Title	Research on face recognition method for beauty industry
Sub Title	
Author	フォン, シュエン チー(Feng, Xuanqi) 小木, 哲朗(Ogi, Tetsuro)
Publisher	慶應義塾大学大学院システムデザイン・マネジメント研究科
Publication year	2023
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2023年度システムエンジニアリング学 第355号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40002001-00002023-0007

Master’s Dissertation 2023

Research on Face Recognition Method for Beauty Industry

Xuanqi Feng
（Student ID Number：82134566）

Supervisor　Tetsuro Ogi

September 2023

Graduate School of System Design and Management,
Keio University
Major in System Design and Management

SUMMARY OF MASTER'S DISSERTATION

Student Identification Number	82134566	Name	Xuanqi Feng

Title

Research on Face Recognition Method for Beauty Industry

Abstract

Pursuing the perception of beauty is a natural instinct for human beings. Makeup, haircuts, and outfits can make people look more exquisite and attractive. Choosing a suitable fashion style is a form of art, and face types play an important role that is closely related to people's style choices.

The importance of human appearance lies in giving people a favorable impression, enhancing positive feelings, and increasing charm. This study aims to meticulously classify people's face types based on their facial features and quickly identify a person's face type, enabling people to understand how to wear makeup and choose outfits to enhance their glamour value.

The paper proposes a 16-face types classification method as the formula for achieving beauty. The study employs machine learning (ML), including data acquisition, pre-processing, feature extraction, and classification, to automatically identify face types among sixteen categories and assess accuracy using different samples.

To design the supervised machine learning system, this study utilizes the vision transformer (ViT) model for feature extraction and the classical classifier random forest algorithm (RF) for classification. Additionally, the results are compared with other classifiers such as support vector machine (SVM), decision tree (DT), Adaboost, and k-nearest neighbor (k-NN).

Keywords (5 words)

Face Style, Face Recognition, Machine Learning, Beauty, Fashion

CONTENTS

List of Figures

List of Tables

1 Introduction

This chapter talks about the research background, why we need face shape classification and its importance, and the existing problem

1.1 Research Background

In In recent years, as people's living standards and consumption levels continue to improve, the pursuit of beauty has also risen. The beauty industry has been growing at an average annual rate of 4.5 percent following the recent increase in people's interest in appearance[1]. Fashion industry which constitutes an aesthetic economy has played an important role in global economy in decades[2].

Pursuing the perception of beauty is a nature for human beings[3]. Over the past centuries, artists and psychologists have been tried to explore the mystery of beauty[4]. Beauty is closely related to people's face shape, face shape classification is considered a common task in beauty and fashion purposes. A good understanding of one's face shape can give people the right direction of how to make up, choose a haircut, dressing style, etc. So face type classification recognition is important, especially under this environment where people like to shopping online.

Therefore, it is an important subject for people to have a better understanding of the style that suits them, so they need to know their face shape. From there they can find the right makeup and products.

1.2 Importance of Face Shape and Its Recognition

Face shape classification is considered a common task in beauty and fashion purposes. The following four points illustrate why it is so important.

1.2.1 Fashion is Closely Related to One's Face Shape

People need to consider their face shape when shopping. Going to the physical store may be helped by a stylist, but most situations require people making their own decisions, such as when shopping online. Studies have shown that picking a suitable eyewear is highly related to face shape.

Hossam, M. et al 's research result shows experts suggest that people with oblong shaped faces consider frames with oversized frames to match long and wide features. Recommendation schemes for hats, make-up, jewelry, and other fashion accessories provide other related applications[3].

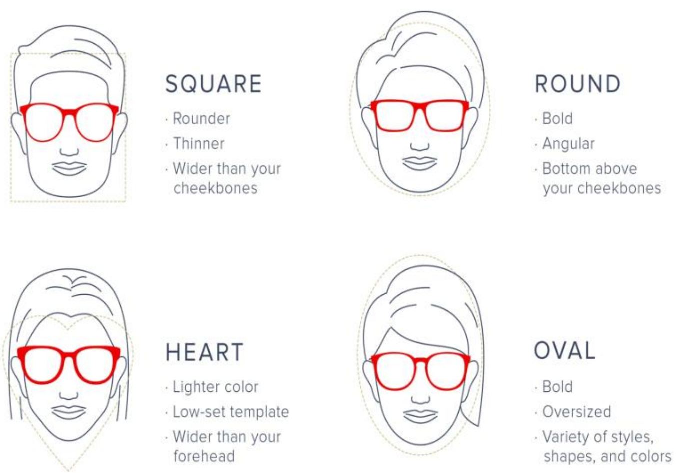

Fig 1.1 best-glasses-for-your-face[4]

It also related to hair style, " it would be better if we know our face shape and features well before our next visit to a salon [5] ". So face shape is important for beauty and attractiveness.

1.2.2 Face and Dress Collectively Affect Beauty

A female may not have very attractive face, but have a good taste of dresses and makeups to match her face shape, which then makes her also very attractive[6]. Therefore, even if a person does not looks pretty, when they have a good sense of dressing, they would also be attractive. So it is important to let people know what is suitable for their face based on the face shape.

1.2.3 Appearance Affect One's Career

It is often assumed that what is beautiful is good, as physically attractive individuals tend to be perceived and treated more positively in daily social interactions than less attractive individuals are[7].

Studies have shown that attractive people are usually hired sooner, get promotions more quickly and are paid more than their less-attractive co-workers[8].

A survey shows physical appearance affects long-term career success for economists[9].

Overall, more physically attractive individuals were viewed both more positively and more accurately in first impressions[7].

1.2.4 Face Shape is Important for Face Recognition

Rough face shape filtering before face recognition can effectively improve the recognition accuracy and speed[10]. So research on machine learning(ML) about face shape recognition might help the process of face recognition.

1.3 Problem Statement

In our daily life, we often see men or women who are not bad in appearance, either in looks or body, but, because of negligence in dressing, they do not look dedicate and attractive.

In this era of big data, the Internet is full of various dressing tutorials and makeup tutorials, and online shopping is also popular. But even so, too much data is more likely to confuse people, and the complexity of work and affairs consumes people's patience at the same time, that makes people hard to spare time to determine what suits them.

Meanwhile, even though some people go to great lengths to dress up, there are many examples of people who can't find their way or dress in the wrong way to look out of place.

People will make their style choices according to their face shape, but the face shape classification is equally ambiguous. It introduces people's face shape in five[5], six[14], seven[12], or nine[13] types.

Fig 1.2 Nine face shapes [30]

"Many people don't fit into just one face shape category. People can have a mix of different face shapes and, in fact, it's quite common ", said by Dr. Shafer[15]. Because of the ambiguity of face shape, people can't be sure of their face shape; also some people don't scientifically follow their face shape to find a style that suits them, so they often don't know what suits them. Wan-Yu Chen et al introduced that the apparel which looks perfect for one person on a fashion photograph may not fit for another person who subscribes the photo[11].

In fact, appearance is the first step for others to get to know us. The clothes we wear every day introduce our aesthetic preferences, professional attributes, and personality traits to everyone we come in contact with. So it's worthwhile to let people know their face shape.

This paper will introduce a more refined classification of sixteen face shapes, aiming to make the classification of face shapes more accurate and people can find their own face shape precisely.

2 Content

This chapter distinguished the research scope, purpose, research objectives, rationale, and research significance.

2.1 Research Scope

There are various ways to apply face types in the beauty and fashion fields. This study focuses on the implementation of ML, to automatically identify face types based on sixteen face shapes and describe their application fields. And will separately give out the accuracy for male and female.

2.1.1 Reason for Include Male in This Research

1. Face style can apply to both genders.
2. Beauty is a melting pot. No longer a single, static idea, this ever-evolving concept of ‘come as you are' is changing cultural perceptions to show that beauty isn't confined or defined by gender[40].
3. Male also have the right to find bend gender norms. For example, In 2018, there was approximately 50,000 attendees respectively at the New York and Los Vegas DragCons, and the beauty market driven by sales at DragCon was worth $8.2 million USD from the New York and Los Angeles alone[41].
4. As chapter 1.2.4 mentioned, face shape recognition can help the speed of face recognition. So both gender is important.
5. An increasing number of male are interested in beauty and makeup, over the past decade, Google searches around male make-up have increased by 67.5%[17], and younger male gives more attention[18]. Jason Chen, general manager for Chinese online retail site Tmall, told Coresight that across China "supply is not able to meet demand for male make-up products. "[39]

6. The more data available, the more accurate or variaty of conclusions can be obtained.

The application field about fashion style of male have not been explained in chapter 3.3.2, it is considered as a future research. But basically, the face shape classification standards(3.1), analysis of sixteen face type's characteristics(3.2), the four groups of face types(3.3.1), and the application field for beauty(3.3.3) is analysed for both gender.

2.2 Research Purpose

This study aims to classify users' face shapes more carefully according to their facial features, and quickly identify a person's face shape through the system so that users can better find the right style for themselves.

2.3 Objectives

Investigate past approaches to face recognition, study face taxonomy, and classify faces in a more detailed way. Investigate previous and related research, find their methods and results, identify their problems, and use alternative solutions for scientific face classification with ML.

2.4 Rationale

Face shape has vital importance for the choice of human appearance and image. The importance of shape is to give people an impression, enhance good feelings, and increase charm. Also, the recognition of face shape is beneficial to the study of face recognition. However, today's taxonomy of face shapes is ambiguous and diverse, and it is often the case that one person covers multiple face shapes and cannot be determined. Therefore, this study helps people to know their own shape more precisely so that they can understand how to wear makeup and dress to enhance their charisma value.

2.5 Research Significance

People's pursuit of beauty is becoming more and more eager as the economic level rises, but the current classification of face shapes fails to accurately give people the direction of fashion choices. So a uniform and detailed face-type classification method should be useful.

This research aims to build an auto-performed face type recognition system to let computer helps people distinguish their type based on their facial features reducing much trouble. This study use ViT to extract people's facial features and used five classifiers to see the accuracy. It aims to classify people's face shape automatically with ML and facilites brands and individuals.

2.5.1 Research Significance for Brands

AI gives convenience to human life. It can automatically analyze and identify people's face shapes, quickly process large datasets and provide instant results. This can reduce the workload for fashion professionals, and allows fashion brands to offer face shape recommendations at scale, reaching a broader customer base. It helps fashion brands and retailers recommend products that are most suitable for different face shapes. It ensures that customers receive personalized suggestions that enhance their appearance and style.

Fashion is highly personalized, and people have different face shapes that influence their overall appearance. By using ML to identify face shapes, fashion brands can provide customized recommendations and styling advice. This personalization enhances the customer experience and increases satisfaction, leading to improved customer loyalty and sales.

For customers of brands, AI-powered face shape identification enhances the overall shopping experience. When customers receive personalized recommendations based on their face shapes, they feel understood and valued by the brand. This level of personalization and attention to detail improves customer engagement, satisfaction, and loyalty.

2.5.2 Research Significance for Individuals

ML can analyze an individual's face shape and provide personalized style guidance. By understanding their face shape, individuals can make informed decisions about clothing styles, hairstyles, and accessories that best complement their features. This guidance helps individuals enhance their overall appearance and develop a personal style that suits them and helps individuals stay updated with the latest fashion trends.

Also, the guidance can encourage individuals to explore and experiment with different fashion styles. With the knowledge of their face shape, individuals can step out of their comfort zones and try new clothing styles, hairstyles, and accessories that they may not have considered before. It adds variety to their fashion choices.

By understanding which styles and accessories are most suitable for them, individuals can feel more self-assured in their fashion choices which can aid individuals in expressing their personal style. This confidence can positively impact their self-image and how they present themselves to others. And as chapter 1.2.3 showed, better appearance enhances workplace competitiveness.

3 Sixteen Facial Styles

The purpose of this research is to classify face shapes meticulously and make 16 facial styles as a formula for becoming beautiful. Once people knew their type, they can look for their style's representatives, and get the scientific direction of what is suitable for them.

3.1 Standard of Sixteen Facial Styles

It has four categories of features: {Defined, Soft}, {Adult, Childish}, {Masculine, Feminine}, {Wide, Narrow}. Do permutate combination of four sets of features, sixteen results can be obtained. The combination method is the same as the Myers-Briggs Type indicator (MBTI)[19], assign a value to each of the four categories, and one letter from each category is taken to produce a four-letter result. For example, a person with defined(D), adult(A), masculine(M), and wide(W) features is DAMW; a person with soft(S), childish(C), feminine(F), and narrow(N) features' type is SCFN.

This classification methodology and its standards combines the previous common rules[46][47], and is used by the stylists at the actual store.

·Defined and Soft

Defined features mean there is a large proportion of facial features on one's face, a strong presence, high three-dimensionality, and a strong sense of facial bones.

Soft features are in the opposite direction with "defined ", which means the person has smaller facial features relative to the proportion of the whole face, weak presence and flattened features.

·Adult and Childish

The adult feature looks more mature, with a long nose, pointed nose tip, long eyes, and hard facial lines.

The childish look is more cute, with a short nose, rounded nose tip, round eyes, and smooth facial lines.

·Masculine and Feminine

The masculine feature is more aggressive and powerful, with stern eyes, narrow eyebrow-eye spacing, and a defined jawline.

Feminine looks non-aggressive, with gentle eyes, soft lined eyebrows, and soft jawline.

·Wide and Narrow

If the whole face looks short and wide, it is wide;

on the contrary, if the face is narrow and pointed, it is narrow.

3.2 Characteristics of each face shape

The table below shows the characteristics of each face shape based on its facial features and contour features. After the table, there's an explanation of why each face type has these characteristics.

DAMW Fierce, flamboyant, aggressive, visually appealing, powerful, and strong feeling. Facial occupancy	**SAMW** Reticent, aggressive, stern-eyed, tenacious, independent.

DAMN Strong visual impact, sense of intelligence, heroic spirit feeling, indifferent, mature	**SAMN** Aggressive but no sense of distance, down-to-earth, atmospheric, powerful
DAFW Gentle, charismatic, dignified, stable, spacious, mature, and charming.	**SAFW** Friendly, easy-going, and approachable. Will look cheesy if has a chubby face.
DAFN High sense of feminine, mature, bright and attractive, flirtatious, focus of the crowd.	**SAFN** Innocent, pure, sensual, mature, cool, subtle, and introspective. Well worth a second look.
DCMW Eye-catching, impactful, rebellious, spontaneous, firm, blunt, wild, vital, noble feeling.	**SCMW** Atmospheric and exquisite. S+C means low aggressive, and M+W is high aggressive, therefore this type has a strong sense of the contract.
DCMN Fresh, pure, refined, elegant, cool, highly malleable	**SCMN** Unisex, juvenile, refreshing, innocent, free and spontaneous, detached

DCFW Looks young, lively, and the most emotionally contagious type. Maybe looks not smart since too cute.	**SCFW** Cute, well-behaved, young state, caricature sense, no offensive, sweet, rich expressions.
DCFN Bright, eye-catching, not aggressive, smooth lines, arousing the desire for protection, charming and pathetic.	**SCFN** Youthful, well-behaved, smooth, affectionate, simple, gentle, sustainable beauty

Table 3.1 Characteristics of each face shape

To explain this table, take defined, childish, masculine, wide(DCMW) as an example, which Lalisa Manobal belongs to.

