Practical Data Mining

Face Recognition Project

Name: James/Peng Li

Date: 03/06/2024

Contents

ABSTRACT

Face recognition is a popular topic and has been studied and researched by many pioneers in this area and even today it is still improved. Along with development of the technology for this application, we also find many applications in bank, security areas etc. High speed and high accuracy and also high security is very important for this application. As a student of practical data mining course, I have learned representative models for analysis and prediction. For this topic will use different models to compare the different recognition results and try different ways to improve this recognition results even the accuracy is not high. Through these trails, either succeed or failed, I have got deeper understanding of disadvantage and advantages of each model. And also got better understanding of the parameters and other preconditions such like data status and quality, which will influence the recognition results. The study is just the first step with basic model settings to have the primary results. More deeper study and adjustment is needed in the future for better applying of these models, not limit to this application, but also for other applications and topics. In general for this topic the raw data standardization can help to improve the data quality and achieve higher recognition accuracy with decision tree and logistic regression. PCA also helps to capture the most variance and increase the operation speed with less data to be dealt with in the models. And we also tried to optimize the defined PCA component number and also find its influence. This trail works for logistic regression. For decision tree we also used pruning to improve the recognition results. The best model is neural network. We tried to prepare the data and defined the parameters for this model. Relative higher recognition accuracy has been achieved with neural network. As the most advanced and also most difficult to be understood model, neural network which is used in this study still needs to be further improved. As we all know there are also further developed neural network which can have better results in this area. In this study, we used the fundamental one for primary study.

1. Introduction

1.1 Main target

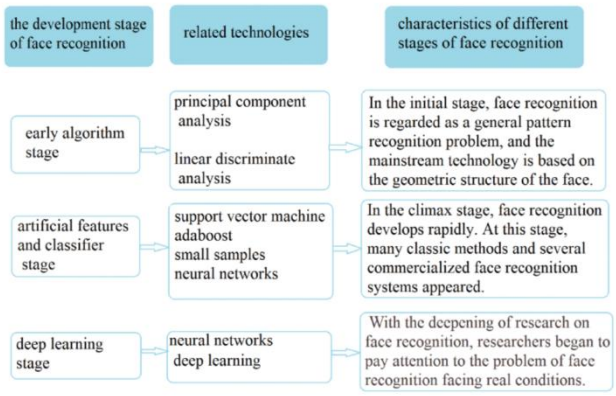
Beyond lecture material and pedagogical examples given during the course, this project aims to practice these knowledge for real project. In this project based on the data provided, we will use different models to deal with this data and analyze different results out of these models. Though this practice to know more deeper of advantage and disadvantages of each model. Last not least, learn to improve the recognition results through different trails, for deeper understanding of these analysis tools, such like PCA, decision tree, logistic regression, neural network etc.

1.2 Face recognition development status and challenges

Face recognition technology is a biometric technology, which is based on the identification of facial features of a person. People collect the face images, and the recognition equipment automatically processes the images[1]. Face recognition is a hot topic and was studied and researched deeply and widely in the past. One main reason is its big commercial application potential. For security system or bank software, now face recognition is widely used. How to improve the recognition accuracy with high speed. This is always one challenge.

If we look back into the history of face recognition technology, it appeared already in 1980s. At that time only some basic research due to the technology limit and the recognition accuracy is also very low. So this technology is only applied in certain filed such like monitoring, access control and etc.

When the algorithm is continuously improved, face recognition technology has also been quickly developed. Around 1990, many advanced algorithms have been developed, such as feature-based algorithms, neural networks, etc. These new algorithms has greatly improved the accuracy of face recognition technology. Blow is the development stage of face recognition, related technologies and characteristics of different stages of face recognition[1].



Currently face recognition has been widely used. Many commercial companies began to apply face recognition technology for payment and other fields according to the new application requirements. Below is the one general summary for face recognition applications[2].

Category	Exemplar application scenarios
Face ID	Driver licenses, entitlement programs, immigration, national ID, passports, voter registration, welfare registration
Access control	Border-crossing control, facility access, vehicle access, smart kiosk and ATM, computer access, computer program access, computer network access, online program access, online transactions access, long distance learning access, online examinations access, online database access
Security	Terrorist alert, secure flight boarding systems, stadium audience scanning, computer security, computer application security, database security, file encryption, intranet security, Internet security, medical records, secure trading terminals
Surveillance	Advanced video surveillance, nuclear plant surveillance, park surveillance, neighborhood watch, power grid surveillance, CCTV control, portal control
Smart cards	Stored value security, user authentication
Law enforcement	Crime stopping and suspect alert, shoplifter recognition, suspect tracking and investigation, suspect background check, identifying cheats and casino undesirables, post-event analysis, welfare fraud, criminal face retrieval and recognition
Face databases	Face indexing and retrieval, automatic face labeling, face classification
Multimedia management	Face-based search, face-based video segmentation and summarization, event detection
Human computer interaction (HCI)	Interactive gaming, proactive computing
Others	Antique photo verification, very low bit-rate image & video transmission, etc.

Recent one new application is from Apple. It uses 3D camera tech to power the thermal infrared-based face ID feature in its iPhone X. Thermal IR imagery maps the patterns of faces derived primarily from the pattern of superficial blood vessels under the skin. Apple also sends the captured face pattern to a ‘secure enclave’ in the device. This ensures the authentication happens locally and that the patterns are not accessible by Apple[3].

On the one hand we see quick development of the face recognition technology along with its wide applications. On the other hand, face recognition is also facing many challenges. The biggest one is safety and faked face information. How to protect our face data and how to avoid that our face information will not be misused for cheating etc. All these lead to higher requirement for this technology. Especially when the generative AI is now so popular as also powerful to automatically generate information which looks so real. So face recognition faces even higher challenge, not just recognition accuracy, but also security.

But for this project, we will primary try the technology to increase the accuracy.

2. Data set description and overall plan for face recognition

2.1 Data set description

Instead of directly utilizing the facial images, we leverage an excel file prepared by our teacher. This file contains crucial information about the facial data.

In total we have 19 celebrities. For each celebrity we have different number of images, statistic as below.:

celeb	
Robert Downey Jr	223
Alexandra Daddario	215
margot robbie	211
amber heard	208
Emma Watson	201
Anne Hathaway	193
scarlett johansson	191
Andy Samberg	186
Henry Cavil	185
Tom Holland	179
Zoe Saldana	176
Jason Momoa	174
Hugh Jackman	169
Chris Pratt	166
Chris Hemsworth	149
Dwayne Johnson	131
Emma Stone	129
Zendaya	128
Taylor Swift	121

And these images have been converted three different pixels (20x20), (50x50), (100x100), with three Excel files. Each row in the Excel file represents a unique data point, with features corresponding to pixel values representing the visual attributes of the facial images. So for pixels 20x20 in one row we have 400 data, for pixels 50x50 in one row we have 2,500 data, for pixels 100x100 in one row we have 10,000 data.

