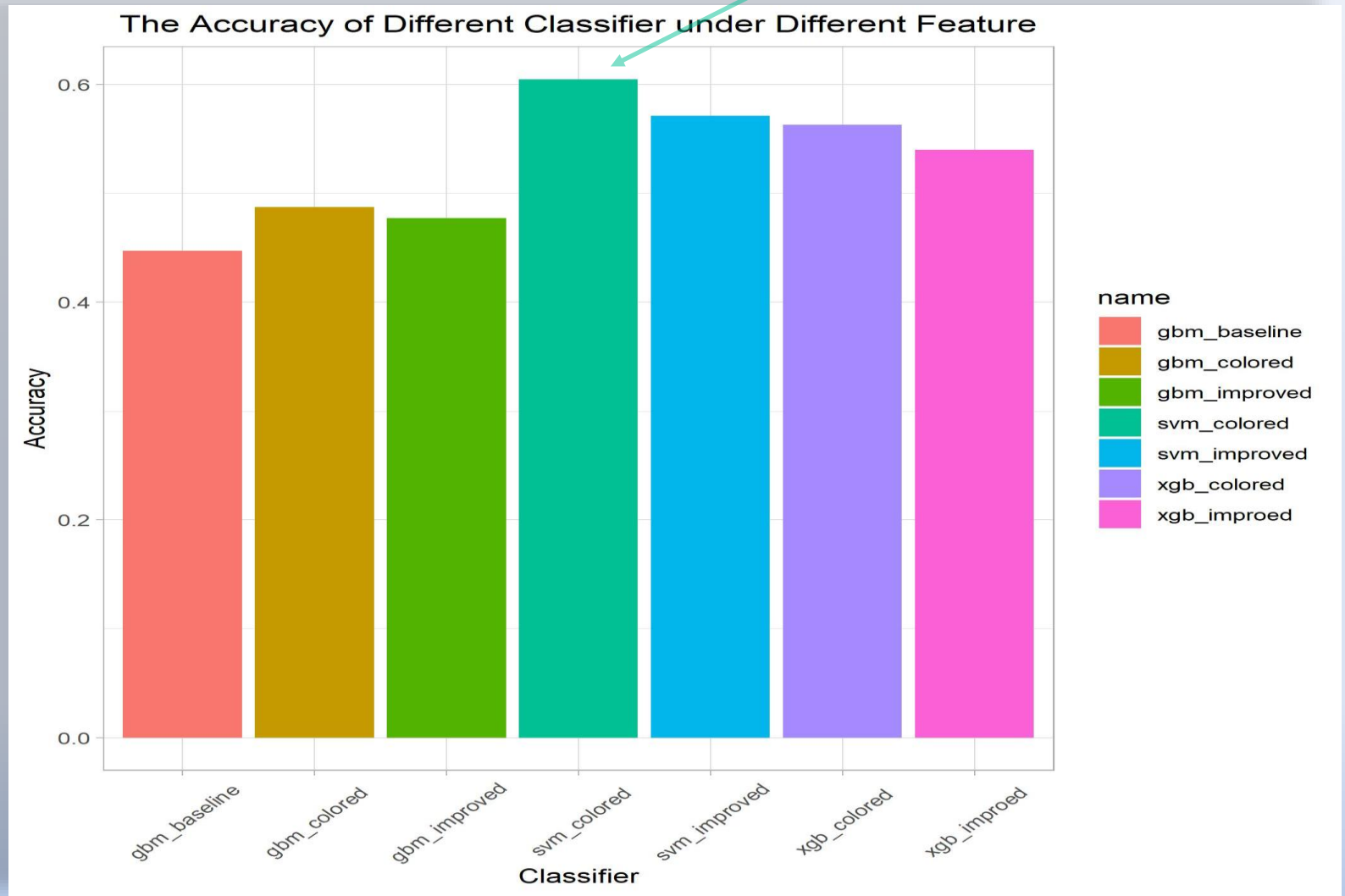




Conclusion - 3

60.46%

Prediction accuracy of different methods



If you re-plan the training and test data set each time, the resulting prediction accuracy will also be different.

Under same training and test data set, the training results of GBM and XGB are different each time. But results of SVM is always the same.

Conclusion - 4

This matrix is the result from the SVM classifier using color features.
Note that the number of test data for each facial expression is not equal.

Which classifications are most confusing?