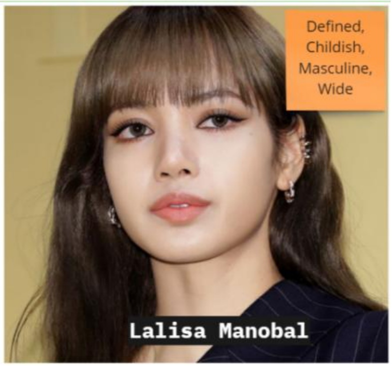

Fig 3.1 DCMW - Lalisa Manobal

·DCMW (Defined, Childish, Masculine, Wide)

This type of face has a combination of D+M(defined and masculine), which makes the face has a large proportion of features and high three-dimensionality, so they got an eye-catching and impactful face.

C+W(childish and wide) makes the face smooth and gentle contour lines and looks more cute.

Combining these features, this type has a heavy sense of color, bone, eyes, and features. As the table mentioned, this type is rebellious, spontaneous, firm, blunt, wild, vital, and noble.

These are the advantages of this type. And its disadvantage is that once this type of person gains weight, the visual proportion of the face to the overall body will be too high, and the lovely and noble mystery feeling will disappear. Another point is because this type has a very "heavy " feeling, they cannot wear complicated clothes, or it will cause visual overload. So they're suitable for more simple dresses.

<u>The other types are explained next in a relatively concise manner.</u>

·DAMW (Defined, Adult, Masculine, Wide)

The most aggressive face shape. D and A imply high maturity and a high proportion of facial features, while M and W imply aggressiveness and masculinity. The whole face is highly defined and visually appealing.

·DAMN (Defined, Adult, Masculine, Narrow)

The defined face has a visual impact, the adult feature looks mature and stable, the masculine feature gives a person a sense of power, while the narrow face reduces the wild feeling of the masculine and improves some of the sophisticated feeling.

·DAFW (Defined, Adult, Feminine, Wide)

The adult feature makes the genre look soft and non-aggressive, but the defined feature has a visual impact. The wide face makes up a dignified atmosphere, so this type is mature and charming.

·DAFN (Defined, Adult, Feminine, Narrow)

The combination of these four characteristics is the highest expression of femininity, the attractiveness will not be low and the ceiling is extremely high. The face has smooth lines, the narrow face looks delicate, mature, and intense.

·DCMN (Defined, Childish, Masculine, Narrow)

D and M represent large facial features with blunt and angular bones, but because of N, the angle is not as obvious as DCMW. N and M bring a masculine refreshing feeling, while the childish feature reduces the aggressiveness of M.

·DCFW (Defined, Childish, Feminine, Wide)

This type is quite similar to DCMW. They both defined, with a large proportion of features, high three-dimensionality. And looks childish, with good skin status, look energetic; and a wide face makes them a high blunt feeling. The difference is because of the feminine feature, this type has much more soft lines, and the conspicuousness of bones is lower than the skin, so looks more cute and lively.

·DCFN (Defined, Childish, Feminine, Narrow)

The defined and feminine feature makes this type has a large proportion of facial features, with smooth contour lines. Childish causes the whole face to remain a little young and sweet feeling.

·SAMW (Soft, Adult, Maculine, Wide)

This type has a smaller proportion of facial features, lighter color, high maturity, and a wide face, so the overall femininity is low and the aggressiveness is strong. So it has the characteristics as Table 1 shown.

·SAMN (Soft, Adult, Maculine, Narrow)

This type of temperament comes from the combination of defined, masculine, and adult. D represents small features, light color, and low facial stereoscopic, and M represents aggressive, sharp contour lines, not smooth, the combination of D and M creates a cold sense of distance. And A represents maturity, thus stable and mature.

·SAFW (Soft, Adult, Feminine, Wide)

A short wide mature face with feminine feature has a soft face contour line, the face is relatively flattened, low three dimensionality, More spare space on the face, and adult features make this type looks mature.

·SAFN (Soft, Adult, Feminine, Narrow)

Soft and feminine's combination gives people a non-aggressive and innocent feeling, with an adult feature that makes the face looks mature, this type looks gentle and intellectual. Compared to wide faces, narrow faces will look more strict. So it gives this type of sense of distance.

·SCMW (Soft, Childish, Masculine, Wide)

Soft and childish indicate low aggressive, masculine and wide means high aggressive, so this type has a sense of contrast. The four features combined together are high sense of bones, high blunt and angular bones, a small proportion of facial features, and the features are biased towards round and young states.

·SCMN (Soft, Childish, Masculine, Narrow)

This face shape has the highest adolescent feeling. The facial features are small in proportion, this type has a juvenile state, with narrow and long features, the sense of visual presence is relatively low, thus increasing the air of freedom and spontaneity.

·SCFW (Soft, Childish, Feminine, Wide)

Soft means small facial features and low facial dimensionality, combine with feminine and childish, generates a very low aggressiveness type. And childish with a wide face let this type has a rounded and wide face contour, making a soft, cute, and sweet type.

·SCFN (Soft, Childish, Feminine, Narrow)

The combination of S and C makes a small proportion of facial features, low three-dimensionality, non-aggressive, and not eye-catching. F and N give a feminine smooth and gentle contour line. Compare to SCFW, this type is less rounded and sweet, but more gentle and approachable.

3.3 Application Field

This chapter will talk about how the sixteen face types can be applied to help people with their makeup and clothing choices.

3.3.1 Four Groups of Sixteen Face Types

The sixteen face types can be divided into four groups, the four face shapes in each group have their common similarities and characteristics.

·Childish and Wide feature (C+W)

This group's people have short, round, and wide faces, making the face visually extend horizontally so look young. So it is not suitable for too mature dressing, nor for long straight hair without bangs, which will make them look dull.

It is recommended to use interesting designs with a strong sense of art, such as round necks, square neck clothes, and other loose tops to make the wide face have a sense of extension.

Hairstyle is recommended to choose fluffy and dynamic hairstyles, such as short hair with layers, medium-length curls, with interesting hair accessories to match the young and cute round wide face.

·Childish and Narrow feature (C+N)

This type generally gives people a refreshing and slightly sweet feeling, and it is easy to gain people's trust. The young and narrow feature makes the face look small and the person is on the youthful side. The youthfulness of this type mainly comes from the shape of the skin and the shape of facial features, but the layout of the features does not have much influence. Therefore, there are more options for dressing.

For dressing, some mainstream styles, such as Chinese millennials style, American, Hong Kong, French, normcore, and sweet style all can be held.

In terms of hairstyles, long straight hair without bangs can show a sense of innocence, big wave curly hair can show sexy, and flowing short hair is light and agile, there are lots of options for this group.

·Adult and Wide feature (A+W)

This type is based on a round and wide face shape with mature facial features, the stretching facial ratio looks atmospheric and stable. A wide face tends to show fleshiness, and a fleshy mature face does not meet the mainstream Asian aesthetic, so is not good to gain weight.

In terms of dressing, it is recommended to use high-quality fabrics and advanced designs to enhance the class and choose a wide and skin-revealing neckline to give the face a sense of extension and increase freshness. Better to avoid the cute sweet styles which are contrary to the dignified and atmospheric face.

·Adult and Narrow feature (A+N)

This group is based on the narrower and longer face with mature features, the overall contour is sharp and clean, with a high degree of maturity and a sense of experience. The features are relatively compact, with smooth and sharp lines. It is not recommended to choose the young state style, such as pink princess dresses, straight bangs, etc., which will be contrary to the qualities of the face.

The dressing style is more selective, Hong Kong style, French, and Chinese style dressing can be held. They can also try a variety of sophisticated or exaggerated accessories, but it is recommended to give preference to mature classic and simple designs.

Hairstyle suggestions are long straight hair without bangs, big curly hair, a simple ponytail, etc. Just need to match to the mature and elegant temperament.

Besides, there are some other application fields as follows two points: fashion and beauty.

3.3.2 Fashion

Fig 3.2 Dressing style based on A-C, F-M axis

The above figure shows the suitable dressing style for different face shapes, based on the latitude of feminine and masculine in horizontal, adult and childish in vertical.

Adult features are suitable for mature style, vintage, elegance, or office style, Childish looks are suibatble for more cute style, lolita, student uniform, etc.

Feminine features tend hot and sexy, like Italy-french style, Hong-kong style, or Thai style, masculine is suitable for gender-neutral dresses, casual wear, sportswear, etc.

After integration, the adult and feminine feature is good for the sexiest styles and childish and masculine suitable for the most youthful styles which look innocent.

3.3.3 Beauty

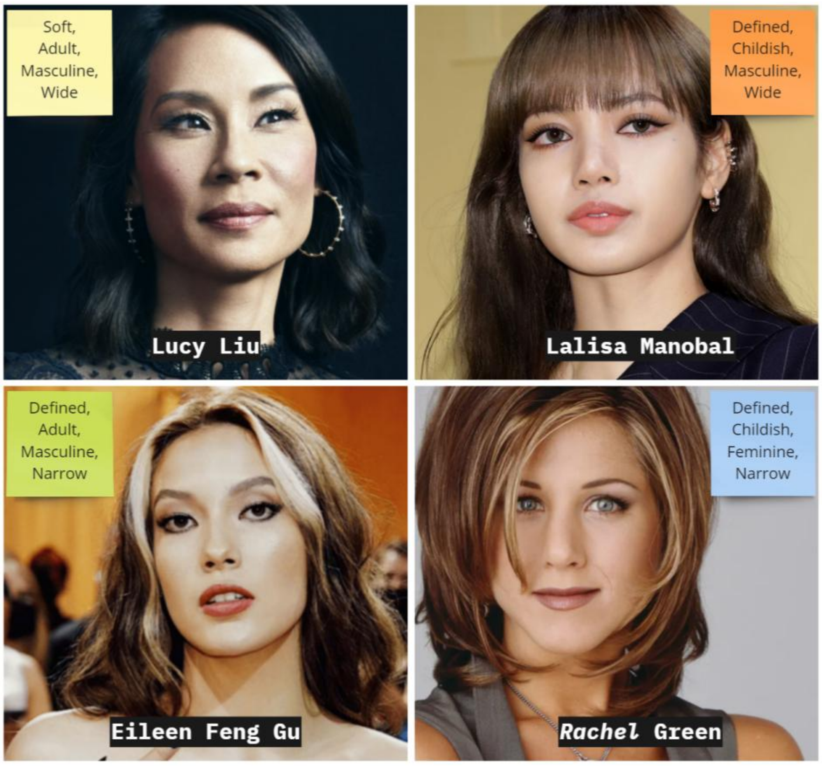

Fig 3.3 Celebraties with various face shapes

For beauty, once people find their own face shape, they can find celebrities that have the same face shape as them as their styles "style representatives " to learn how to makeup and know what is good.

4 Literature Review

This chapter examines the past literature on ML face recognition, surveying the neural networks, models, face feature extraction and classification methods they use, and analyzing them.

4.1 For Feature Extraction

Most related work used CNN(Convolutional Neural Network) and HOG(Histogram of Oriented Gradient) for feature extraction.

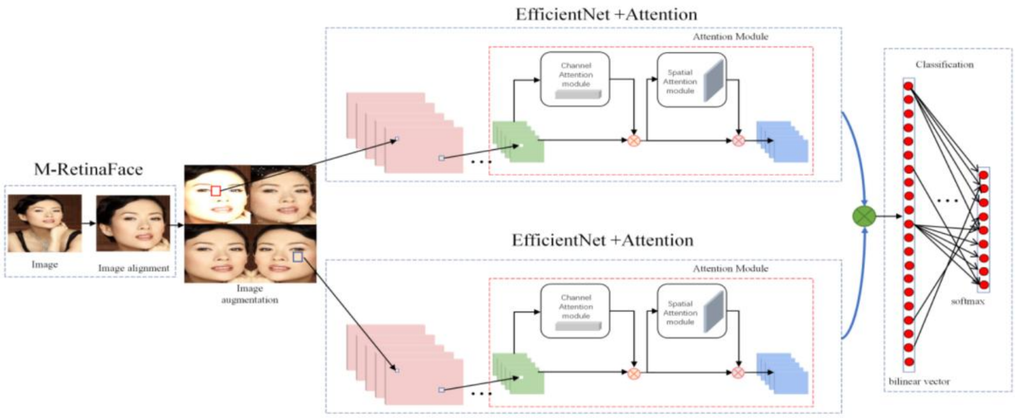

Figure 1. The framework of the proposed algorithm.

Fig 4.1 The framework of the proposed algorithm[10]

The above research combined EfficientNet and Attention mechanism and proposed AB-CNN network for extract features. It compared B-CNN and AB-CNN, and AB-CNN shows a better training result.

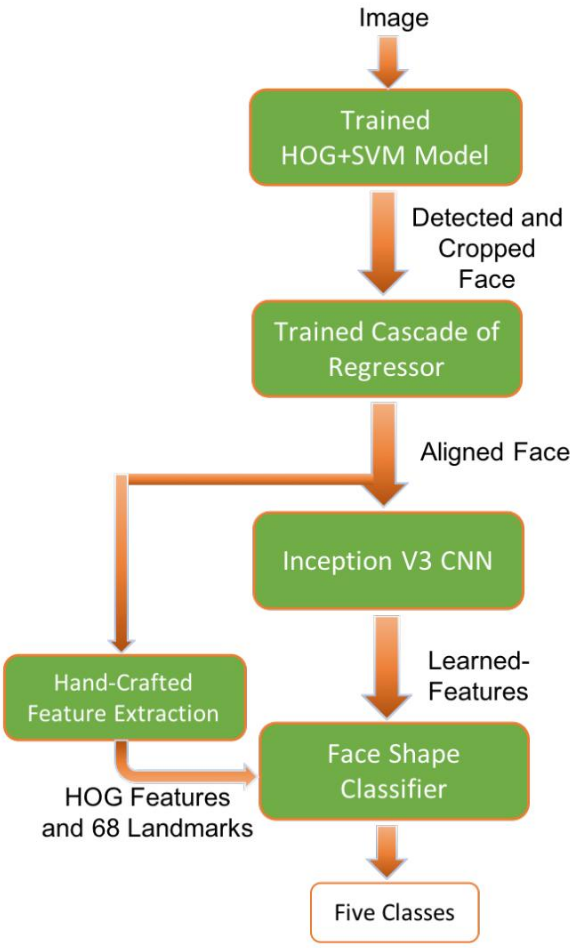

Fig 4.2 Block diagram of proposed face shape classification system[20]

The above research used V3 CNN for learning features. It trained HOG+SVM to crop the image and then detacted landmark and align face. This study used 500 images.

Besides, Ji, W. and Jin, L.[21], Darugar, M.J. and Kiong, L.C.[22], and Tabassum, F. et al.[23] all did their researches used CNN.

Except the study that mentioned above[10] used HOG, Darugar, M.J. and Kiong, L.C.[22], Alashkar, T., Jiang, S. and Fu, Y. [24], Sharzeel Saleem.et.al.[25], Chandrakala, M. and Durga Devi, P. [26]'s research all used HOG features.

LBP(Local Binary Patterns) also be widely used in face recognition field. [24][27][11].

These researches mentioned above all published in 2021 to 2022, except two published in 2017: [22], [24], one in 2014[11], and one in 2020[27].

Some other algorithms also been used in variaty of researches like attention mechanism, LDA, PCA[28], AAM[14][29], etc.

4.2 For Classification

Most researches mentioned in chapter 4.3 used SVM(Support Vector Machines) for the classification work. It is the most classical model for classification. Except [26] cascading k-NN followed by SVM to increase the accuracy.

4.3 Analysis of Machine Learning Mechanism

Through these researches, it is obvious that LBP is not usually used in recent years, the reason is it does the extraction of localized features as a basis for judgment, and it is very sensitive for directional information.

HOG is not suitable for large datasets since it has a large latitude of features and calculation speed is slow.