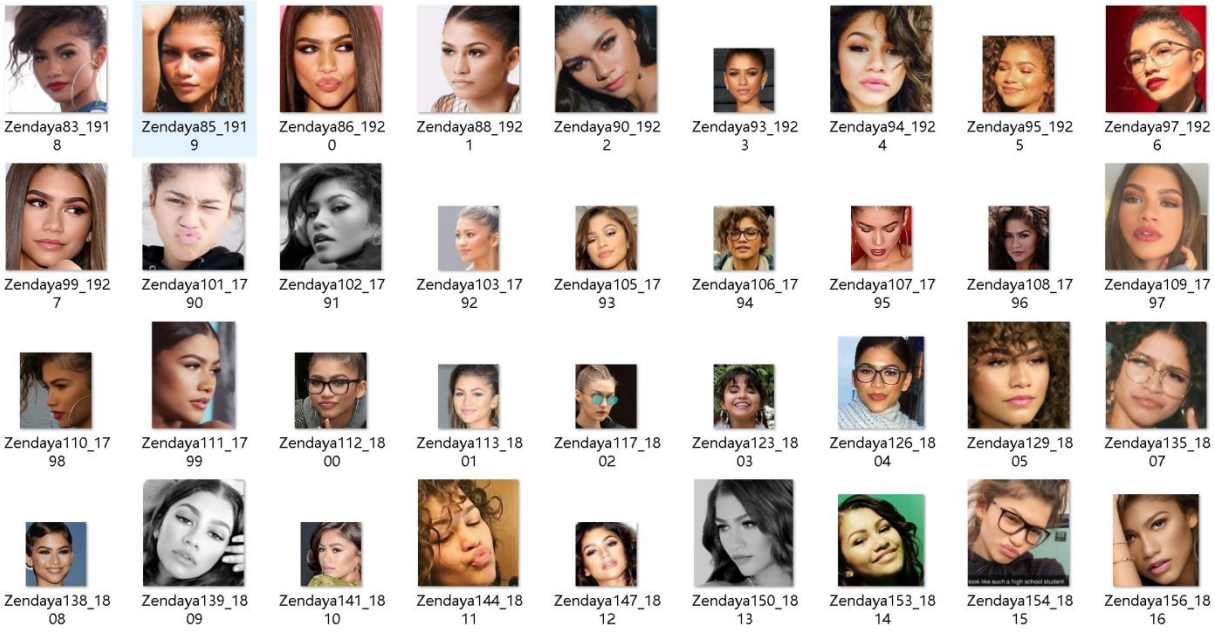
Additionally, each data set includes a column designated for target labels, where each entry signifies the identity of the celebrity depicted in the respective image. Part of the data statistic please see below:

Statistical Summary:

	r1c1	r1c2	r1c3	r1c4	r1c5
count	3335.000000	3335.000000	3335.000000	3335.000000	3335.000000
mean	0.468713	0.465971	0.462059	0.459661	0.455295
std	0.320964	0.320200	0.319412	0.319089	0.317978
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.170647	0.173843	0.169725	0.168510	0.163588
50%	0.420627	0.424235	0.408000	0.408314	0.405529
75%	0.777725	0.773196	0.770490	0.766314	0.760843
max	1.000000	1.000000	1.000000	1.000000	1.000000

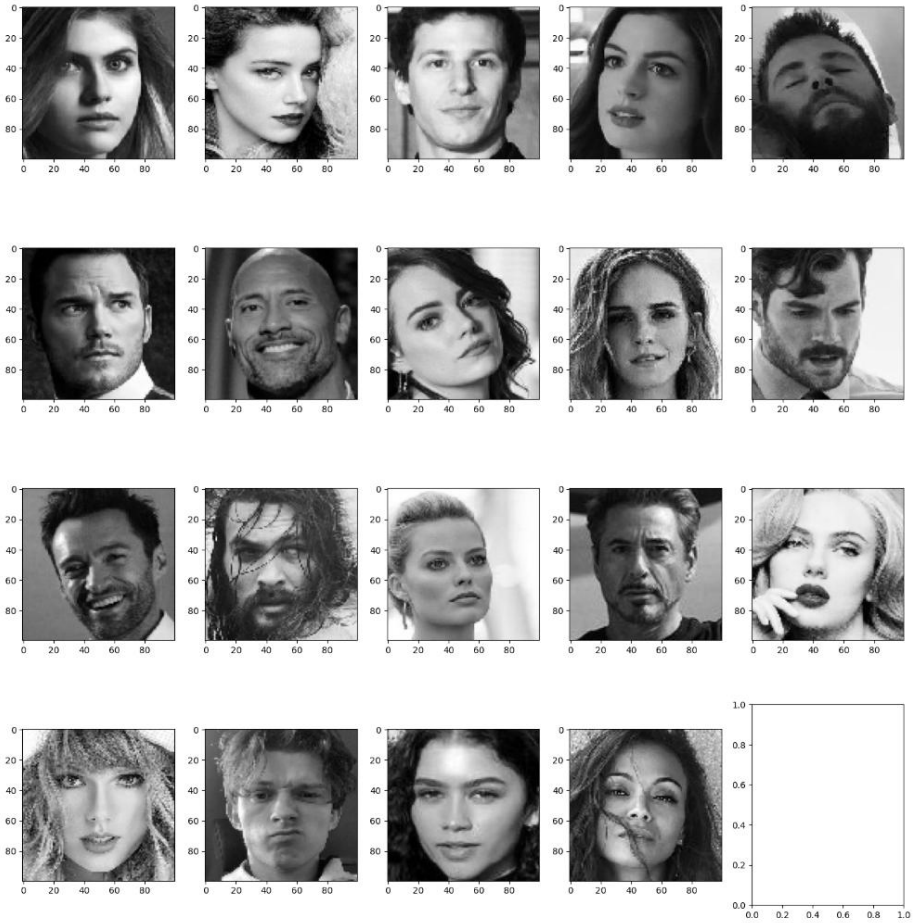
From above chart, we can find that the all the data are between 0 and 1. And no missing data. So the data is well organized and prepared.

If we check the original images of each celebrity, we can find that many images are quite different from each other. Let's take "Zendaya" as an example in below. We can find that these images are taken under different illumination, with different hair styles or even with glasses, varied pose angles, different emotions accompanied with different mouse or eyes shapes, and also variant backgrounds etc. All these lead to big challenges for face recognition.