Confusion Matrix and Statistics

Prediction \ Reference	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	19	0	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	19	0	0	0	0	0	5	3	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	0	14	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0	0	1	1
4	0	0	0	14	0	1	0	0	0	7	1	1	1	0	0	1	0	0	0	0	0	0
5	0	0	0	0	16	0	1	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0
6	0	0	0	2	0	13	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	2
7	0	0	0	0	0	0	16	0	0	0	0	0	0	0	2	1	2	0	1	0	1	0
8	0	1	0	0	0	0	0	16	0	0	0	0	0	0	0	0	2	0	0	0	0	0
9	0	1	0	0	0	0	0	0	17	1	0	0	0	0	0	0	0	0	0	1	1	1
10	0	0	1	1	0	0	0	0	0	11	1	0	2	0	0	0	0	0	0	0	1	0
11	0	0	0	1	0	1	0	0	0	1	11	2	1	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	1	0	0	0	1	7	12	5	0	0	0	0	0	0	0	0	3
13	0	0	1	3	0	3	0	0	0	1	1	2	4	1	0	0	0	0	0	1	1	2
14	0	0	0	0	0	1	0	0	0	0	0	1	0	18	0	0	0	0	1	0	1	0
15	0	0	0	1	0	0	1	0	0	0	0	0	0	3	9	0	0	0	2	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	17	5	2	2	3	0
18	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	6	9	1	0	0	0
19	0	0	4	0	0	0	1	0	0	1	0	0	0	1	2	0	0	0	7	1	1	0
20	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	10	1	2
21	0	0	0	0	0	1	2	0	0	1	0	0	0	1	1	0	1	0	5	4	14	2
22	0	0	2	0	0	2	0	0	0	1	0	2	1	0	1	1	0	0	1	3	1	11

As you can see from the confusion matrix on the left, the number 10 expression and the number 4 expression are highly confused. These two number indicate ‘Sadly angry’ and ‘Angry’. The confusing classifications are as follows：

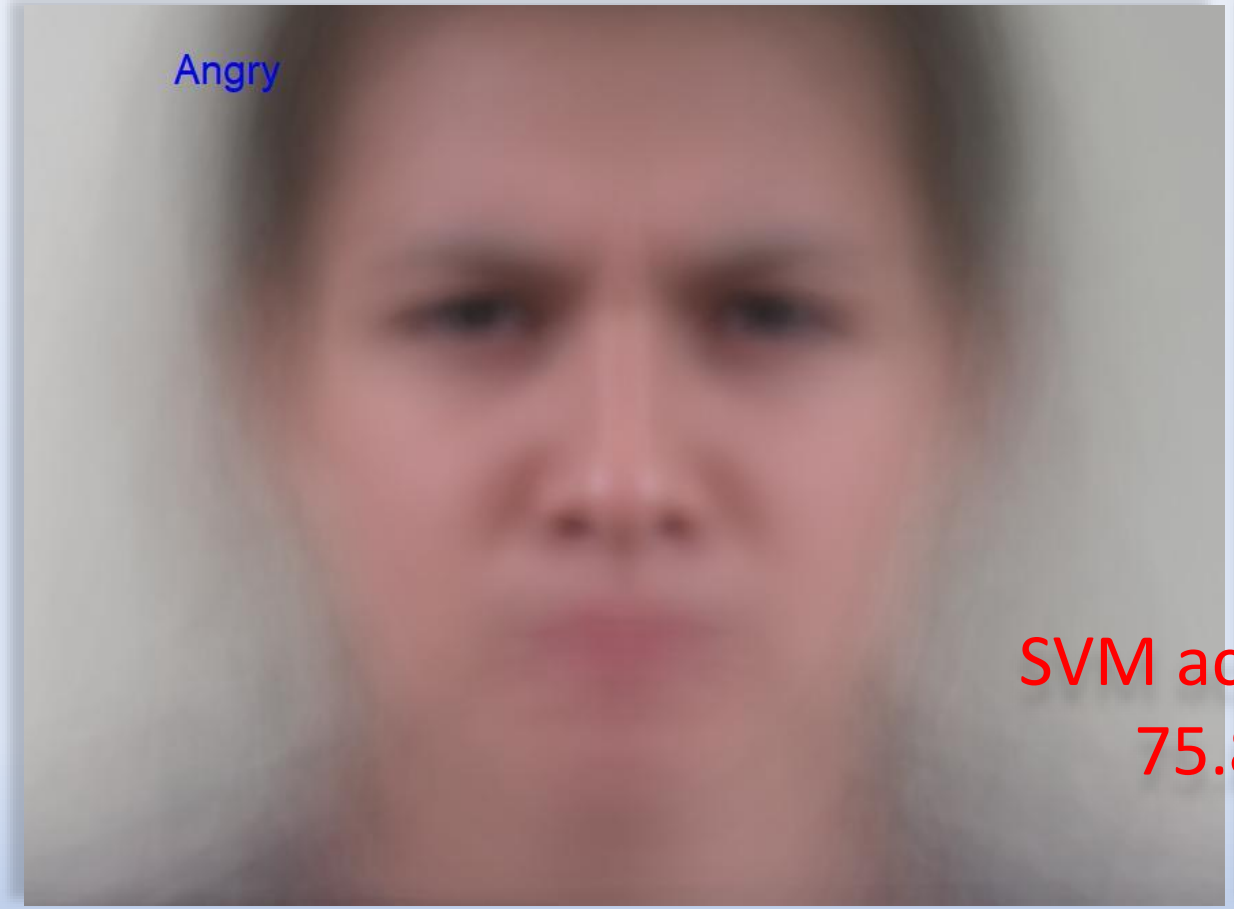
‘Angry’ and ‘Sadly angry’;
‘Appalled‘ and ’Hatred’;
‘Awed’ and ‘Fearfully surprised’;
‘Happy’ and ‘Happily surprised’;
‘Sadly fearful’, ‘Fearfully disgusted’ and ‘Fearfully angry’ ;
……