Using CNN to extract features and SVM as a classifier is the most commonly used supervised ML approach on face recognition.

4.4 Face Shape Classification Method

The previous classical classification method of face shape is always inaccurate as chapter 1.3 mentioned.

The most common face shape classification method is ambiguous. It introduces people's face shape in five[5], six[14], seven[12], or nine[13] types, the nine types of face shapes included round, oval, square, rectangle, inverted rectangle, triangle, heart, diamond and oblong.

But the face classification rules for six face shapes are ambiguous and hard to represent mathematically, therefore it can lead to misclassification[14]. Besides, "Many people don't fit into just one face shape category. People can have a mix of different face shapes and, in fact, it's quite common", said by Dr. Shafer[15]. Also, board certified plastic and reconstructive surgeon Dr. K. Roxanne Grawe, M.D says most faces "are a hybrid of shapes[16]. " This is the disadvantage of the traditional face shape classification method.

4.5 Problem of Existing Researches

Feature extraction plays an important role in face shape recognition, but the existing research pays more attention to the design of classifiers, and less attention to the feature extraction process.

The existing researches always use a single classifier which might not get the best result since it might be one-sided.

Another thing is, the traditional face shape classification method is imprecise, Which makes people hard to know what is suitable for them.

5 Research Methodology

This chapter will introduce the research method, process, feature extraction algorithm, and classifier that has been used. Use five kinds of classifiers for the classification work and show the result of this study.

5.1 Machine Learning Process

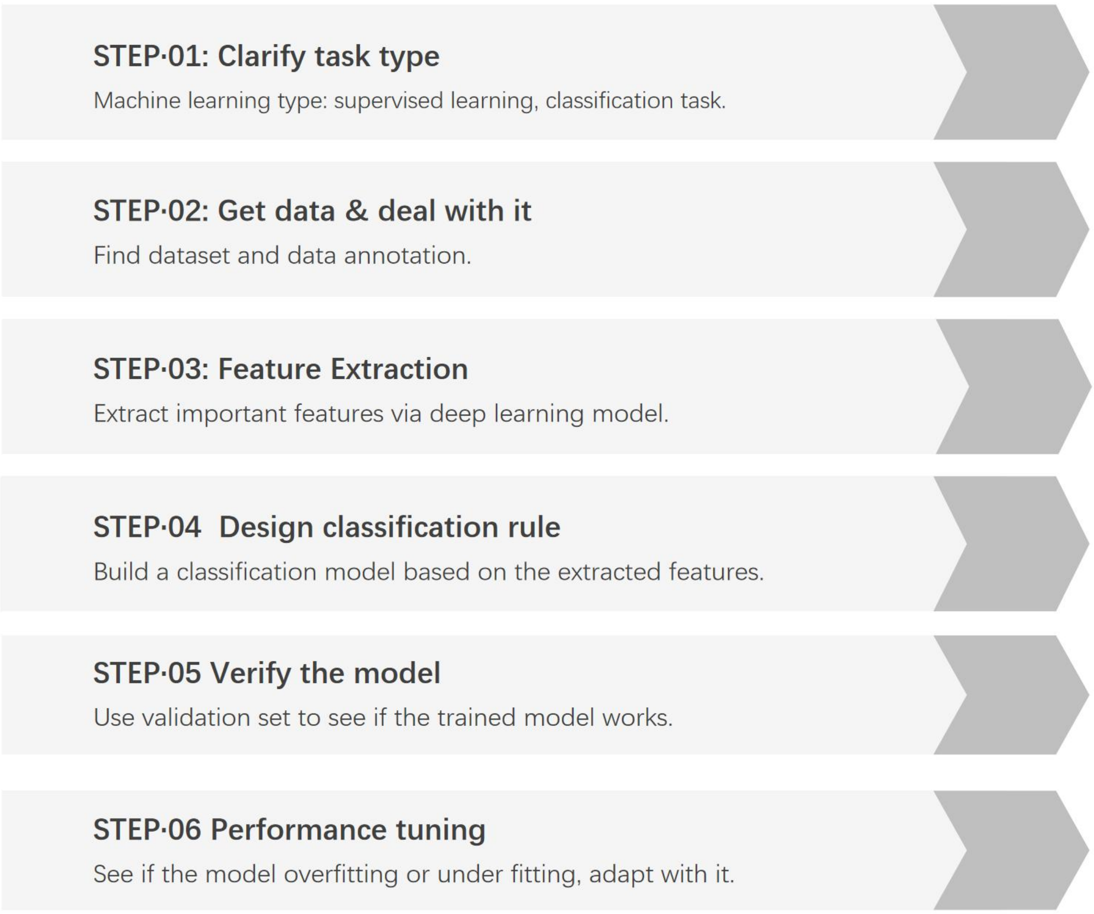

Fig 5.1 Machine Learning Process

The above image shows the basic process of this research. It will use supervised ML, create the dataset and do data annotation, make a system to let the computer do the feature extraction and classification work, verify the model by running it and see the accuracy, and debugging the system, enhance the system, and finally analysis which faces type people belongs to.

5.2 Dataset

5.2.1 Introduction of Datasets

There are numerous large datasets of faces, with different features and focuses, Sifting through the datasets to find the best fit for a given project can take time and effort[32]. Since this study requires a non-synthetic face-focused database of real people, four available datasets were filtered: Japanese Female Facial Expression(JAFFE)[33], VGGFace2[34], Chinese Academy of Sciences - Peking University - Asian Face (CAS-PEAL)[35] and CelebFaces Attributes Dataset(CelebA)[36]. For this research, since the standards and attributes has never been used in the related researches, there is no dataset specifically publicly available that is well labeled with face types, all of which require manual labeling.

JAFFE: This dataset has ten Japanese female expressers, seven posed face expressions includes six basic facial expressions: anger, disgust, fear, happy, sad, surprise, and one neutral expression. 213 images in total. It is small size and focus on female facial expressions.

VGGFace2: A large-scale benchmark dataset for face recognition and verification tasks. The dataset has 3.31 million images of 9131 identities, with large variations in pose, age, illumination, ethnicity and professions[37].

CAS-PEAL: This dataset primarily focused on Asian faces. It has 99,594 images of 595 males and 445 females, PEAL means pose, expression, accessory, and lighting. Its subset CAS-PEAL-R1 includes 30,900 images of 1040 individuals.

CelebA: A widely used benchmark dataset in the field of computer vision and ML, particularly for tasks related to face recognition, facial attribute analysis, and other facial image-related tasks. It has 202,599 celebrity images of 10,177 number of identities, each with 40 attribute annotations. CelebA has large diversities, large quantities, and rich annotations[36].

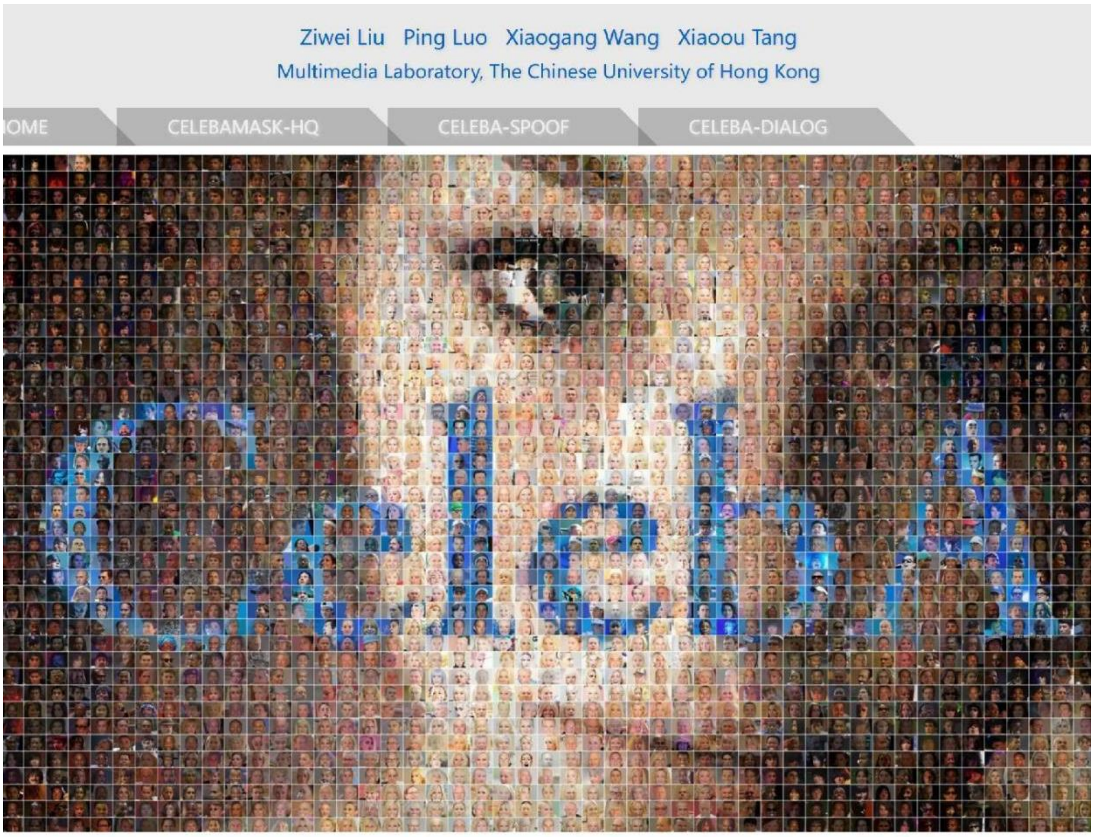

Fig 5.2 CelebFaces Attributes Dataset (CelebA) [36]

5.2.2 Reasoning for Dataset Choice

This research used CelebA dataset.

JAFFE and CAS-PEAL's data is not enough for this research. Another thing is, since JAFFE is a dataset for face expression, the face shape cannot be identified properly, for example, the surprised expression is easy to be misjudged as a melon seed face shape[38].

VGGFace2 and CelebA have similar identities, but compared CelebA dataset, VGGFace2 has more images. Since this study was formulated to require different people's face as data, so CelebA is selected as the dataset for this study.

5.2.3 Data Labeling

Since the standards and attributes have never been used in the related research, there is no dataset specifically publicly available that is well labeled with face types, all of which require manual labeling. This research selected 9,998 male images and 10,182 female images from the CelebA dataset for training. The pictures chosen are front faces, without teeth showing, and without glasses. The formers of each attribute are denoted by 1 and the latter are by -1. There are 20,180 images in total, so 80,720 data were labeled for this research.

The majority of this dataset is western faces, eastern faces might have a different training result, which is considered as future research.

All the data has been labeled manually, through the four sets of sixteen facial styles: wide and narrow, defined and soft, adult and childish, masculine and feminine. The formers of each attribute are denoted by 1 and the latter are by -1.

	Defined/ Soft	Adult/ Childish	Masculine/ Feminine	Wide/ Narrow
00001 .jpg	1	1	1	-1
00095 .jpg	-1	-1	-1	1
00008 .jpg	1	1	1	-1
00064.jpg	-1	-1	-1	1

Table 5.1 Example of data pre-processing method

Fig 5.3 Examples in the dataset, the two female images have the contrary features, and so do the male images

To take images 00001 and 00095 as examples, defined facial features means there is a large proportion of facial features on one's face, strong presence, high three-dimensionality, and a strong sense of facial bones, soft features means the person has smaller facial features relative to the proportion of the whole face, weak presence, and flattened features. So the left female image is 1, defined, and the right image is -1, soft.

Male (9998)								
	1,1,1,1	1,1,1,-1	1,1,-1,1	1,1,-1,-1	1,-1,1,1	1,-1,1,-1	1,-1,-1,1	1,-1,-1,-1
Number	108	1052	142	1462	222	440	241	879
Percent	0.011	0.105	0.014	0.146	0.022	0.044	0.024	0.088
	-1,1,1,1	-1,1,1,-1	-1,1,-1,1	-1,1,-1,-1	-1,-1,1,1	-1,-1,1,-1	-1,-1,-1,1	-1,-1,-1,-1
Number	202	**2651**	**46**	1930	70	224	55	274
Percent	0.020	0.265	0.005	0.193	0.007	0.022	0.006	0.027

Table 5.2 Data analysis for male data

Female (10182)								
	1,1,1,1	1,1,1,-1	1,1,-1,1	1,1,-1,-1	1,-1,1,1	1,-1,1,-1	1,-1,-1,1	1,-1,-1,-1
Number	481	1338	296	**2392**	572	432	**155**	387
Percent	0.047	0.131	0.029	0.234	0.056	0.042	0.015	0.038
	-1,1,1,1	-1,1,1,-1	-1,1,-1,1	-1,1,-1,-1	-1,-1,1,1	-1,-1,1,-1	-1,-1,-1,1	-1,-1,-1,-1
Number	493	1046	259	1171	385	343	183	249
Percent	0.049	0.103	0.025	0.115	0.038	0.034	0.018	0.024

Table 5.3 Data analysis for female data

Male + Female (20180)								
	1,1,1,1	1,1,1,-1	1,1,-1,1	1,1,-1,-1	1,-1,1,1	1,-1,1,-1	1,-1,-1,1	1,-1,-1,-1
Number	589	**2390**	438	**3854**	794	872	396	1266
Percent	0.029	0.118	0.022	0.191	0.039	0.043	0.020	0.063
	-1,1,1,1	-1,1,1,-1	-1,1,-1,1	-1,1,-1,-1	-1,-1,1,1	-1,-1,1,-1	-1,-1,-1,1	-1,-1,-1,-1
Number	695	**3697**	305	**3101**	455	567	**238**	523
Percent	0.035	0.183	0.015	0.153	0.023	0.028	0.012	0.026

Table 5.4 Data analysis for all the samples

The above tables analyze the data on males, females, and both gender respectively.

The yellow background means the category with the least number of people, green means the largest number.

For females, "-1,1,1,-1 " has the largest number of people, which means narrow, defined, adult, feminine (DAFN); For males, the category "1,1,1,-1 " has the largest number of people, which means wide, defined, childish, feminine (DCFW).

For the groups with the least number of people, for male-only 47 faces includes in "-1,1,-1,1 ", which means narrow, soft, adult, masculine(SAMN); for females, only 156 faces includes in "1,-1,-1,1 " which means wide, defined, childish, masculine(DCMW).

For the total number, "1,1,-1,-1 ", wide, defined, childish, feminine(DCFW) has the largest number, "-1,-1,-1,1 ", narrow, soft, childish, masculine(SCMN) has the least number, only 242 over 20,180 faces.

The four groups marked as red color has been separately trained since they have more samples.

5.3 Research Method

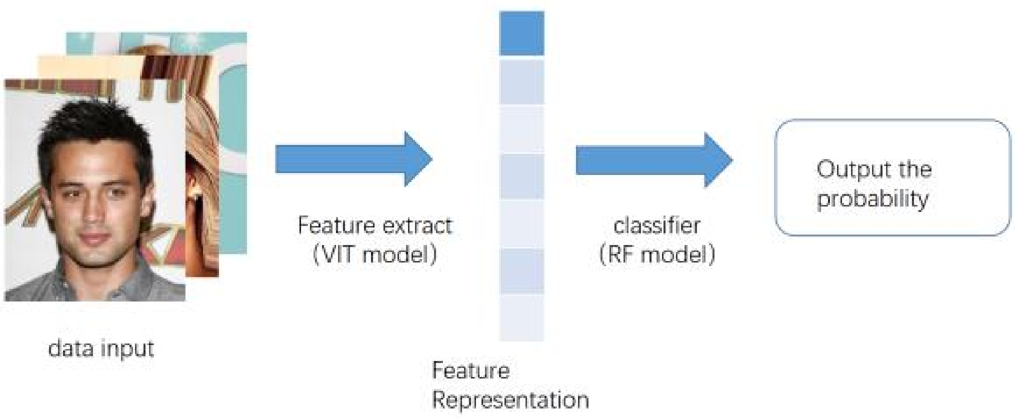

Fig 5.4 Research method

This research used the Vision transformer (ViT) model for feature extraction, and RF(random forest) algorithm for classification. After classification, the system will output the probability of each face shape.