In this study the data we got is already the transformed data from original images. If we recover the images from the pixels and grayness. We can find the first image of each celebrity as below. I think the special setting during the data transformation from original images to data and elimination of the colors can help solve the challenges listed above, about the diversity and big difference of the original images for same celebrity. This way reduces computational complexity, computing resources and processing time. It also eliminates color interference to capture main facial features and structures, and also avoid interference caused by changes of lighting conditions, skin color or background color. This also helps for algorithm stability which is easier to normalize and make the recognition algorithm more robust to illumination changes and simplify the data preprocessing for model training etc.

In general, this way of data transformation from images can improve the efficiency and robustness of the algorithm, while reducing the computational complexity, which is a practical and effective method.



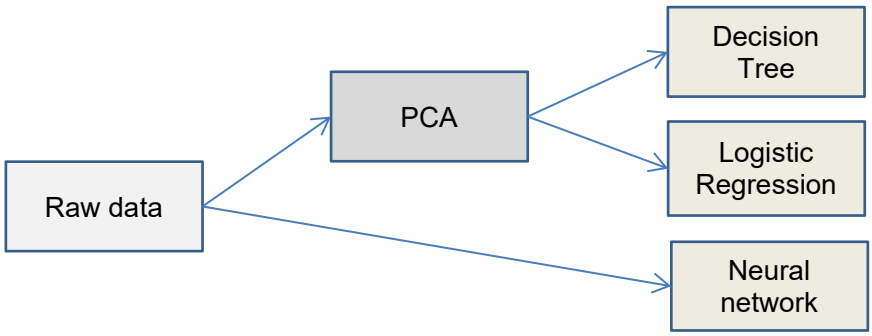
2.2 Overall plan for face recognition

For pixels 20x20 we have 400 variables, for pixels 50x50 we have 2,500 variables, and for pixels 100x100 we have 10,000 variables. So our first target is to decrease the dimensions.

- Firstly we will use PCA to capture the main features. The main reason is to reduce the dimension and this also saves the operation time, thinking that for 100x100, we have 10,000 data for one image. If we uses all the data, it will take long time to recognize the image. Of course, after using PCA, we will lose some features and the recognition accuracy will be influenced a bit. But thinking of application requirement, we need also to have high recognition speed during real application circumstance. So we will PCA first for data preparation. Based the PCA results, we will use three different ways for analysis: decision tree, logistic regression and neural network.
- In general decision maybe not a good way or first way for face recognition if we want to achieve high accuracy. But we would like also try to check its results comparing to other two ways.
- Logistic regression could be a better way, as according to its methodology, it will transform the data into 0 and 1, and the original data already locates in this range. So the logistic regression is considered to a be suitable way for face recognition.

- Neural network is the currently primarily considered to be the most suitable technology for face recognition. It has very good historical results for this area. We will try to build up the basic model and try to find the suitable parameters to get a good result. As it is so complex, so for this study we are just trying the basic model for better understanding its methodology.

This is overall primary plan, we will adjust it based on our exploring findings. For decision tree and logistic regression, we will based the PCA analysis results with reduced dimensions. And for the neural networks, we will directly use the original data for analysis. And we will use Python program for this project.



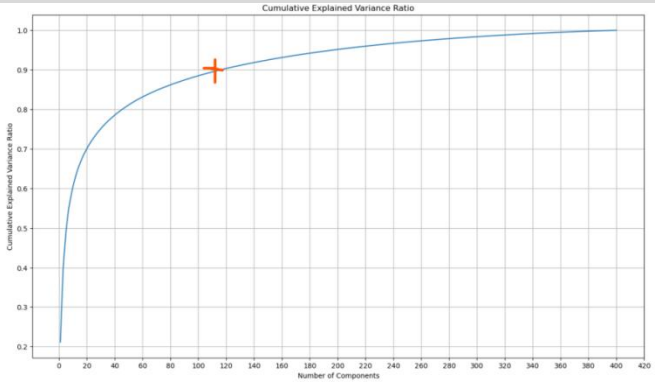
3. PCA

Principal component analysis (PCA) is a linear dimensionality reduction technique with applications in exploratory data analysis, visualization and data preprocessing. The data is linearly transformed onto a new coordinate system such that the directions (principal components) capturing the largest variation in the data can be easily identified[4]. Through this way we can capture the main feature and reduce the dimension for next step data analysis.

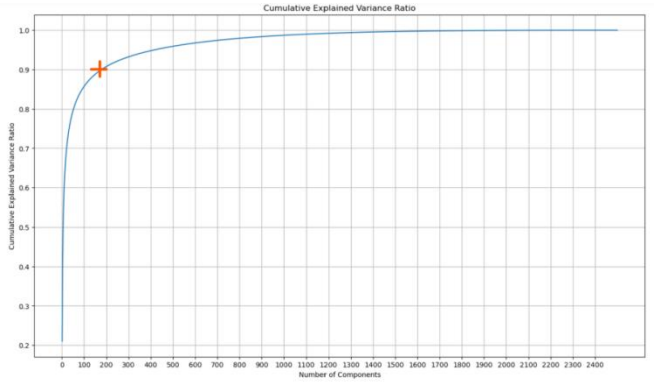
Step 1: Split the three Excel files (20x20), (50x50), (100x100) into two parts: training data 80% and validation data 20%. We tried to keep more data for training to achieve high quality trained model.

Step 2: for the training data we used PCA analysis to define the PCA components number to reduce the dimension.

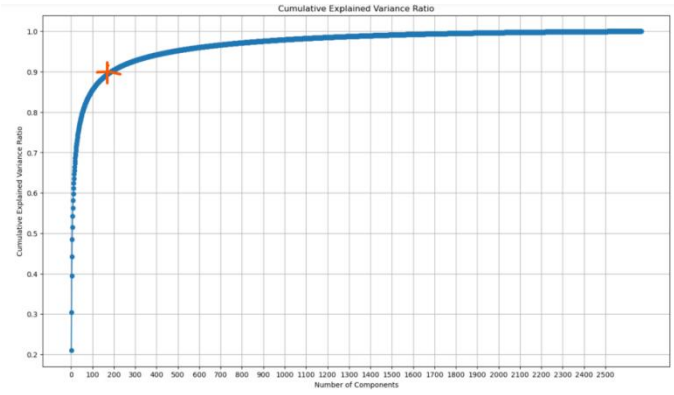
- 20x20 pixel results as below. Based on the principle to capture 90% of the variance, we chose 110 PCA components.



- 50x50 pixel results as below. Based on the principle to capture 90% of the variance, we roughly chose 160 PCA components.