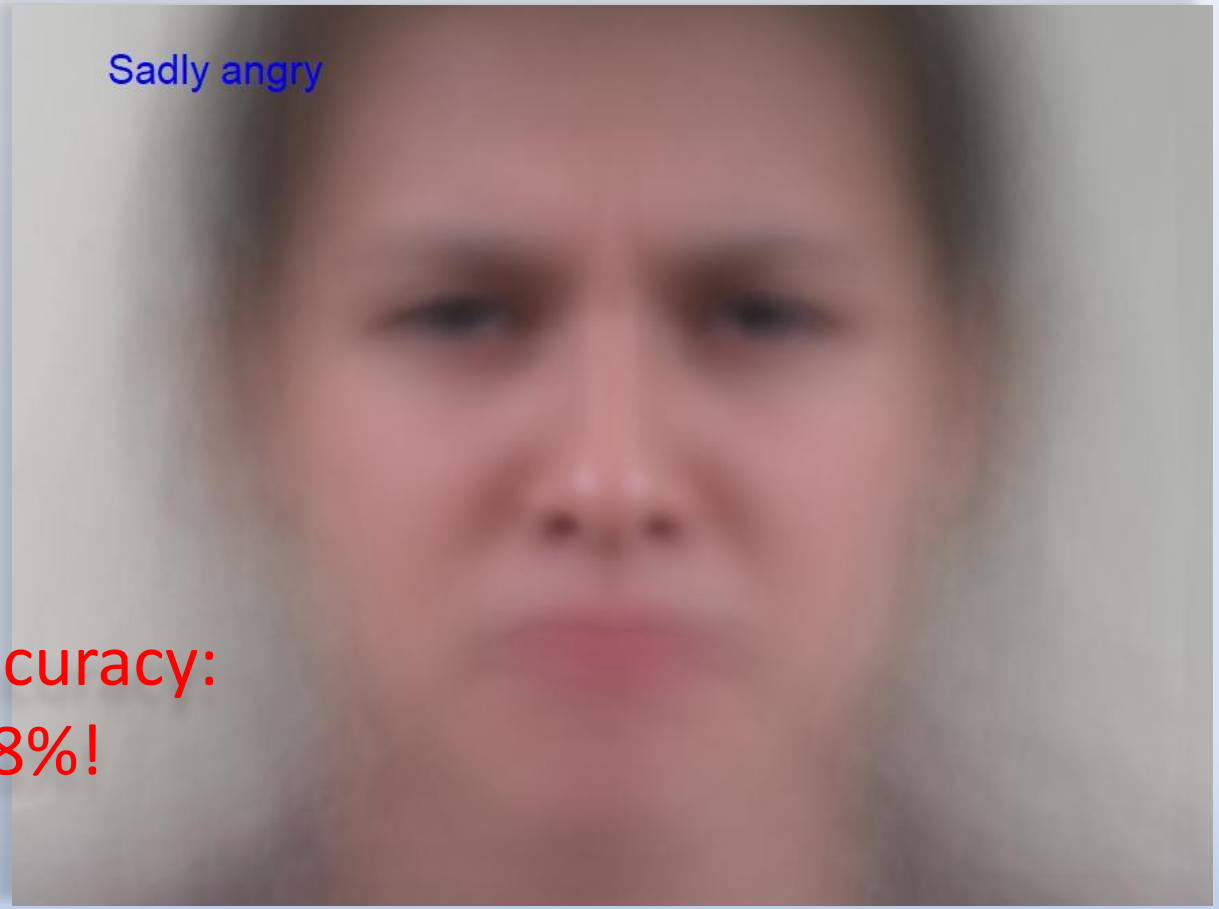
Conclusion - 5

The two images are slightly different near the eyes, and the darkness of the nasolabial folds is different.
But I'm not sure if I can correctly classify by staring them.

What do these confusing classifications look like?

‘Angry’ vs ‘Sadly angry’





SVM accuracy:
75.8%!

Conclusion - 6

Computer configuration:
System: Microsoft Windows 10 x64 | GPU: NVIDIA GeForce RTX 2070 SUPER (8192 MB)
CPU: Intel(R) Core(TM) i7-9700K CPU @ 3.60GHz(3600 MHz) | RAM: 16.00 GB (2400 MHz)

Prediction of in-class test

We choose color features and SVM classifier to make predictions. The test time of about 500 test sets is 40s, so we predict that the test time of 2500 test sets is 200s. Under the same computer configuration as above, the calculation time is 200s. If you use a MacBook air, the time will be doubled to 400s.

Since the test accuracy is related to the selection of the training and test sets, it is difficult to make accurate prediction. After comprehensive consideration, we believe that the accuracy of the in-class test will be around 58%.


THANK YOU