5.3.1 Feature Extraction - ViT model

ViT is a deep learning architecture that revolutionized computer vision tasks, particularly in image recognition, object detection, and segmentation. It is the most new emerging model, it is the one that has been used in the underlying structure of ChatGPT. It is suitable for large data sets.

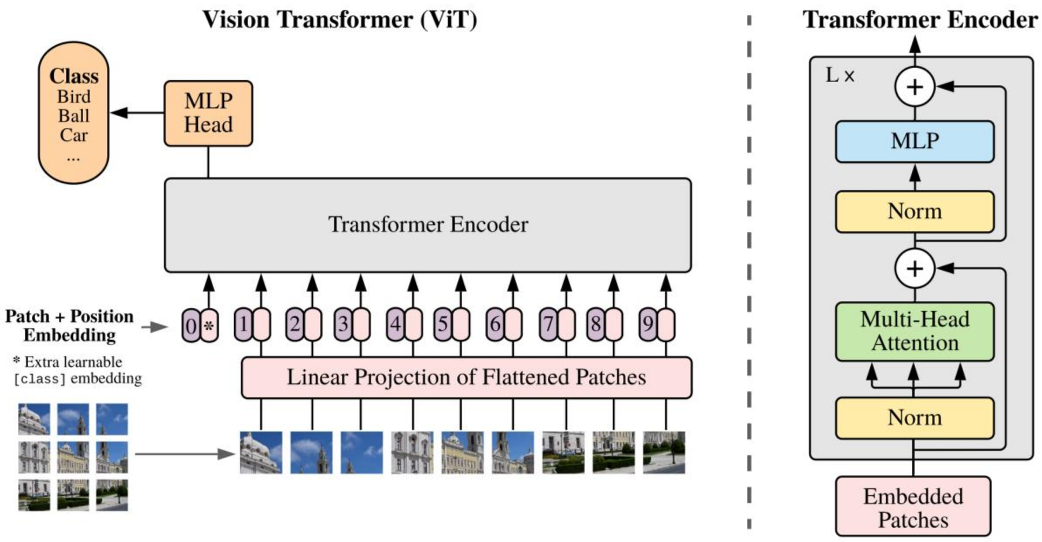

Fig 5.5 ViT model overview[31]

The main idea behind ViT is to treat an image as a sequence of 1D tokens, where each token represents a specific patch or region of the image. These patches are flattened and linearly embedded to form a sequence, similar to how words are represented in natural language processing tasks. The transformer architecture then operates on this sequence of tokens to perform various computer vision tasks.

5.3.1.1 Key Components of ViT Extractor

Image Patching: The input image is divided into small, fixed-size patches, typically in the form of square grids. Each patch is treated as an individual token and fed into the transformer model. This allows the model to process images of variable sizes consistently.

Embeddings: Each patch is embedded into a lower-dimensional vector representation, called an embedding. These embeddings are learned during the training process and capture the visual features of the individual patches.

Positional encodings: Since transformers do not have inherent positional information, positional encodings are added to the patch embeddings. These positional encodings provide information about the spatial location of each patch within the original image.

Transformer encoder: The heart of ViT. It consists of multiple layers of self-attention and feedforward neural networks. The self-attention mechanism allows the model to capture dependencies between different patches and learn global context information. The feedforward neural networks provide additional non-linear transformations to the patch embeddings.

Multi-head self-attention: In the transformer encoder, the self-attention mechanism is applied with multiple attention heads. Each head can learn different relationships between patches, enhancing the model's ability to capture various types of visual patterns.

Layer normalization: Layer normalization is applied after each layer in the transformer encoder. It helps stabilize training and improves the convergence of the model.

5.3.1.2 Algorithm of ViT Model

N is the number of patches, *H, W* is resolution of the image, and (P, P) is the resolution of each patch. is the resulting number of patches. This research used 224 x 224 resolution images, invert the images to patch sequences, and each patch's size is 16 x 16 pixels. So each image has 224 x 224 / 16 x 16 = 196 patches.

Each patch is a token in 16 x 16 x 3 = 768-dim. After add the [cls] token($z_0^0 = x_{class}$), the size of the position embeddings of the first layer of the transformer encoder is 197x768. After linear projection and positional encoding, each patch has been given a different weight. Layer normalization(LN) is given to every block (Eq.1,2,3). Multiheaded self-attention(MSA) maps the input to query, key, and value(q,k,v), this research is 12-headed, each head's qkv is 197x64, and connect the 12 outputs, the final dimension is still 197x768(Eq.1), MLP zoom in the dimension of image and zoom back out(Eq.2), from 197x768 to 197x3072 then back to 197x768. [cls]'s output (z_L^0) represents the feature of image y(Eq.3).

$$z'_\ell = MSA(LN(z_{\ell-1})) + z_{\ell-1}, \quad \ell = 1 \ldots L \tag{1}$$

$$z_\ell = MLP(LN(z'_\ell)) + z'_\ell, \qquad \ell = 1 \ldots L \tag{2}$$

$$y = LN(z_L^0) \tag{3}$$

5.3.1.3 ViT Compares to Other Networks

And previous study has shown that this model showed good train result.

	Ours-JFT (ViT-H/14)	Ours-JFT (ViT-L/16)	Ours-I21k (ViT-L/16)	BiT-L (ResNet152x4)	Noisy Student (EfficientNet-L2)
ImageNet	**88.55** ± 0.04	87.76 ± 0.03	85.30 ± 0.02	87.54 ± 0.02	88.4/88.5*
ImageNet ReaL	**90.72** ± 0.05	90.54 ± 0.03	88.62 ± 0.05	90.54	90.55
CIFAR-10	**99.50** ± 0.06	99.42 ± 0.03	99.15 ± 0.03	99.37 ± 0.06	–
CIFAR-100	**94.55** ± 0.04	93.90 ± 0.05	93.25 ± 0.05	93.51 ± 0.08	–
Oxford-IIIT Pets	**97.56** ± 0.03	97.32 ± 0.11	94.67 ± 0.15	96.62 ± 0.23	–
Oxford Flowers-102	99.68 ± 0.02	**99.74** ± 0.00	99.61 ± 0.02	99.63 ± 0.03	–
VTAB (19 tasks)	**77.63** ± 0.23	76.28 ± 0.46	72.72 ± 0.21	76.29 ± 1.70	–
TPUv3-core-days	2.5k	0.68k	0.23k	9.9k	12.3k

Fig 5.6 Comparison with state of the art on popular image classification benchmarks[31]

The smaller ViT-L/16 model pre-trained on JFT-300M outperforms BiT-L, and ViT pre-trained on the smaller public ImageNet-21k dataset performs well too.

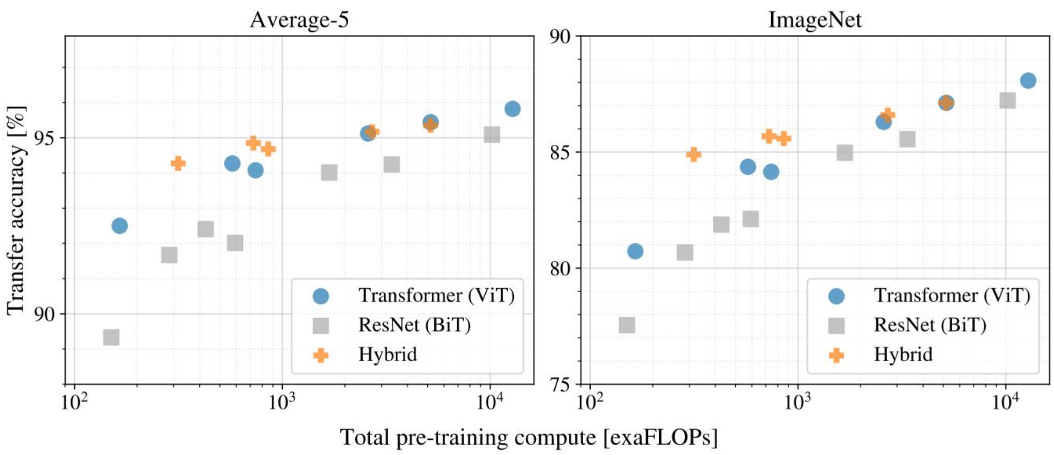

Fig 5.7 Performance versus cost for different architectures[31]

Before ViT, the most widely used computer vision models were based on convolutional neural networks (CNNs). While CNNs have been highly successful, they have some limitations, especially when it comes to scalability and capturing long-range dependencies in images.

The BiT CNNs outperform ViT on ImageNet, but with the larger datasets, ViT overtakes.

ViT to model global context and preserve the structural information compared to

CNN[42].

And it is demonstrated favorable merits of ViTs over CNNs for occlusion handling, robustness to distributional shifts and patch permutations, automatic segmentation without pixel supervision, and robustness against adversarial patches, sample specific adversarial attacks and common corruptions[43].

5.3.1.4 Benefits of ViT

End-to-end vision and language modeling [65]: ViT leverages the transformer architecture, which was initially designed for natural language processing tasks, and applies it directly to the visual domain. This unified framework enables the possibility of seamless integration between vision and language models, allowing for more efficient multimodal learning.

Scalability: ViT's self-attention mechanism allows it to process image patches independently, which significantly reduces computational complexity compared to traditional CNNs. This scalability enables ViT to handle larger image sizes and datasets efficiently, making it easier to train on more extensive image datasets.

Global context awareness: By applying self-attention mechanisms, ViT can capture long-range dependencies and global context information in an image. This ability is particularly useful in recognizing objects or patterns that may be spread across different parts of the image.

Fewer inductive biases: It is based on self-attention, has fewer inductive biases, allowing it to be more flexible and adaptive to various vision tasks.

Interpretable Attention: The attention mechanism in ViT can be visualized to understand which image regions contribute more to specific predictions. This interpretability makes it easier to troubleshoot and analyze the model's decisions.

Handling Variable-Sized Inputs: ViT can handle variable-sized inputs through the image patching technique. It allows ViT to process images of different resolutions without additional complexities.

5.3.1.5 Disadvantages of ViT

For the face shape classification work, ViT works better than other networks. But it still has some limitations for this task.

Firstly, ViT models tend to be **computationally more expensive** compared to CNNs [65], especially for larger image sizes and datasets. This is primarily due to the self-attention mechanism, which requires more computations and memory. Training ViT models from scratch can be time-consuming and resource-intensive.

Secondly, ViT models often **require more memory**, both during training and inference, due to the large number of parameters and the need to store attention matrices for computing self-attention. This can be a limitation, particularly when deploying the model on devices with limited memory, such as mobile devices or edge devices.

Thirdly, ViT's performance can be **sensitive to the choice of hyperparameters**, such as the number of attention heads, layers, and patch size. Finding the optimal hyperparameters can be a challenging and time-consuming process.

Fourthly, while ViT excels in capturing global context and long-range dependencies, it might be **less effective in capturing local patterns**, which are often crucial in tasks like edge detection, texture recognition, or local object characteristics.

Besides, it is **hard to understand**. While the attention mechanism in ViT can provide some level of interpretability, understanding the model's decision-making process and the importance of individual patches can be challenging.

Despite these disadvantages, the ViT model has demonstrated remarkable performance in various computer vision tasks.

5.3.2 Classifier

This research didn't choose ViT as the classifier but used Random Forests, Decision trees, Adaptive Boosting, Support-Vector Machine, and K-Nearest Neighbor also be used to make comparisons.

5.3.2.1 Reason of Not Using ViT as Classifier

Using ViT for feature extraction and other classifier is a common application of this model. In this context, ViT acts as a feature extractor, allowing to obtain high-level visual representations from images, which can then be used as inputs to other ML models for classification.

The choice of classifier depends on factors like dataset size, complexity, interpretability needs, and available computational resources.

The following points are advantages of Random Forest (RF) compared to ViT as a classifier:

· RF can be computationally efficient and easier to train, but they might not capture complex relationships in data as well as deep learning models like ViT or CNNs.

· RFs are generally more interpretable than deep learning models like ViT due to their decision tree-based structure.

· Random Forests can handle moderately-sized datasets efficiently, but they might not scale as well to very large datasets as deep learning models. The dataset is not that large so RF can handle it.

· RFs can handle moderately-sized datasets efficiently, and their scalability is generally better compared to deep learning models like ViT for specific use cases.

RF is a valuable option for certain use cases, particularly when computational efficiency and scalability are critical factors. The choice between ViT and RF depends on the specific requirements of the application, available resources, and the nature of the data.

5.3.2.2 Random Forests

RF [48] is a popular and powerful ML algorithm used for classification tasks. It is an ensemble learning method that combines multiple decision trees to make more accurate and robust predictions.

5.3.2.2.1 Key Components of RF

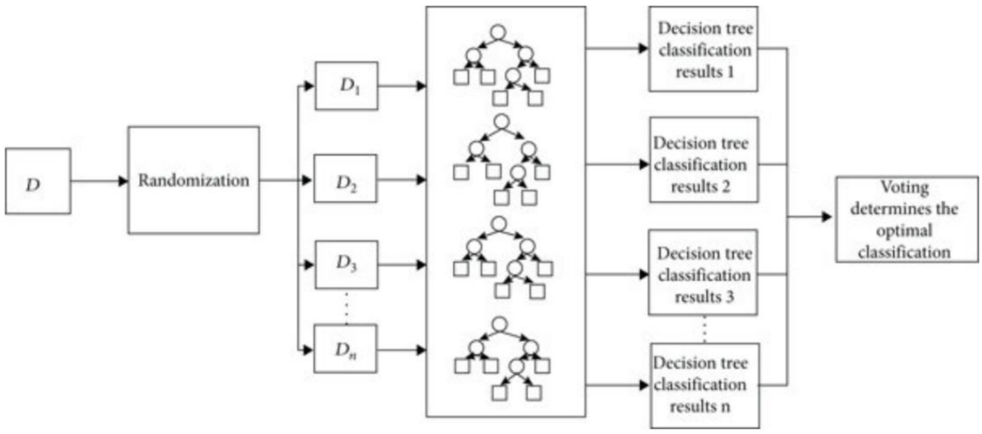

Fig 5.8 Structure of Random Forests [44]

The overall process of RF uses bootstrap resampling to draw multiple samples from the original sample, models a decision tree (DT) [49] for each bootstrap sample, and then combines these decision trees to arrive at the final classification by voting. Numerous theoretical and empirical studies have demonstrated that RF has high prediction accuracy, good tolerance to outliers and noise, and is not prone to overfitting.

Decision Tree: a tree-like model that recursively splits the data into subsets based on the values of input features. It learns a hierarchy of if-then-else decision rules from the data to make predictions.

Ensemble learning refers to the process of combining multiple individual models to improve performance. The idea behind ensemble methods is that the collective wisdom of multiple models can often lead to better generalization and accuracy than a single model alone.

Bootstrap Aggregating(Bagging): This research used sampling with replacement. During the construction of each DT, RF randomly selects a subset of the training data with replacement. This means that some data points may appear multiple times in the same subset, while others may not appear at all. Each tree in the forest is built on a random subset of the training data and a random subset of features. This randomness helps to introduce diversity among the trees, making them less correlated and more diverse in their predictions.