- 100x100 pixel results as below. Based on the principle to capture 90% of the variance, we also roughly chose 160 PCA components.



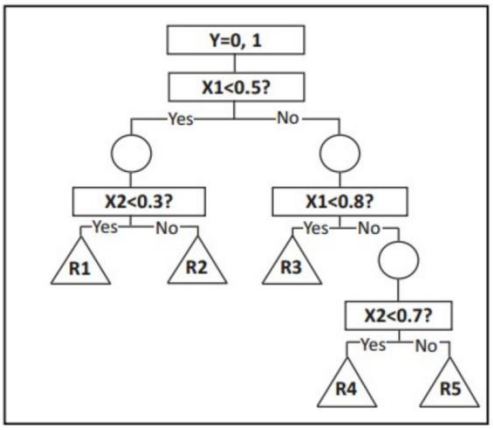
So through this step, we successfully reduce the dimension to simplify the access.

The images can be constructed from the specification of gray levels at pixel locations. For 20x20 pixel in the python code, we used imshow function to reconstruct the images. Through below images, we can find that the first few images already captured the main features and further PCA components just add more details into it. Similar results for 50x50 pixel and 100x100 pixel.

4. Decision Tree

A decision tree is a decision support hierarchical model that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tool in machine learning[5].

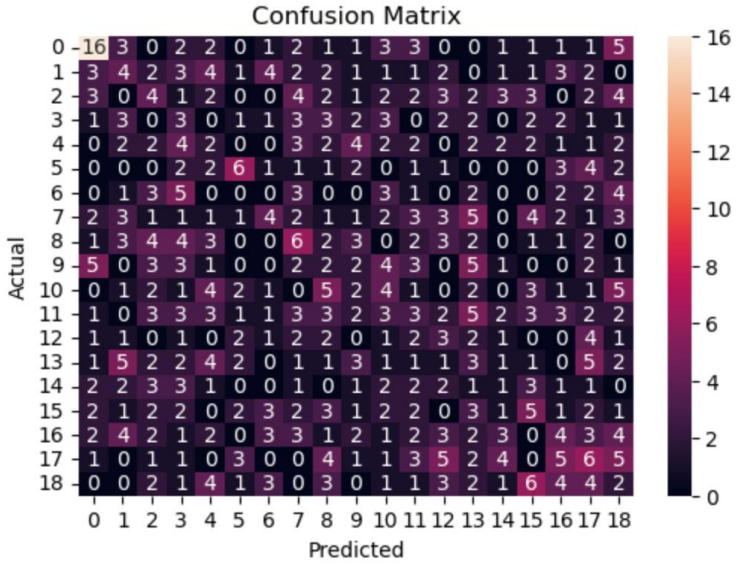
Decision tree methodology is a commonly used data mining method for establishing classification systems based on multiple covariates or for developing prediction algorithms for a target variable. This method classifies a population into branch-like segments that construct an inverted tree with a root node, internal nodes, and leaf nodes[6].



For decision tree, we will based on the PCA analysis results to run the decision tree analysis. Below is the flowchart:

Original results: With decision tree model for the test data, we got below results:

	20*20	50*50	100*100
Original data	0.13	0.13	0.11
PCA components	110	160	160

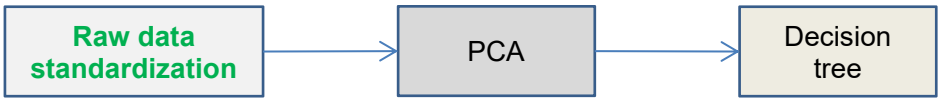


In the confusion table, we can find different prediction results for different celebrity.

	precision	recall	f1-score	support
Alexandra Daddario	0.31	0.30	0.31	43
Andy Samberg	0.25	0.19	0.22	37
Anne Hathaway	0.21	0.18	0.19	38
Chris Hemsworth	0.05	0.07	0.06	30
Chris Pratt	0.07	0.06	0.07	33
Dwayne Johnson	0.19	0.27	0.22	26
Emma Stone	0.04	0.04	0.04	26
Emma Watson	0.00	0.00	0.00	40
Henry Cavil	0.07	0.08	0.07	37
Hugh Jackman	0.10	0.09	0.09	34
Jason Momoa	0.21	0.23	0.22	35
Robert Downey Jr	0.10	0.09	0.09	45
Taylor Swift	0.20	0.17	0.18	24
Tom Holland	0.15	0.17	0.16	36
Zendaya	0.00	0.00	0.00	26
Zoe Saldana	0.11	0.11	0.11	35
amber heard	0.11	0.10	0.10	42
margot robbie	0.14	0.14	0.14	42
scarlett johansson	0.17	0.21	0.19	38
accuracy			0.13	667
macro avg	0.13	0.13	0.13	667
weighted avg	0.13	0.13	0.13	667

From above table, we can even more clearly to see the different prediction result for different celebrities. The highest one is 0.31 for Alexandra Daddario, and the lowest is 0 for Emma Stone and Zendaya. So there are difference of the prediction results for different celebrities. Essentially this is depends on the original images' features. If we check the original images for each celebrity, we can find that for some celebrity, based our human recognition experience, there many quite different from each other, based on face light, the ankle of faces, hair style, with glasses or not, clearness. If the images are not identical, it will be hard to recognize them and may easily lead to misclassification.

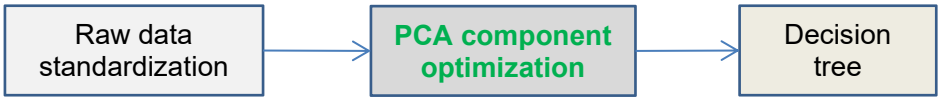
Improvement trial 1: original data standardization. As explained before, the original data already locate between 0 and 1. With standardization, we wish to further increase the data quality for data preparation. As we all know, data quality is the fundamental basis for data analysis, so we wish to through this way to increase the quality and check if it improves the recognition results or not. After that we found that there is nearly no big change of the PCA components curve, so we keep the same PCA components number for comparing the improvement results.



	20*20	50*50	100*100
Original data	0.12	0.12	0.12
Standard data	**0.15**	**0.13**	**0.13**
PCA components	110	160	160

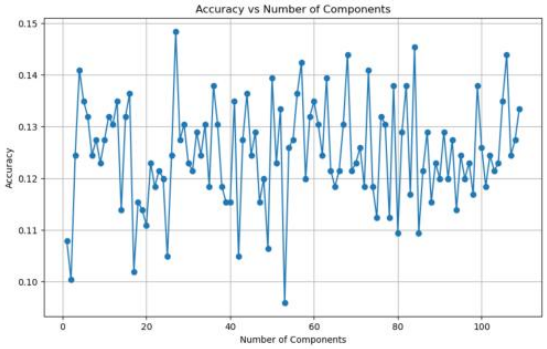
With the standardized data and based on the PCA analysis results, we ran the decision tree again, and for the three data sets, we all see good improvements. For 20x20, it has big increase from 0.12 to 0.15. For 50x50 and 100x100, it also some increase from 0.12 to 0.13. So the improvement with standardization works. For 20x20, this is already 25% increase of accuracy. So this is somehow good result.