Voting: Once all the decision trees are built, predictions are made by combining the outputs of all the individual trees. In classification tasks, the final prediction is obtained through majority voting (the class with the most votes). In regression tasks, the final prediction is the average of all the tree predictions.

5.3.2.2.2 Benifits of RF

High predictive accuracy: Random Forests generally provide high predictive accuracy and robust performance. By aggregating predictions from multiple decision trees, RF reduces the risk of overfitting and improves generalization to unseen data.

Non-linear handling: Random Forests can handle non-linear relationships between features and the target variable effectively. This is achieved by combining multiple decision trees, each capable of capturing different aspects of the data's non-linear patterns [64].

Robustness to outliers: RF is robust to outliers and noisy data points. The majority vote or averaging mechanism in the ensemble reduces the impact of individual erroneous predictions, leading to more reliable results.

Reduced overfitting: The bagging and random feature selection processes help to reduce overfitting. By averaging predictions from multiple trees, RF reduces variance and provides more stable and accurate results.

Handles high-dimensional data: RF can effectively handle both numerical and categorical features, It can handle a large number of high-dimensionality features without much feature engineering or dimensionality reduction.

Parallelization: Training individual decision trees in Random Forests can be parallelized, leading to faster training times on multi-core processors or distributed computing environments.

No need for feature scaling: Random Forests are not sensitive to feature scaling, meaning there's no need to scale the features to a common range before training the model.

5.3.2.2.3 Disadvantage of RF

Model complexity: Random Forests can be computationally expensive and memory-intensive. Training a large ensemble of trees may require significant computational resources and time [63].

Lack of interpretability: Random Forests are not as interpretable as single DT.

Imbalanced data: Random Forests can struggle with imbalanced datasets, where one class significantly outnumbers the other. The algorithm may favor the majority class, leading to biased predictions.

Out-of-range predictions: Random Forests may not perform well in extrapolating beyond the range of the training data. Predictions for values outside the observed range of the training data may not be reliable.

Large storage requirements: Storing the trained Random Forest model can require substantial memory, especially if the dataset and number of trees are large.

Despite these disadvantages, RF remains an effective choice for a wide range of ML problems. Its ability to handle complex datasets, reduce overfitting, and provide feature importance rankings makes it valuable in many practical applications. Therefore, this research chose RF as the classifier.

5.3.2.3 Introduction of Other Classifiers

Besides RF, this experiment also used other 4 classifiers for comparison, they are Decision Tree (DT), Adaptive Boosting (AdaBoost), Support-Vector Machine (SVM), and k-nearest neighbor (k-NN).

5.3.2.3.1 Decision Tree

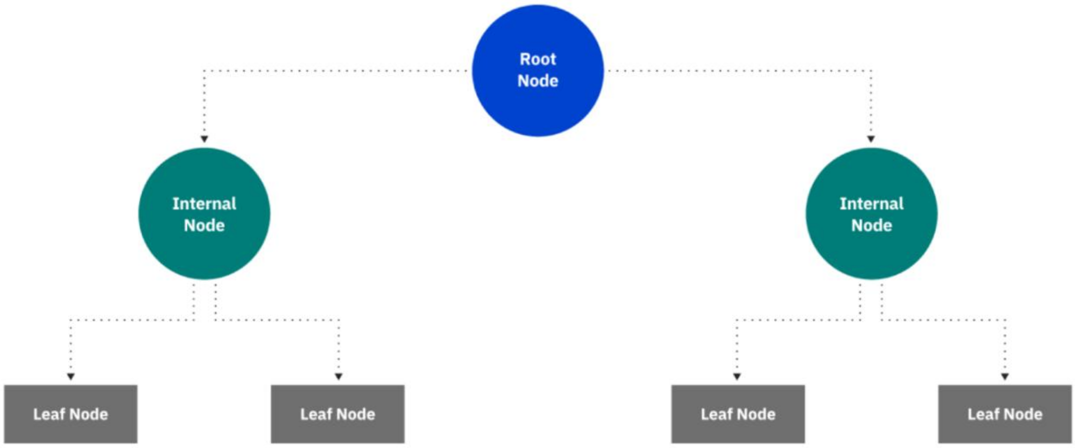

Fig 5.9 Structure of Decision Trees [58]

Decision Trees (DT) [50][51] are a popular and intuitive ML algorithm used for classification tasks. It is a type of supervised learning model that can be applied to a wide range of problems, including pattern recognition, data classification, and decision-making tasks.

The decision tree algorithm works by recursively partitioning the input data into subsets based on feature values, ultimately creating a tree-like structure of decision nodes and leaf nodes. Each decision node represents a test on a specific feature, and the edges leading from the node correspond to the possible outcomes of that test. The leaf nodes represent the final predictions or decisions made by the model.

The benefits of DT include, it is easy to interpret, Little to no data preparation is required, is more flexible [58], can handle missing values, and can provide a measure of feature importance.

The disadvantage of DT is, it is easy to overfit, cannot deal with linear relationships, and lacks smoothness [57]. It is less stable, and in classification tasks, decision trees tend to favor features with more unique values or those that have a larger number of instances in the dataset.

5.3.2.3.2 Adaptive Boosting

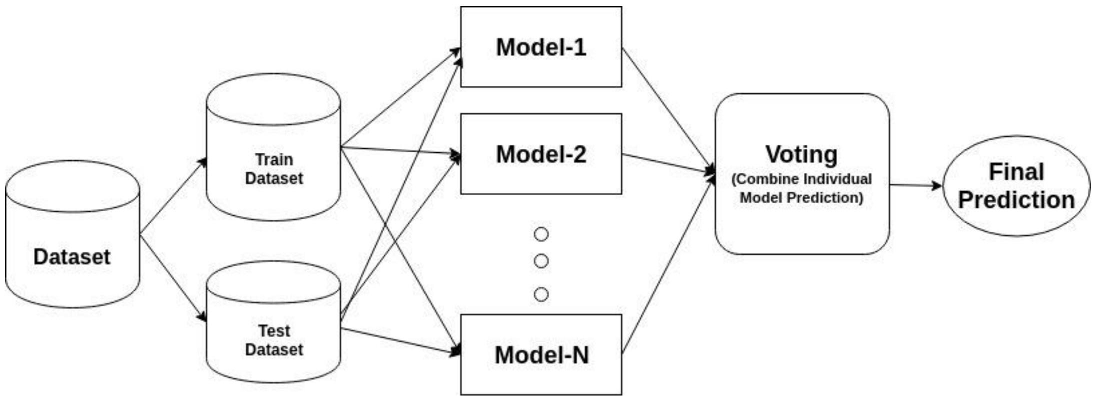

Fig 5.10 Structure of Adaptive Boosting [59]

Adaptive Boosting (AdaBoost) [53] is an ensemble ML algorithm that is used for both classification and regression tasks.

The main idea behind AdaBoost is to combine the predictions of multiple weak learners to create a strong learner. The weak learners are trained sequentially, and more emphasis is given to misclassified samples in each subsequent iteration to improve the overall performance. Different from RF which reduces model variance, AdaBoost focuses on reducing model bias [60].

It is a powerful algorithm that often performs well in practice. It can adapt to complex decision boundaries and handle noisy datasets.

The disadvantage of AdaBoost includes, it is sensitive to noise data and outliners, has the potential of overfitting, the calculation speed is slower compares to RF, potentially limiting the diversity of the ensemble and reducing its overall performance, and is difficult to parallelize.

5.3.2.3.3 Support-Vector Machine

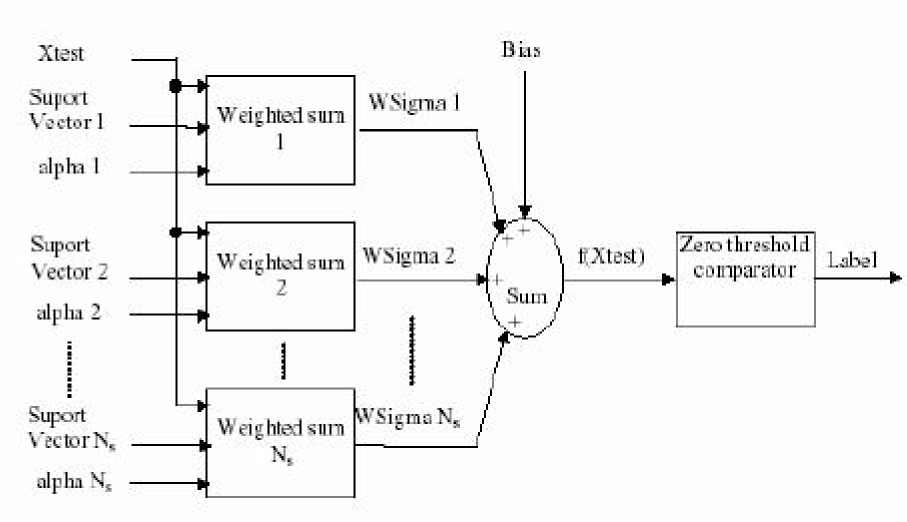

Fig 5.11 Block diagram of the time-continuous SVM classifier [61]

Support Vector Machine (SVM) [54] is a popular supervised ML algorithm used for classification tasks. The primary goal of SVM is to find the optimal hyperplane that best separates data points of different classes in a feature space. In binary classification, this hyperplane is a decision boundary that maximizes the margin between the two classes, while in multi-class classification, SVM constructs multiple decision boundaries to separate different classes.

It is effective in high-dimensional spaces and with small sample sizes[62], can handle both linear and non-linear decision boundaries using the kernel trick. It works well for binary and multi-class classification tasks. And it is robust against overfitting, especially with appropriate parameter tuning.

The disadvantage of SVM includes, it is computationally expensive, especially for large datasets. It requires careful selection of the kernel function and tuning of hyperparameters. And it can be sensitive to noisy or overlapping data.

5.3.2.3.4 K-Nearest Neighbor

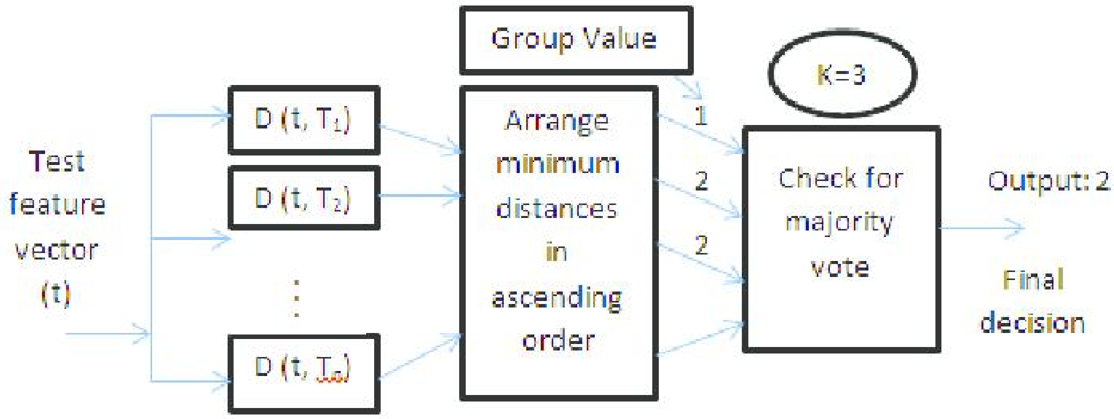

Fig 5.12 Block diagram for k-NN classifier [66]

K-Nearest Neighbor (k-NN) [52] is a simple and intuitive ML algorithm used for classification tasks. It is a non-parametric, instance-based learning algorithm, meaning it doesn't make strong assumptions about the underlying data distribution. Instead, it makes predictions based on the similarity between the input data and the labeled data points in the training set.

It is simple and easy to implement. It has no training phase, as the entire training dataset is used directly during prediction [67]. It is effective for both classification and regression tasks, especially in cases where data distributions are non-linear or complex.

The disadvantage of k-NN includes it is expensive during prediction, sensitivity to the choice of distance metric and K value, and it is not suitable for high-dimensional data due to the "curse of dimensionality," where the distance metric becomes less meaningful in high-dimensional spaces.

5.3.3 Cross-Validation

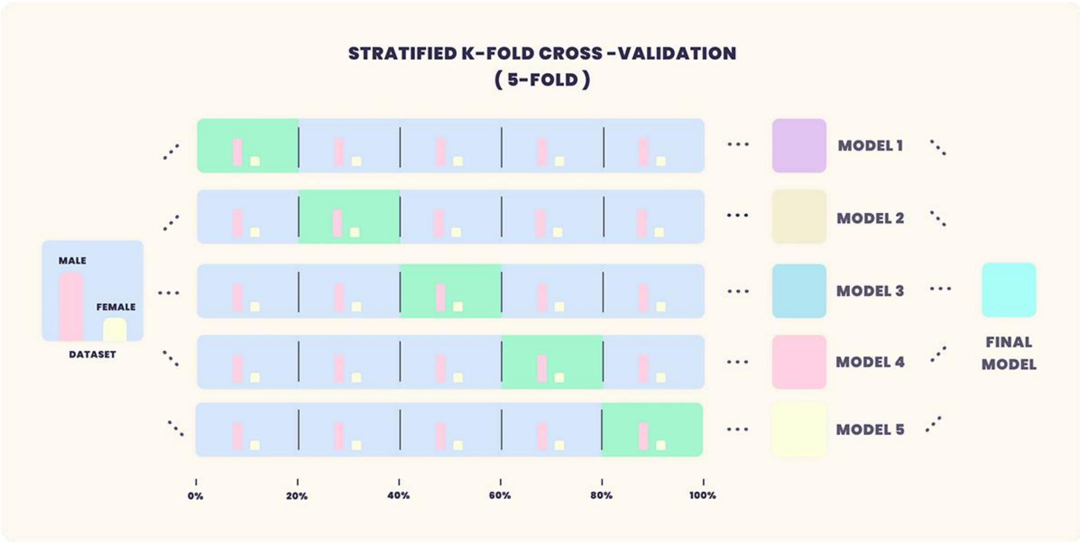

Fig 5.13 Stratified k-Fold Cross Validation (5-fold) [56]

Cross-validation[45][55] is a popular technique used to evaluate the performance of an ML model. It helps to assess how well the model generalizes to unseen data and provides a more reliable estimate of the model's performance compared to a single train-test split. The process involves splitting the dataset into ten subsets, or "folds," and iteratively training and testing the model on different combinations of these subsets.

This research used K-fold cross-validation for experiments, K = 10 is the most commonly used way. It is a practical method for dividing data samples into smaller subsets, each fold has its training data and test data, and each fold gives out its result. The final result is the average. For example, the entire dataset be divided into {A1, A2, A3, ..., A10}, the first fold might choose A1-A9 as training data, and A10 as validation data; the second fold might choose A9 as validation data, others training data. So each fold has different results.

5.3.3.1 Process of 10-Fold Cross-Validation

Data preparation: Start by dividing the original dataset into ten equal-sized subsets or folds. Each fold represents roughly 1/10th of the entire dataset.

Training and testing iterations: For each iteration (or "fold"), one fold is used as the test set, and the remaining nine folds are used as the training set. The model is trained on the training set and evaluated on the test set using a chosen evaluation metric (e.g., accuracy, precision, recall, or mean squared error, depending on the type of task - classification or regression). The performance metric obtained during this iteration is recorded.

Aggregating results: After all ten iterations are completed, there will be ten different performance metrics, one for each fold. The performance metrics are then averaged to obtain a single, overall performance metric, which is a more robust estimate of the model's performance.