Improvement trial 2: original data standardization + optimal PCA components number. In PCA part, we defined the optimal PCA number according to the principle to capture 90% of the variance. But is this really optimal? We will test different kinds of PCA Components number based on the standardized data, to the different prediction accuracy with logistic regression model.



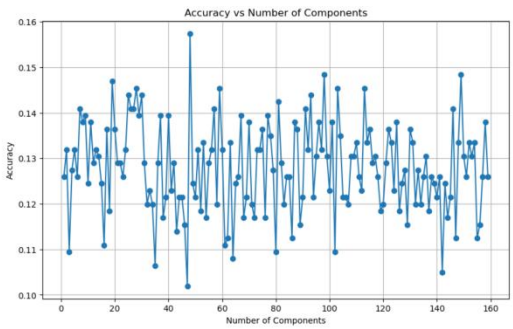
With the standardized data, below is the logistic regression results accuracy based on different PCA component number from 0 to 110.

20x20 accuracy with different PCA number

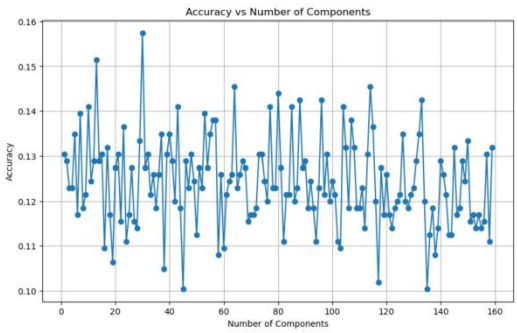


For 20x20, we can not see clear trend after changing the PCA components number. There is a big variation up and down. What we can learn from here is that, PCA number do influence the prediction results. Based on this we will keep the original PCA component number. So this trial does not work.

50x50 accuracy with different PCA number



100x100 accuracy with different PCA number



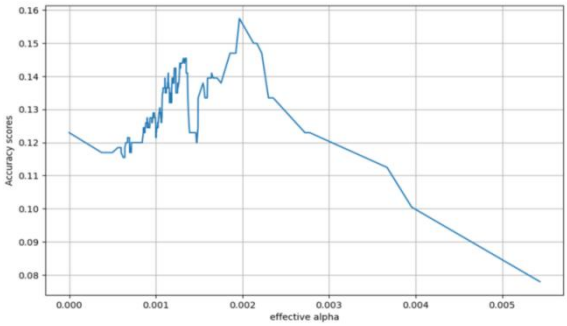
Improvement trial 3: original data standardization + PCA (component number according to capture 90% variance)+Pruning.

For this part, firstly we will compute the cost-complexity pruning path and extract the alphas and impurities for next step.

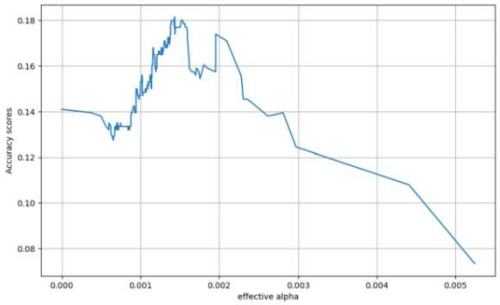
Secondly, with different alpha values we will train and prune the decision tree. We will valuate and select the best pruned decision tree.

Thirdly, we will evaluate the accuracy of the pruned decision trees on the test set.

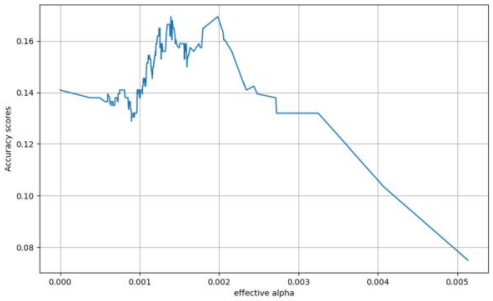
20x20 accuracy after pruning



50x50 accuracy after pruning



100x100 accuracy after pruning



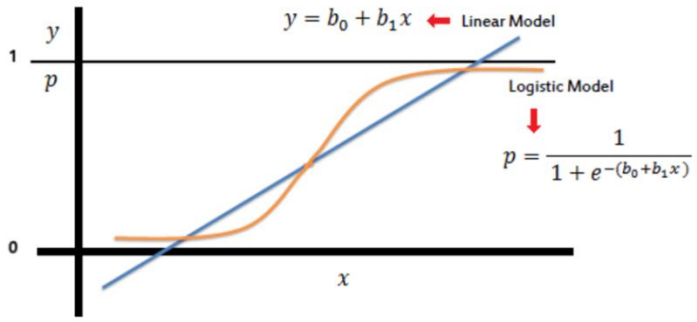
	20*20	50*50	100*100
Original data	0.12	0.12	0.12
Standard data	**0.15**	**0.13**	**0.13**
Standard data+Pruning	**0.16**	**0.18**	**0.17**
PCA components	110	160	160

After the pruning, we can find the further improvement of the recognition accuracy. Relative good improvements for 50x50 and 100x100 data sets. So the pruning step works.

5. Logistic Regression

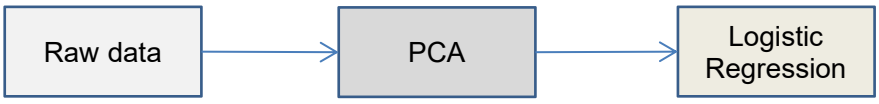
"Logit model" redirects here. Not to be confused with Logit function. In statistics, the logistic model (or logit model) is a statistical model that models the log-odds of an event as a linear combination of one or more independent variables. In regression analysis, logistic regression (or logit regression) is estimating the parameters of a logistic model (the coefficients in the linear combination)[7].

We use the log to transform below linear model into logistic model, to locate the y between 0 and 1.