5.3.3.2 Benifits of 10-Fold Cross-Validation

Better model evaluation: Cross-validation provides a more reliable estimate of the model's performance compared to a single train-test split. It reduces the risk of obtaining overly optimistic or pessimistic estimates of the model's performance.

Reduced overfitting: By evaluating the model on multiple validation sets, cross-validation helps detect and mitigate overfitting.

Optimal parameter tuning: It allows assessing how different hyperparameter settings impact the model's performance and helps in selecting the best hyperparameters that generalize well.

Maximizing data utilization: In traditional train-test splits, a portion of the data is set aside for testing, potentially reducing the amount of data available for training. Cross-validation ensures that all data points are used for both training and validation.

More robust model assessment: The results of cross-validation are less sensitive to the randomness in data partitioning compared to a single train-test split. This makes the model evaluation more robust and reliable.

Insight into model variance: Cross-validation provides information about how stable the model's performance is across different data subsets. A high variance between folds may indicate that the model is sensitive to the data distribution.

Handling imbalanced datasets: In scenarios with imbalanced class distributions, cross-validation can ensure that each fold contains a representative proportion of each class, leading to a more accurate performance evaluation.

Efficient use of data for small datasets: In situations where the dataset is limited, cross-validation allows better utilization of available data to obtain a more realistic assessment of model performance.

6 Research Result

This research did four experiments: for sixteen face types, for the four groups with the largest amount of data, the four types of four features, and a binary classification of each set of features.

6.1 Research Result for 16 Face Types

For both gender, the RF classifier's result is, the average is 28%, and the forth fold with the highest accuracy is 34.6%, lowest is the ninth fold, which got 21.1%. Adaboost works ok which got 26.4% accuracy as average.

Fold	Female + Male				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.148662	0.271556	0.254708	0.191278	0.18781
2	0.16551	0.32111	0.315659	0.190783	0.182854
3	0.166501	0.339445	0.31665	0.190783	0.204658
4	0.169475	**0.345887**	0.326561	0.190783	0.204163
5	0.1556	0.317641	0.323092	0.191278	0.191774
6	0.152131	0.300297	0.27106	0.191278	0.199207
7	0.150149	0.23885	0.214073	0.191774	0.16997
8	0.14668	0.230426	0.203171	0.191278	0.187314
9	0.140238	**0.210605**	0.200198	0.190783	0.170961
10	0.152626	0.231913	0.216056	0.191278	0.166501

Final	0.1547572	0.280773	0.2641228	0.1911296	0.1865212

Table 6.1 Training result for sixteen facial style for all samples

The red color means the highest accuracy in the ten folds of all the classifiers, and green means the lowest accuracy for RF. Yellow is the highest average accuracy for all 5 classifiers of all folds' results.

For both gender, the RF classifier's result is, the average is 28%, the fourth fold with the highest accuracy is 34.6%, lowest is the ninth fold, which got 21.1%. Adaboost works ok which got 26.4% accuracy as average.

DT, SVM, and k-NN all didn't reach 20%. DT works badly at this task, the accuracy is only 15.6% and the highest fold has 14.0% accuracy.

Fold	Female				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.135427	0.275761	0.251227	0.234544	0.190383
2	0.158979	0.2895	0.266928	0.234544	0.18842
3	0.165029	0.292731	0.25835	0.234774	0.209234
4	0.147348	0.286837	0.272102	0.234774	0.18664
5	0.160118	0.292731	0.280943	0.234774	0.201375
6	0.142436	**0.314342**	0.27112	0.234774	0.200393
7	0.141454	0.291749	0.277014	0.235756	0.176817
8	0.165029	0.295678	0.277996	0.236739	0.172888
9	0.133595	0.274067	0.24558	0.235756	0.189587
10	0.0962672	**0.234774**	0.224951	0.234774	0.136542
Final	0.14456822	**0.284817**	0.2626211	0.2351209	0.1852279

Table 6.2 Training result for sixteen facial style for female

SVM works better for both females only and both genders compare to other classifiers. DT still works the worst. All classifiers' results for the two experiments are similar except SVM raises 4.4% for females only. And SVM gets the most stable result for each fold, but its overall performance is not as good as RF.

In summary, RF got the best result, then is AdaBoost. DT works the worst, the second worst is k-NN. The final accuracy for each classifier in both experiments is similar.

6.2 Research Result for the 4 Types with Over 2,000 Data

From the previous part of this paper(5.2.3), four categories of facial styles have more than 2,000 people in them (Table 5.4), the four sets are (1,1,1,-1), (1,1,-1,-1), (-1,1,1,-1), (-1,1,-1,-1). They have to be trained individually in order to see how the result goes.

Fold	Female + Male				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.343295	0.439847	0.407663	0.295785	0.378544
2	0.340996	0.510345	0.500383	0.295785	0.380843
3	0.366564	0.532975	0.506902	0.296012	0.407975
4	0.368865	**0.534509**	0.520706	0.296012	0.423313
5	0.33819	0.516104	0.480828	0.296012	0.389571
6	0.332822	0.47546	0.43635	0.296012	0.375767
7	0.281442	0.378067	0.360429	0.296779	0.32362
8	0.315184	0.355828	0.338957	0.295245	0.338957
9	0.288344	0.341258	0.330521	0.295245	0.319018
10	0.302147	**0.370399**	0.375	0.295245	0.316718
Final	0.3277849	**0.4454792**	0.4257739	0.2958132	0.3654326

Table 6.3 Training result of both male and female four facial styles with with over 2,000 data

The training result is about 16% higher than the result for the sixteen types since the four groups all have more samples. The best fold from RF obtained 53.5%, but the worst got 37.0% of accuracy. SVM is still stable for each folder, but it works the worst this time, only achieving about 29.6%.

Fold	Female				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.386555	0.462185	0.442017	0.403361	0.394958
2	0.371429	0.487395	0.452101	0.403361	0.403361
3	0.366387	0.515966	0.478992	0.401681	0.447059
4	0.391597	0.502521	0.477311	0.401681	0.405042
5	0.381513	0.492437	0.497479	0.401681	0.421849
6	0.389916	**0.515966**	0.467227	0.401681	0.418487
7	0.363025	0.510924	0.480672	0.401681	0.448739
8	0.383838	0.520202	0.473064	0.40404	0.424242
9	0.340067	0.451178	0.429293	0.40404	0.39899
10	0.282828	**0.400673**	0.3367	0.402357	0.326599
Final	0.3657155	**0.4859447**	0.4534856	0.4025564	0.4089326

Table 6.4 Training result of female only for four facial styles with over 2,000 data

The experience result is about 20% higher than the result for the sixteen types. The fold with the highest accuracy of RF reached 51.6%.

In summary, RF and AdaBoost give the top 2 results. DT still works the worst since it is the most simple classifier with high instability, and the calculations involved can also become complex compared to other algorithms. SVM is still the most stable classifier, and its overall performance is still not as good as RF and AdaBoost.

From the comparison of females only and both genders' results for all the classifiers, it can be assumed that the dataset for males works worse than for females.

6.3 Research Result for the 4 Types of 4 Features

The previous chapter of this paper(3.3.1) shows there are four groups of the sixteen types, they used only two sets of features, {adult, childish} and {wide, narrow}, each attribute has 1 and -1, so the four groups are (1,1), (1,-1), (-1,1), (-1,-1).

Fold	Female + Male				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.314668	0.443013	0.429138	0.295342	0.338949
2	0.317641	0.47671	0.477205	0.295342	0.352329
3	0.324083	**0.494054**	**0.496036**	0.295342	0.368186
4	0.299306	0.412289	0.385035	0.295837	0.319623
5	0.268087	0.349851	0.357284	0.294846	0.323588
6	0.278989	0.388503	0.379584	0.295342	0.318632
7	0.263627	0.280971	0.278989	0.294351	0.282953
8	0.258176	**0.276511**	**0.251239**	0.294846	0.276511
9	0.272052	0.39445	0.400396	0.294846	0.339941
10	0.252725	0.283944	0.289891	0.294846	0.272547
Final	0.2849354	**0.3800296**	**0.3744797**	0.295094	0.3193259

Table 6.5 Training result of both gender, for {adult, childish} and {wide, narrow} feature

For {adult, childish} and {wide, narrow} facial features of both genders, since RF and AdaBoost's accuracy are similar, the significant values for both classifiers are marked with colors.

Adaboost got the best result at the second fold, there's only a very slight difference compared to RF, only about 0.2%. RF still gets the best result for overall accuracy, though it's only about 0.56% higher than AdaBoost.

Fold	Female				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.321884	0.443572	0.417076	0.316977	0.363101
2	0.315996	0.458292	0.416094	0.316977	0.342493
3	0.308448	0.469548	0.448919	0.317289	0.340864
4	0.314342	**0.51277**	0.47446	0.318271	0.359528
5	0.302554	0.444008	0.391945	0.318271	0.361493
6	0.292731	0.345776	0.350688	0.317289	0.300589
7	0.239686	0.355599	0.348723	0.317289	0.293713
8	0.297642	0.358546	0.353635	0.317289	0.278978
9	0.246562	**0.277996**	**0.275049**	0.317289	0.281925
10	0.24165	0.295678	0.280943	0.317289	0.256385
Final	0.2881495	**0.3961785**	0.3757532	0.317423	0.3179069

Table 6.6 Training result of female only, for {adult, childish} and {wide, narrow} feature

For this experience of females only, RF got the best result, Adaboost has 2% lower than RF. The eighth fold got the lowest accuracy for both RF and AdaBoost, at only 27.8% and 27.5%. SVM and k-NN both exceed 30%.

Fold	Male				
CV	DT	RF	AdaBoost	SVM	k-NN

1	0.293	**0.377**	0.363	0.315	0.274
2	0.26	0.356	0.344	0.315	0.272
3	0.3	0.356	0.349	0.314	0.296
4	0.289	0.359	0.343	0.315	0.257
5	0.274	0.376	0.345	0.315	0.28
6	0.283	0.35	0.353	0.315	0.277
7	0.279	0.358	0.325	0.314	0.273
8	0.266	0.326	0.314	0.314	0.263
9	0.268268	**0.276276**	0.253253	0.314314	0.259259
10	0.234234	0.276276	0.263263	0.314314	0.258258
Final	0.2746502	**0.3410552**	0.3252516	0.3145628	0.2709517

Table 6.7 Training result of male only, for {adult, childish} and {wide, narrow} feature

For males only, Adaboost and RF both obtained relatively higher results compared to other classifiers.

Besides, females only obtained a higher result, the highest fold get about 51.2% by RF model and about 40% for the result, which is higher than the result for both gender. Males get lower accuracy at about 34%. This time, DT works better than k-NN.

In summary, the research result is similar to the previous two sets of experiments. RF works best, then is Adaboost. DT works the worst, and k-NN works worse than DT for males only.

The result for the female is 5.5% higher than males, so the assumption in chapter 6.2 should be right, that female's data is labelled slightly better than males.

6.4 Binary Classification

Since all the previous experiments did not get well results, it is important to see the result of each set of features. So binary classification has been implemented for the experiment. The following are the results of the experiment.

Fold	Defined / Soft				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.616947	0.746283	**0.749257**	0.746779	0.681863
2	0.625372	0.74331	0.744797	0.746779	0.681863
3	0.644698	**0.746779**	0.735382	0.746779	0.68335
4	0.60555	0.74331	0.739841	0.746779	0.684341
5	0.601586	0.743806	0.734886	0.746779	0.659564
6	0.632309	0.745292	0.746779	0.746779	0.680377
7	0.61893	0.736373	0.726462	0.746779	0.66551
8	0.612488	0.74331	0.741328	0.746779	0.682359
9	0.60555	0.740833	0.729931	0.746779	0.673935
10	0.622894	0.745292	0.744797	0.746283	0.677403
Final	0.6186324	**0.743458**	**0.739346**	**0.7467294**	0.6770565

Table 6.8 Binary classification of defined and soft feature

For individual folds of defined and soft facial features, AdaBoost got the best result at 74.9%, which is slightly higher than RF (74.8%). The result is quite stable for each fold for RF, AdaBoost, k-NN, and SVM, especially SVM. DT is not stable because it is easy to be overfitting, small variations in the data might result in a completely different tree being generated [68].

Fold	Adult / Childish				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.567393	0.677899	0.671952	0.502478	0.625372
2	0.602081	**0.727453**	**0.727948**	0.501487	0.642716
3	0.589197	0.726462	0.709118	0.502478	0.652626
4	0.585728	0.720515	0.712587	0.502478	0.632805
5	0.55996	0.630823	0.627849	0.501487	0.583746
6	0.511397	**0.546581**	0.546581	0.501487	0.521308
7	0.575818	0.690783	0.680872	0.501982	0.629832
8	0.59217	0.71556	0.703171	0.500496	0.620416
9	0.563925	0.606541	0.607037	0.501487	0.557978
10	0.531219	0.598612	0.603072	0.501982	0.548563
Final	0.5678888	**0.664122**	0.6590187	0.5017842	0.6015362

Table 6.9 Binary classification of adult and childish feature

For individual folds for adult and childish facial features, AdaBoost got the best result at 72.8%, which is slightly higher than RF (72.7%). For the final result, RF works the best (66.4%), which is slightly higher than AdaBoost (65.9%). K-NN also exceeded 60%. The result is quite stable for each fold for SVM, though it works the worst.

Fold	Masculine / Feminine				
CV	DT	RF	AdaBoost	SVM	k-NN

1	0.616452	0.806244	0.797324	0.806244	0.681863
2	0.655104	**0.806739**	0.796333	0.806244	0.70664
3	0.660059	0.804757	0.803766	0.806244	0.73439
4	0.686819	0.806244	0.800793	0.806244	0.740833
5	0.671457	0.806244	0.803271	0.806244	0.730426
6	0.706145	0.806244	0.804262	0.806244	0.776016
7	0.692765	0.805253	0.800793	0.806244	0.75669
8	0.734886	0.805748	0.804757	0.806244	0.801288
9	0.716551	0.805748	0.804757	0.805748	0.800297
10	0.673935	0.805748	0.803271	0.805253	0.730426
Final	0.6814173	**0.805896**	**0.8019327**	**0.8060953**	0.7458869

Table 6.10 Binary classification of masculine and feminine feature

The accuracy for masculine and feminine facial features reached 80% with RF, Adaboost, and SVM. SVM is slightly higher than RF, about 0.02%, and they're 0.4% higher than Adaboost.

The result of each fold doesn't differ much for RF, Adaboost, and SVM. DT is from 61.6% at the first fold as the lowest to 73.5% at the eighth fold as the highest. K-NN's lowest accuracy is the first fold which got 68.2% and the highest is the eighth fold, 80.0%.

Fold	Wide / Narrow				
CV	DT	RF	AdaBoost	SVM	k-NN
1	0.542121	0.617939	0.616452	0.524281	0.558969
2	0.524281	0.658077	0.648662	0.525273	0.5555

3	0.526759	0.683845	0.668484	0.525273	0.558969
4	0.521804	0.654113	0.651635	0.525768	0.560951
5	0.518831	0.602577	0.603568	0.525273	0.551041
6	0.49554	0.501982	0.517344	0.525768	0.512389
7	0.475223	0.392963	0.407334	0.524777	0.461843
8	0.431615	0.286422	0.335481	0.525273	0.408821
9	0.527255	0.61893	0.593657	0.524777	0.546085
10	0.485134	0.454906	0.451437	0.525273	0.49554
Final	0.5048563	0.547175	0.5494054	0.5251736	0.5210108

Table 6.11 Binary classification of wide and narrow feature

For wide and narrow facial features, AdaBoost got the best result (54.9%), which is slightly higher than RF (54.7%). The best fold belongs to the third fold of RF, which is 68.4%. And the worst fold also belongs to RF, which is 28.6%.