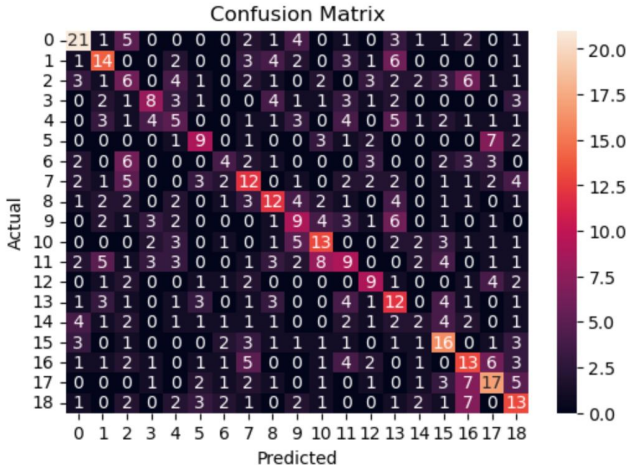
The advantage of logistic regression is that it is simple and interpretable. Performs well with linear relationships. Provides probabilistic outputs. Disadvantage is that it is limited to linear relationships. Sensitive to outliers and multicollinearity. May underperform with complex data.

We will use the logistic regression model based on the PCA analysis results.



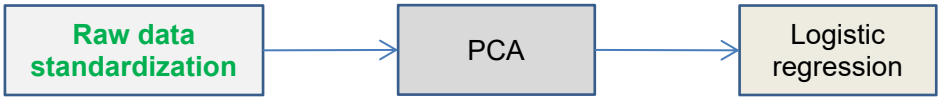
Original results: With logistic regression model for the 20% test data, we got below results:

	20*20	50*50	100*100
Original data	0.29	0.31	0.31
PCA components	110	160	160



In the confusion table, we can find different prediction results for different celebrity.

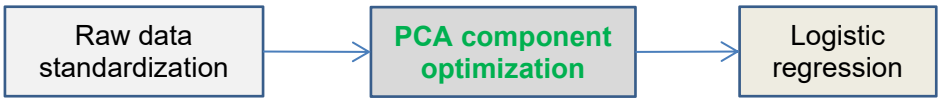
Improvement trial 1: original data standardization. We will try to standardize the original data, and after that we found that there is nearly no big change of the PCA components curve, so we keep the same PCA components number for comparing the improvement results.



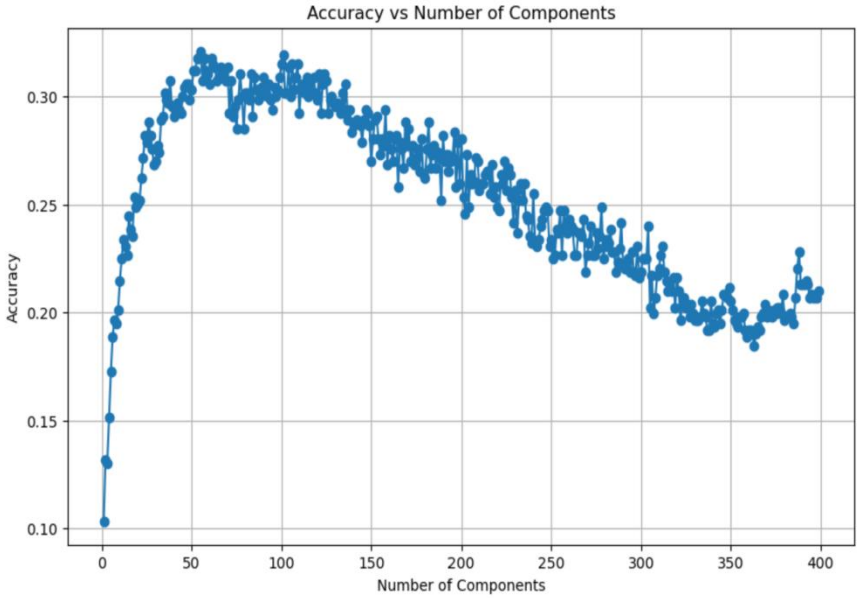
	20*20	50*50	100*100
Original data	0.29	0.31	0.31
Standard data	**0.31**	**0.31**	**0.31**
PCA components	110	160	160

For three data sets, we only see relative good improvement for 20x20 pixel data. The reason is that for the logistic regression, we will transform the original data into 0 and 1, so the standardization will not help much for improvement of logic regression results.

Improvement trial 2: original data standardization + optimal PCA components number. In PCA part, we defined the optimal PCA number according to the principle to capture 90% of the variance. But is this really optimal? We will test different kinds of PCA Components number based on the standardized data, to the different prediction accuracy with logistic regression model.

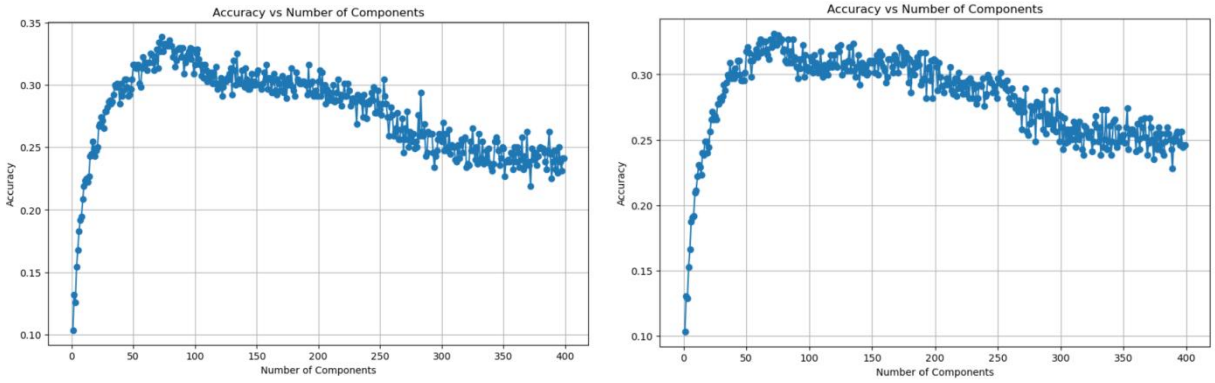


Below is the logistic regression results based on different PCA component number from 0 to 400:



The program also automatically output the highest accuracy 0.32 with 55 PCA components. Compared to 0.31 with 110 PCA components, it is not a big increase. But it is interesting to find the reduced PCA components number in this way. Of course, even 55 components number is theoretically the optimal number, but it is better to assume that the range between 50 and 125 are all high accuracy range with above 0.3 accuracy. So we can reduce the PCA components number significantly nearly to half of the original chosen one. Last but not least, we can find the clear trend of accuracy increase according to the increase of PCA components number and then decrease after certain number of PCA components. The peak area is also obvious. This is different from decision for the same trail. In decision tree we can find big variation. Here there is also some changes, but not so big. Overall we can see the clear pattern of this curve. So the result can be accepted.