All the classifiers work the worst for the eighth fold except k-NN. However, the result run directly from the code, so the training set and testing set didn't record, and it is hard to analyze which testing set belongs to the eighth fold.

In summary, {masculine, feminine} has the highest accuracy for this experience (81%), then {defined, soft} also worked well (74%). {adult, childish} achieved 66%. The result for {wide, narrow} is not good, it gets 55%. The overall accuracy is an average of the four sets, 69%.

The binary classification works the best of the four terms of experiences, second place is the experiment with the four sets with the largest data. The 16 types one has the lowest accuracy.

7 Discussion

This chapter will summarize the key findings, compare the results to previous studies, discuss the reason for unexpected results, restate the research objectives, and talk about its application field.

7.1 Comparison of Previous Studies

Due to the differences in categorization, this study cannot be fully compared with previous studies. Here just show three examples for broad comparison.

Shape	MLP	Naïve Bayes	KNN	SVM-LIN	SVM-RBF	Random Forest	AdaBoost
Heart	76%	54%	58%	60%	74%	68%	59%
Long	92%	64%	77%	83%	95%	83%	79%
Oval	66%	26%	49%	59%	78%	63%	37%
Round	78%	51%	47%	67%	75%	66%	44%
Square	89%	48%	62%	78%	86%	66%	56%
Overall	80%	49%	59%	70%	82%	70%	59%

Fig 7.1 Test accuracies comparison using dataset with 4960 images [3]

This research has six categories, SVM-RBF works the best, the overall accuracy achieved 82%, and long shape got 95% of accuracy.

Algorithm	Accuracy(%)	
	Training Set	Test Set
LDA	64.89	58.00
ANN	66.89	60.00
SVM-Linear	66.67	64.00
SVM-RBF	**70.67**	**72.00**

Fig 7.2 PERFORMANCES OF MACHINE LEARNING ALGORITHMS USING THE PROPOSED FEATURES [14]

The above research used five face shape classifications (oval, heart, square, oblong, round), it also used SVM-RBF, and the accuracy is about 71%.

Method	Accuracy	Dataset
Inception V3 [8]	84.4%	500 image [8]
Region Similarity, Correlation and Fractal Dimensions [6]	80%	Caltech 450 [16], Cuhk 260 [17], Lfw 200 [11]
Active Appearance Model (AAM), segmentation, and SVM [2]	72%	1000 image [2]
Hybrid approach VGG and SVM [7]	70.3%	500 image [7]
3D face data and SVM [3]	73.68%	209 3D image [3]
Geometric features [4]	80%	400 image [4]
Probability Neural Network and Invariant Moments [5]	80%	120 image [5]
Inception3 CNN	75.2%	500 image [8]
Inception3 CNN+HOG	76.9%	
Inception3 CNN+LMs	78.2%	
Inception3 CNN+HOG+LMs	**81.1%**	

Fig 7.3 THE PERFORMANCE OF PROPOSED FACE SHAPE CLASSIFICATION METHODOLOGY COMPARING TO THE EXISTING METHODS IN LITERATURE. [20]

The above research used the same shape classification method as Fig 7.2, its accuracy is about 81%.

Compare with the previous research's result, this study works worse. However, another thing is, Number of categories is related to the final accuracy. Here's an example.

	Method	Exact	1-off
1	Shan 2010 [74]	55.9	87.7
2	Alnajar et al. 2012 [54]	59.5	–
	Our system, no alignment		
3	LBP	56.4 ± 0.4	92.7 ± 0.2
4	FPLBP	60.2 ± 0.6	92.0 ± 0.2
5	LBP+FPLBP	57.5 ± 0.5	93.7 ± 0.2
6	LBP+FPLBP+Dropout 0.5	62.7 ± 0.6	94.5 ± 0.1
7	LBP+FPLBP+Dropout 0.8	65.6 ± 0.6	94.0 ± 0.2
	Our system, including alignment		
8	LBP	58.0 ± 0.1	94.1 ± 0.2
9	FPLBP	61.0 ± 0.5	92.2 ± 0.2
10	LBP+FPLBP	59.0 ± 0.4	**95.3** ± **0.2**
11	LBP+FPLBP+PCA 0.5	46.8 ± 0.6	90.1 ± 0.3
12	LBP+FPLBP+PCA 0.8	41.3 ± 0.4	83.9 ± 0.3
13	LBP+FPLBP+Dropout 0.5	64.3 ± 0.6	**95.3** ± **0.3**
14	LBP+FPLBP+Dropout 0.8	**66.6** ± 0.7	94.8 ± 0.3

Fig 7.4 AGE CLASSIFICATION ON THE GALLAGHER BENCHMARK. MEAN ACCURACY (± STANDARD ERRORS) OVER THE SEVEN AGE CATEGORIES IN THE GALLAGHER BENCHMARK. BOLDFACE ARE BEST SCORING [69]

	Method	Exact	1-off
1	LBP	41.4 ± 2.0	78.2 ± 0.9
2	FPLBP	39.8 ± 1.8	74.6 ± 1.0
3	LBP+FPLBP	44.5 ± 2.3	**80.7** ± 1.1
4	LBP+FPLBP+PCA 0.5	38.1 ± 1.4	75.5 ± 0.9
5	LBP+FPLBP+PCA 0.5	32.9 ± 1.6	67.7 ± 1.1
6	LBP+FPLBP+Dropout 0.5	44.5 ± 2.2	80.6 ± 1.0
7	LBP+FPLBP+Dropout 0.8	**45.1** ± 2.6	79.5 ± 1.4

Fig 7.5 AGE CLASSIFICATION ON THE ADIENCE BENCHMARK. MEAN ACCURACY (± STANDARD ERRORS) OVER THE EIGHT AGE CATEGORIES IN THE ALIGNED ADIENCE SET. BOLDFACE ARE BEST SCORING [69]

The above research is a study using facial recognition to identify age. "exact " is the mean accuracy, across all age groups, of predicting the true age label. "1-off " implies counting labeling errors, one age group removed from the true label, as correct

[69]. Fig 7.4 used seven age categories, Fig 7.5 is eight. And the accuracy has dropped by about 20%. For 1-off, it also dropped about 15%.

Therefore the final accuracy for sixteen types is explainable, though it is low. And here are some other reasons for low accuracy.

7.2 Possible Reasons for Low Accuracy

Ambiguity in face shapes: Certain face shapes can exhibit ambiguity, making it challenging even for human experts to categorize them accurately. During the data labeling process, there was a problem that many images' facial features are in between, hard to determine which feature they belongs to, making it difficult for the algorithm to make precise classifications.

Increased complexity: As the number of face types increases, the classification task becomes more complex. It might be hard for machines to distinguish subtle differences between closely related face shapes, resulting in misclassifications and reduced accuracy.

Class imbalance: Here are 16 face types, the data is imbalanced. For example, only 238 data in the (-1,-1,-1,1) dataset, is significantly fewer than others, leading the model to be biased towards the majority classes and affecting its ability to accurately classify the minority classes.

Feature extraction challenges: With more types, identifying specific features that distinguish each face shape becomes more complex, potentially leading to errors in feature extraction and, consequently, misclassifications.

Overfitting: The model can be too specific to the training data and fails to generalize well to unseen data.

7.3 Research Objectives and Application Fields

This research is aiming to develop a face shape recognition system for personalized beauty and fashion recommendations and enhanced customer experience. Combine with the accuracy of the binary classification result, masculine/feminine got 80%, and defined/ soft got 74%. Though the 16 styles result is not good enough, these 2 sets work not badly, and these feature classification methods already be widely used in Asia. One feature set also has its valid application areas.

Start with defined and soft, Shiseido is marketing its cosmetics that are suitable for defined facial features[70] and soft facial features[71]. Also, makeup professionals give detailed explanations of what kind of makeup is suitable for the defined feature[73] and soft feature[72]. So this feature set can be used efficiently.

Then, for masculine and feminine facial features, they also have detailed explanations from makeup and fashion professionals[74][75].

Also, these two sets of features can be combined like masculine x defined, feminine x soft, etc[76]. There are also many instructions on how to makeup of the combined features [46][77].

All references mentioned in this chapter have used a lot of space on how to distinguish which type people belong to. So this research can be valid and helpful for this situation.

8 Future Work

This chapter will explain how the existing shortcomings of this study can be improved in the future, how the missing pieces can be filled in, and the future extensibility of this study.

8.1 For Improve the Accuracy

Face type is closely related to fashion, beauty, and attractiveness, people always get confused about which shape they belong to, so an automatic face type recognition system is useful. But because of the low accuracy, there is still much work that needs to do, a higher accuracy is important for a system, otherwise, it is not useful.

Here are the next steps.

Data annotation: The data has a big problem since there are no existing labeled datasets available. Due to funding and time constraints, I will shrink the dataset to 1,000 images, and let more than 5 people vote for the result for data annotation, to get a relatively correct dataset.

Keep the data balance: Balance of data is important, each feature should have a similar number of data.

Decrease the complexity: Use the new dataset to do the binary classification again. The research should be step-by-step, it cannot proceed to the complex classification method unless the binary classification's accuracy is good enough.

Model selection: Choose the classifier that is more suitable for binary classification. Hybrid models are also a valuable choice. Investigating the combination of multiple classification algorithms may help improve accuracy.

By addressing these aspects in future research, it is possible to develop an accurate and reliable face shape recognition system.

8.2 For Application Field

Though some previous research talked about the application fields of males, it is not enough. Therefore finding more application areas that serve men is a necessity.

In addition, since the research methodology of this study is mainly about machine learning for face shape recognition, while it is intended to be put into practical applications. So a user-oriented program for computers or smartphones is necessary.

Furthermore, there are more application scenarios for face type recognition. For example, virtual try-on, makeup simulation, virtual styling, and fashion consultation, etc. There are many ways to integrate face shape recognition technology into the beauty and fashion industry, And all of these applications help to increase customer satisfaction, which in turn increases people's happiness value.

8.3 Conclusion

Personalization has become a cornerstone of the beauty and fashion industry, and ML driven facial analysis plays a pivotal role in this aspect. By accurately identifying individual face shapes and features, beauty brands can offer tailored recommendations, ensuring that customers find products and styles that complement their specific facial characteristics, ultimately leading to higher satisfaction and brand loyalty. Individuals can make informed decisions about clothing styles, hairstyles, and accessories that best complement their features. Enhance their overall appearance and develop a personal style that suits them, and help individuals stay updated with the latest fashion trends.

This study delves into the impact of automatic face style recognition on the beauty and fashion industry, analyzing face types with ML and giving the corresponding application areas, and analyzes the importance and usefulness of integrating ML face style recognition and the fashion and beauty industry.

In conclusion, the integration of ML in face style recognition has paved the way for a more inclusive, personalized, and engaging beauty and fashion experience, benefiting both consumers and industry stakeholders. This thesis contributes to the growing body of knowledge in this field and serves as a foundation for continued exploration and progress in the dynamic intersection of machine learning, beauty, and fashion.

9 Acknowledgement

First and foremost, I am deeply grateful to my professor and my research supervisor Tetsuro Ogi for his continuous and patient teaching during the two years. He gave me lots of valuable advice and suggestions for my research from topic selection to completing the thesis paper. He has helped me to observe problems from a rational and objective perspective, his contribution has a huge impact on my research and life.

Thanks to my secondary advisor Niitsuma Masahiro for his extremely valuable advice. It gives me a new understanding of my experiments.

Thanks to all SDM faculties for their passionate and serious education.

Thanks to everyone in Ogi lab that listen to my presentations and gave me ideas.

Thanks to all of my friends in SDM for giving me advice on thesis writing, and encouraging me.

Thanks to my undergraduate college roommates for inspiring me to choose a topic for my thesis.

Thanks to Fengling Li for allowing me to do my research using the face categorization method she summarized.

Thanks to my mother that sponsored me to do the thesis and support me throughout the two years of my master's study, she has given me immeasurable emotional value.

Without your collective efforts and support, this project would not have been possible. Thank you for being an essential part of this endeavor.

Sincerely,

Xuanqi Feng

August. 2023

References

[1] Roberts, R. (2022) 2022 Beauty Industry Trends & Cosmetics Marketing: Statistics and strategies for your ecommerce growth, Common Thread Collective. Available at: https://commonthreadco.com/blogs/coachs-corner/beauty-industry-cosmetics-marketing-ecommerce

[2] Joanne Entwistle, "The aesthetic economy: The production of value in the field of fashion modelling, " in Journal of Consumer Culture, 2002.

[3] Hossam, M. et al. (2021) "A comparative study of different face shape classification techniques, " 2021 International Conference on Electronic Engineering (ICEEM) [Preprint]. Available at: https://doi.org/10.1109/iceem52022.2021.9480638

[4] Kaur, J. (2023) The best glasses for Your face shape shape: Jins Eyewear, JINS. Available at: https://www.jins.com/us/blog/best-glasses-for-your-face

[5] ZergNet. "Find the perfect cut for your face shape, " Available: http://www.instyle.com/hair/find-perfect-cut-your-face-shape

[6] Nguyen, T.V. et al. (2012) "Sense beauty via face, dressing, and/or voice, " Proceedings of the 20th ACM international conference on Multimedia [Preprint]. Available at: https://doi.org/10.1145/2393347.2393385

[7] Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. Journal of Personality and Social Psychology, 24, 285–290.

[8] TodayShow (2012) Hey, Good Lookin', your job outlook lookin' good, TODAY.com. TODAY. Available at: https://www.today.com/id/wbna49350507#.URiLAKX6MUU

[9] Research shows physical appearance affects career success in Economics (2021) UC Santa Cruz News. Available at: https://news.ucsc.edu/2021/04/attractiveness-career-success-economics.html

[10] Duan, J. et al. (2022) "Face shape classification based on bilinear network with attention mechanism, " Journal of Physics: Conference Series, 2278(1), p. 012041. Available at: https://doi.org/10.1088/1742-6596/2278/1/012041

[11] Chen, W.-Y., Chen, J.-L. and Chen, L.-G. (2014) "On-the-fly fashion photograph recommendation system with robust face shape features, " 2014 IEEE International Conference on Consumer Electronics (ICCE) [Preprint]. Available at: https://doi.org/10.1109/icce.2014.6776105

[12] Seven basic face shapes - westmorewestmore Westmore. Available at: https://www.westmoresofhollywood.com/seven-basic-face-shapes

[13] Simon, M. (2014) Scientists have identified nine distinct face shapes five new groups, Daily Mail Online. Available at: https://www.dailymail.co.uk/news/article-2795014/scientists-identified-nine-distinct-face-shapes-five-new-groups-traditional-oval-round-heart-square.html

[14] Sunhem, W. and Pasupa, K. (2016) ‘An approach to face shape classification for hairstyle recommendation', 2016 Eighth International Conference on Advanced Computational Intelligence (ICACI) [Preprint].

doi:10.1109/icaci.2016.7449857.