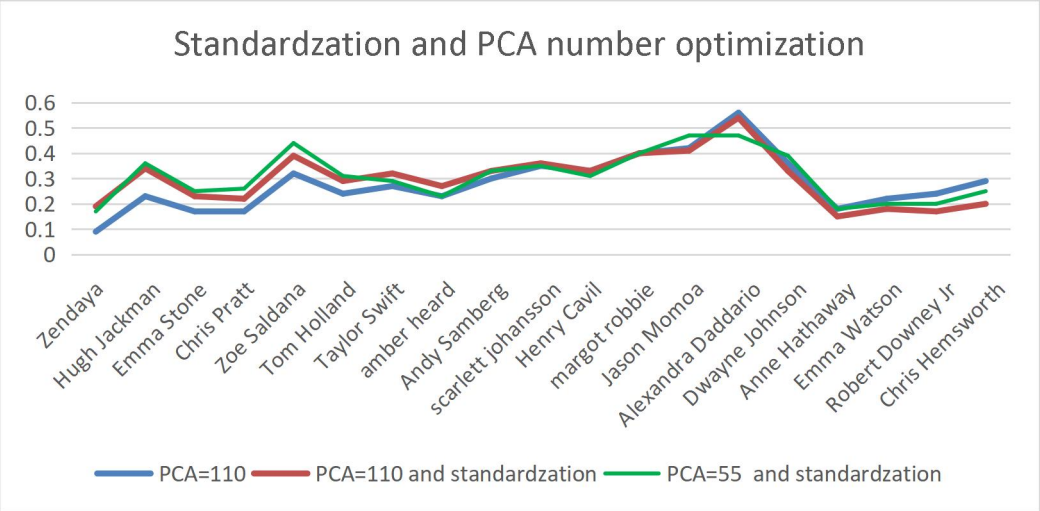
For 50x50 and 100x100 data set, we can also got similar results.



And the summary is as below, we can find relative good improvement for 50x50 data set.

	20px	50px	100px
Original data	0.29(110)	0.31(160)	0.31(160)
Standardized data	0.31(110)	0.31(160)	0.31(160)
Adjusted PCA number	**0.32(55)**	**0.34(73)**	**0.33(72)**

Further study: from the confusion table for 20x20 above we can find that there is different prediction results for different celebrities. So we wish to know what is the contribution for different celebrities.



From above chart for 20x20 pixel, we can find that:

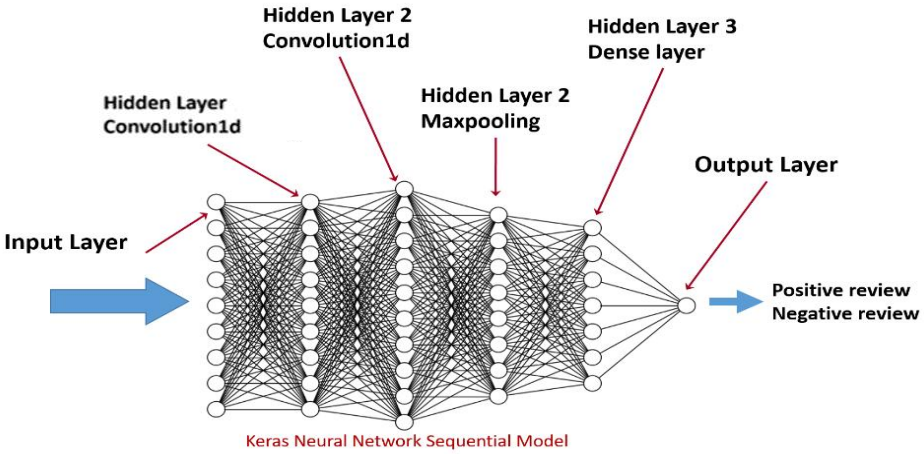
- After standardization, there is clear improvement for most of the celebrities, especially about Zendaya with around 100% and Hugh Jackman with around 40%. So standardization improves their data quality hugely. On contrary, we also see negative influence for some celebrity, such like Chris Hemsworth with around 30% accuracy reduction.
- After reduction the PCA components number with the optimal one. We can also find different impact on the prediction accuracy of different celebrities.

So in general, if we take each celebrity as a separate sub data set, we can find that standardization and optimal PCA component number does not have the consistent impact on each celebrity. This is depends on the data difference, and fundamentally depends on the photos of different celebrities. But overall we see the improvement through the two trails.

6. Neural Network

A neural network is a group of interconnected units called neurons that send signals to one another. Neurons can be either biological cells or mathematical models. While individual neurons are simple, many of them together in a network can perform complex tasks[8].

Sequential model will be used. In a Sequential model, data flows through each layer in a fixed order, starting from the input layer and ending with the output layer. Each layer can perform a different type of transformation on the data, such as convolutions, pooling, dropout, or dense connections. By stacking these layers together, a sequential model can learn complex patterns and representations from the data, making it a powerful tool for many different types of machine learning tasks[8].



Keras Neural Network Sequential Model

For neural network, we will follow below steps:

We will not have the PCA procedure, but we need to prepare and transfer the original data for neural network analysis.

Data preparation:

Neural networks, especially with libraries like Keras, typically it requires input data in the form of NumPy arrays. This conversion from pandas DataFrame to NumPy array is a necessary for neural network. And convolutional Neural Networks need the input data to be in a specific shape.

- Convert the training and validation data from a pandas DataFrame format to a NumPy array.
- Reshape the data into the format which is suitable for input into the CNN.
- Additionally also replace the label of celebrity names with unique label number 0 to 18.

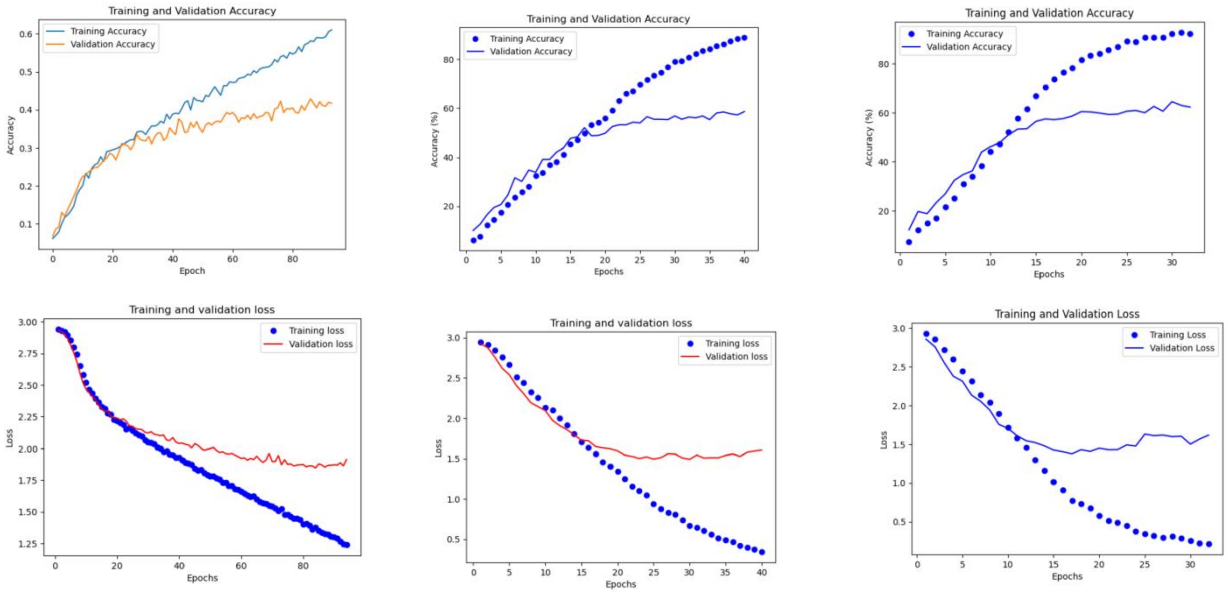
Neural network:

Based on the prepared data, we can use the neural network for face recognition.

- Firstly we need to define and construct a Convolutional Neural Network (CNN) model using Keras. Keras is a popular deep learning framework. We tried and defined the parameters inside this function. Such like Conv2D Layer (1st Layer), Conv2D Layer (2nd Layer), Flatten Layer, Dense Layer (1st Fully Connected Layer), Dropout Layer (1st Dropout Layer), Dense Layer (2nd Fully Connected Layer), Dropout Layer (2nd Dropout Layer), Dense Layer (Output Layer).
- Secondly, we compiled the neural network model. This an important step before training the model. We set the learning rate with 0.0001, which will help the model learn and update weights effectively during the training.
- Thirdly, we begin to train the neural network model on the training data. We set the epochs=150. This defines how many times the entire training dataset will pass through the neural network during the training. A higher number of epochs means the model will have more opportunities to learn from the data, but it also increases the risk of overfitting. We set call back as early_stopping. It is used to stop training early if the model's performance on the validation set stops improving. It helps prevent the overfitting. And it can save time and keep the efficiency.

Based these steps, we got results as below, the accuracy for training data and validation data:

20x20 **50x50** **100x100**



	20px	50px	100px
Training accuracy	0.59	0.91	0.92
Validation accuracy	0.42	0.63	0.62

So we can find that with the neural neural network can achieve the highest accuracy. The rigid up and down movement of the validation accuracy is affected by the dropout parameters since not all data was used after every epoch cycle. When Training Accuracy is high, Training Loss tends to go low.

7. Summary

- For the raw data, we tried to have standardization for improving the data quality. In general this works for decision tree and logistic regression. For neural network, we did not standardize the raw data.
- PCA step is also used for decision tree and logistic regression. This step helps to reduce the dimension and increase the recognition speed. For the neural network, we did not use PCA step, but directly analyze the raw data.
- Decision tree is easy to understand and intuitive. It holds no assumptions of the data and it's insensitive to the distribution of the raw data. Therefore, PCA and other preprocessing method didn't help much. It's also not a good model for distinguishing faces because the pixel variables are too complicated for the model. It can't catch the relationship effectively. So we the recognition accuracy is under 0.2 even after pruning.
- Logistic regression has a relatively good performance in the traditional method. It's simple to understand and interpret. As a classification method, it benefits from PCA and standardization methods. PCA and standardization increases the efficiency and simplifies the subsequent modeling. Optimal component iteration provides a range of number for choosing the right number of components in traditional machine learning. However, it's more compatible with data with linear relationship but pixel number of images is not the case. So the highest accuracy is only around 0.34 with some trails of improvements.
- Unlike traditional machine learning, NN model apply filters upon images itself. It captures spatial relationships and automatically learns hierarchical features. Therefore, PCA and standardization didn't do much for the NN model. And unlike traditional method, the higher resolution actually help on the accuracy. So we can achieve relative higher accuracy with more the 0.6 for the validation data.

In general, through this study I practiced the models we learned in the practical data mining. And also used these models for a popular topic the face recognition. Although the recognition accuracy with decision tree and logistic regression is not so high, this also help me to better understand these models and their limits. Neural network is the newest and most advanced model and is also the most used tool for this topic the face recognition. As we all all know even for neural network, there are many improved and highly trained models which are far more powerful than the model we used in this study. I did tried to adjust some parameters of this model to get good result, but this is still very basic. So we can find this from the recognition accuracy which is not so high. The highly trained neural network model can achieve much higher accuracy. After rigorous analysis of various state-of-art techniques it is reported that the highest accuracy rate of 99.5% is achieved on AR database using wavelet based feature extraction whereas 99.78% is attained using convolutional neural network accuracy[9]. This is far more higher accuracy than we can achieve in this study.

Through these trials, I got the first impression of applications of these models. Because of time limit and capability, I can not get a very high recognition accuracy for this study currently. So this study is very basic based on the methodologies I have learned and the models in this study still need to be further improved for better results.

Reference:

[1] *A review of face recognition technology*. (2020). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/9145558

[2] Li, S. Z., Jain, A. K., Huang, T., Xiong, Z., & Zhang, Z. (2005). Face recognition applications. *Handbook of Face Recognition*, 371-390. https://link.springer.com/content/pdf/10.1007/0-387-27257-7_17.pdf

[3] *Facial recognition history.* (2023, June 12). Thales Group. https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/inspired/history-of-facial-recognition

[4] Wikipedia contributors. (2024, May 16). *Principal component analysis*. Wikipedia. https://en.wikipedia.org/wiki/Principal_component_analysis

[5] Wikipedia contributors. (2024, May 20). *Decision tree*.Wikipedia. https://en.wikipedia.org/wiki/Decision_tree

[6] Decision tree methods: applications for classification and prediction. (2015). PubMed. https://doi.org/10.11919/j.issn.1002-0829.215044

[7] Wikipedia contributors. (2024, May 20). *Logistic regression.* Wikipedia. https://en.wikipedia.org/wiki/Logistic_regression

[8] Wikipedia contributors. (2024, March 26). *Neural network.* Wikipedia. https://en.wikipedia.org/wiki/Neural_network

[9] Dalal, T., & Yadav, J. (2024). TIFd-FR: Trends, Issues and Future directions of feature extraction in Face Recognition. *Procedia Computer Science*, 235, 1386–1398. https://doi.org/10.1016/j.procs.2024.04.130