[15] Gould, W.R. (2023) How to find your face shape, Real Simple. Available at: https://www.realsimple.com/beauty-fashion/what-is-my-face-shape

[16] Harmeyer, C. (2022) How to find your face shape in 4 easy steps, according to experts, HelloGiggles. Available at: https://hellogiggles.com/what-is-my-face-shape/

[17] In-Cosmetics (2020) Male grooming and men's makeup on the rise: In-cosmetics connect, in. Available at: https://connect.in-cosmetics.com/regions/in-cosmetics-asia/male-grooming-and-mens-makeup-on-the-rise/

[18] トレンダーズ株式会社 (2022)【第1回 メンズ美容定点調査】男性が化粧をすることは「世間に受け入れられている」と半数近くが回答 実施率は「スキンケア:66.5%」「メイク:14.6%」, プレスリリース・ニュースリリース配信シェアNo.1｜PR TIMES. Available at: https://prtimes.jp/main/html/rd/p/000000091.000056551.html

[19] The Myers &Briggs Foundation - MBTI® basics. Available at: https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/

[20] Alzahrani, T., Al-Nuaimy, W. and Al-Bander, B. (2019) "Hybrid feature learning and engineering based approach for face shape classification, "2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS) [Preprint]. Available at: https://doi.org/10.1109/isacs48493.2019.9068910

[21] Ji, W. and Jin, L. (2021) "Face shape classification based on MTCNN and FaceNet, " 2021 2nd International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI) [Preprint]. Available at: https://doi.org/10.1109/ichci54629.2021.00042

[22] Darugar, M.J. and Kiong, L.C. (2017) "Gender estimation based on supervised hog, action units and unsupervised CNN feature extraction, " 2017 Artificial Intelligence and Robotics (IRANOPEN) [Preprint]. Available at: https://doi.org/10.1109/rios.2017.7956439

[23] Tabassum, F. et al. (2022) "Human face recognition with combination of DWT and machine learning, " Journal of King Saud University - Computer and Information Sciences, 34(3), pp. 546–556. Available at: https://doi.org/10.1016/j.jksuci.2020.02.002

[24] Alashkar, T., Jiang, S. and Fu, Y. (2017) "Rule-based facial makeup recommendation system, " 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) [Preprint]. Available at: https://doi.org/10.1109/fg.2017.47

[25] Sharzeel Saleem, Jeba Shiney, B. Priestly Shan, Vishesh Kumar Mishra, Face recognition using facial features, Materials Today: Proceedings, 2021, ISSN 2214-7853, https://doi.org/10.1016/j.matpr.2021.07.402

[26] Chandrakala, M. and Durga Devi, P. (2021) "Two-stage classifier for face recognition using hog features, " Materials Today: Proceedings, 47, pp. 5771–5775. Available at: https://doi.org/10.1016/j.matpr.2021.04.114

[27] Zhu, Y. and Jiang, Y. (2020) "Optimization of face recognition algorithm based

on deep learning multi feature fusion driven by Big Data, " Image and Vision Computing, 104, p. 104023. Available at: https://doi.org/10.1016/j.imavis.2020.104023

[28] Benedict, S.R. and Kumar, J.S. (2016) "Geometric shaped facial feature extraction for face recognition, " 2016 IEEE International Conference on Advances in Computer Applications (ICACA) [Preprint]. Available at: https://doi.org/10.1109/icaca.2016.7887965

[29] Benedict, S.R. and Kumar, J.S. (2016) "Geometric shaped facial feature extraction for face recognition, " 2016 IEEE International Conference on Advances in Computer Applications (ICACA) [Preprint]. Available at: https://doi.org/10.1109/icaca.2016.7887965

[30] Bhati, S. (2023) How to find your face shape - 9 types of face shapes, Fabbon. Available at: https://fabbon.com/articles/makeup/types-of-face-shapes

[31] Dosovitskiy, A. et al. (2021) An image is worth 16x16 words: Transformers for image recognition at scale, arXiv.org. Available at: https://arxiv.org/abs/2010.11929

[32] Datagen (2023) Top 10 face datasets for facial recognition and analysis, Datagen. Available at: https://datagen.tech/blog/face-datasets/

[33] Papers with Code - Jaffe Dataset. Available at: https://paperswithcode.com/dataset/jaffe

[34] Ox-vgg/vgg_face2, GitHub. Available at: https://github.com/ox-vgg/vgg_face2

[35] The peal face database. Available at: http://www.jdl.link/peal/index.html

[36] Liu, Z. et al. Large-scale celebfaces attributes (celeba) dataset, CelebA Dataset. Multimedia Laboratory. Available at: https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

[37] Cao, Q. et al. (2018) ‘VGGFACE2: A dataset for recognising faces across pose and age', 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018) [Preprint]. doi:10.1109/fg.2018.00020.

[38] L. Li, J. So, H.-C. Shin, and Y. Han. "An AAM-based face shape classification method used for facial expression recognition, " International Journal of Research in Engineering and Technology, vol. 2, pp. 164–168, 2013.

[39] Warfield, N. (2019) Men are a multibillion dollar growth opportunity for the Beauty Industry, CNBC. CNBC. Available at: https://www.cnbc.com/2019/05/17/men-are-a-multibillion-dollar-growth-opportunity-for-the-beauty-industry.html

[40] Punjabi, N.S. (2018) Is beauty defined by gender? these people prove that it isn't, Vogue India. Available at: https://www.vogue.in/magazine-story/is-beauty-defined-by-gender-these-people-prove-that-it-isnt/

[41] Innovators choose wonder (2020) Drag Queen Research - United States. Available at: https://start.askwonder.com/insights/drag-queen-research-united-states-z49v7ev27

[42] Yuxin Mao, Jing Zhang, Zhexiong Wan, Yuchao Dai, Aixuan Li, Yunqiu Lv,

Xinyu Tian, Deng-Ping Fan, and Nick Barnes. Transformer transforms salient object detection and camouflaged object detection. arXiv.org. Available at: https://arxiv.org/abs/2104.10127

[43] Naseer, M. et al. (2021) Intriguing properties of vision transformers, arXiv.org. Available at: https://arxiv.org/abs/2105.10497

[44] The structure of the RF algorithm. | download scientific diagram. Available at: https://www.researchgate.net/figure/The-structure-of-the-RF-algorithm_fig1_363447998

[45] Pandian, S. (2022) K-fold cross validation technique and its essentials, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2022/02/k-fold-cross-validation-technique-and-its-essentials/

[46] 男顔？女顔？あなたが美人に！可愛く！なれる「メイクのルール」って？｜今を一番輝く時に。30代を楽しむビューティメディア「byBirth」 2017) byBirth. Available at: https://bybirth.jp/press/archives/90615

[47]【顔タイプ】子供顔、大人顔の見分け方・特徴 (2023) 愛知・名古屋・オンラインの顔タイプ診断、骨格診断 ｜ ACCORD. Available at: https://accord-a.net/face-type/cild-adult-face

[48] Breiman, L. (2001) Random forests - machine learning, SpringerLink. Available at: https://link.springer.com/article/10.1023/a:1010933404324

[49] What is a decision tree IBM. Available at: https://www.ibm.com/topics/decision-trees

[50] J. R. Quinlan. 1986. Induction of Decision Trees. Mach. Learn. 1, 1 (March 1986), 81–106. https://doi.org/10.1023/A:1022643204877

[51] Ronny Kohavi and J. Ross Quinlan. 2002. Data mining tasks and methods: Classification: decision-tree discovery. Handbook of data mining and knowledge discovery. Oxford University Press, Inc., USA, 267–276.

[52] Duda, R.O. (2004) Pattern classification 2nd edition with Computer Manual 2nd Edition set. John Wiley & Sons.

[53] Freund, Yoav and Robert E. Schapire. "Experiments with a New Boosting Algorithm. " International Conference on Machine Learning (1996).

[54] Cortes, C., Vapnik, V. Support-vector networks. Mach Learn 20, 273–297 (1995). https://doi.org/10.1007/BF00994018

[55] Seymour Geisser & William F. Eddy (1979) A Predictive Approach to Model Selection, Journal of the American Statistical Association, 74:365, 153-160, DOI: 10.1080/01621459.1979.10481632

[56] 交差検証 (クロスバリデーション) : Cross Validation - AI・機械学習ナレッジセンター VARISTA. Available at: https://www.varista.ai/knowledge/cross-validation/

[57] Molnar, C. (2023) Interpretable machine learning, 5.4 Decision Tree. Available at: https://christophm.github.io/interpretable-ml-book/tree.html

[58] What is a decision tree. IBM. Available at: https://www.ibm.com/topics/decision-trees#:~:text=A%20decision%20tree%20i

s%20a,internal%20nodes%20and%20leaf%20nodes.

[59] Navlani, A. (2018) AdaBoost classifier algorithms using python Sklearn tutorial, DataCamp. Available at: https://www.datacamp.com/tutorial/adaboost-classifier-python

[60] Kurama, V. (2021) A guide to understanding AdaBoost, Paperspace Blog. Available at: https://blog.paperspace.com/adaboost-optimizer/

[61] Oltean, Gabriel & Gordan, Mihaela. (2005). TOWARDS ANALOG IMPLEMENTATION OF SUPPORT VECTOR MACHINES: A TIME-CONTINUOUS FORMULATION OF THE CLASSIFICATION PHASE.

[62] K, D. (2023) Top 4 advantages and disadvantages of support vector machine or SVM, Medium. Available at: https://dhirajkumarblog.medium.com/top-4-advantages-and-disadvantages-of-support-vector-machine-or-svm-a3c06a2b107

[63] Mbaabu, O. (2020) Introduction to random forest in machine learning, Section. Available at: https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

[64] Paul, M. (2018) Nonlinear Classification. Available at: https://cmci.colorado.edu/classes/INFO-4604/files/slides-9_nonlinear.pdf

[65] Takyar, A. (2023) How is a vision transformer (VIT) model built and implemented?, LeewayHertz. Available at:

https://www.leewayhertz.com/vision-transformer-model/

[66] Gupta, Sheifali & Singla, Chaitanya. (2016). Segmentation, Feature Extraction and Classification of Astrocytoma in MR Images. Indian Journal of Science and Technology. 9. 10.17485/ijst/2016/v9i36/102154.

[67] Kumar, N. (2019) Advantages and disadvantages of KNN algorithm in Machine Learning, Available at: http://theprofessionalspoint.blogspot.com/2019/02/advantages-and-disadvantages-of-knn.html

[68] Gupta, P. (2017) Decision trees in machine learning, Medium. Available at: https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052

[69] Eidinger, E., Enbar, R. and Hassner, T. (2014) ‘Age and gender estimation of unfiltered faces', IEEE Transactions on Information Forensics and Security, 9(12), pp. 2170–2179. doi:10.1109/tifs.2014.2359646.

[70] Shiseido Co., Ltd. (2021) 濃い顔をナチュラルに見せるメイクテク。2つのコツで簡単に印象チェンジ！: 美容の情報, ワタシプラス／資生堂. Available at: https://www.shiseido.co.jp/sw/beautyinfo/DB009486/

[71] Shiseido Co., Ltd.(2017) 2点を盛るだけでさりげなく華やかに！薄顔＆地味顔解消メイクテク: 美容の情報, ワタシプラス／資生堂. Available at: https://www.shiseido.co.jp/sw/beautyinfo/DB007213/

[72] 薄い顔はメイク映えする！『ナチュラルに盛れる』メイク術をパーツごとに伝授 (2021) MINE（マイン）｜ファッション動画マガジン. Available at:

https://mine-3m.com/articles/105546

[73] ぴよ (2022) 派手顔メイクはケバくなる？濃い顔を活かすナチュラルなメイク方法＆おすすめコスメをお届け, LIPS. Available at: https://lipscosme.com/articles/6376

[74] RAXY Style 編集部 (2023) 自分は男顔？女顔？特徴を活かした垢抜けメイク方法とおすすめの似合う髪型6選, RAXY(ラクシー). Available at: https://raxy.rakuten.co.jp/beautytopics/articles/2022/article453/

[75] RAXY Style 編集部 (2023)【女顔】の特徴ってどんなの？男顔との違いや華やか見えするメイク方法をお届け！, RAXY(ラクシー). Available at: https://raxy.rakuten.co.jp/beautytopics/articles/2022/article545/

[76] Taguchi, H. (2020) あなたは男顔?女顔?タイプ別似合うメイクのやり方, point house. Available at: https://point-house.jp/beauty/taguchi180211/

[77]【あなたは女顔or男顔？】目鼻立ちや骨格など、個性が活きる最強メイクまとめ: マキアオンライン(maquia online) (2020) MAQUIA ONLINE. Available at: https://maquia.hpplus.jp/makeup/matome/50695/16/

Appendix

Code for Data Analysis

Male Statistics

```
value_male = male.values

value_male_1 = value_male[:,1:]

male_count = count_type(value_male_1)
```

Female Statistics

```
value_female = female.values

value_female_1 = value_female[:,1:]

female_count = count_type(value_female_1)
```

Male and Female Statistics

```
value_male_female = np.vstack((value_male_1,value_female_1))

male_female_count = count_type(value_male_female)
```

Code for the System

Feature extraction by ViT model

```
import cv2

from glob import glob as gl

import os

import numpy as np

from transformers import ViTFeatureExtractor, ViTModel

import re


glob = lambda k: sorted(gl(k))

pjoin = os.path.join
```

```
path = '.\\male'
data = []
ind = []
for pt in glob(pjoin(path, '*.jpg')):
    ind.append(int(re.sub('\D','',pt)))
    img = cv2.resize(cv2.imread(pt), (200, 200))
    data.append(img)

feature_extractor =
ViTFeatureExtractor.from_pretrained('google/vit-base-patch16-224-in21k')
model = ViTModel.from_pretrained('google/vit-base-patch16-224-in21k')
image_feature = []
for j in range(len(data)):
    print(j)
    inputs = feature_extractor(images=data[j], return_tensors="pt")
    outputs = model(**inputs)
    feature = outputs.pooler_output.data.numpy()
    image_feature.append(feature)
image_feature = np.array(image_feature).reshape(len(data),-1)

np.savetxt('male_data_vit.txt',image_feature)
```

Deal with Data

```
import pandas as pd
import numpy as np
import re
import os
from glob import glob as gl
```

```
def count_type(array):
    count = np.zeros((1,16))
    for i in range(len(array)):
        if array[i][0]==1:
            if array[i][1]==1:
                if array[i][2]==1:
                    if array[i][3]==1:
                        count[0,0]+=1
                    else:
                        count[0,1]+=1
                else:
                    if array[i][3]==1:
                        count[0,2]+=1
                    else:
                        count[0,3]+=1
            else:
                if array[i][2]==1:
                    if array[i][3]==1:
                        count[0,4]+=1
                    else:
                        count[0,5]+=1
                else:
                    if array[i][3]==1:
                        count[0,6]+=1
                    else:
                        count[0,7]+=1
        else:
            if array[i][1]==1:
                if array[i][2]==1:
                    if array[i][3]==1:
```

```
                        count[0,8]+=1
                    else:
                        count[0,9]+=1
                else:
                    if array[i][3]==1:
                        count[0,10]+=1
                    else:
                        count[0,11]+=1
            else:
                if array[i][2]==1:
                    if array[i][3]==1:
                        count[0,12]+=1
                    else:
                        count[0,13]+=1
                else:
                    if array[i][3]==1:
                        count[0,14]+=1
                    else:
                        count[0,15]+=1
    return count


def label_type(array):
    label = np.zeros((len(array),1))
    for i in range(len(array)):
        if array[i][0]==1:
            if array[i][1]==1:
                if array[i][2]==1:
                    if array[i][3]==1:
                        label[i,0]=1
```

```
            else:
                label[i,0]=2
        else:
            if array[i][3]==1:
                label[i,0]=3
            else:
                label[i,0]=4
    else:
        if array[i][2]==1:
            if array[i][3]==1:
                label[i,0]=5
            else:
                label[i,0]=6
        else:
            if array[i][3]==1:
                label[i,0]=7
            else:
                label[i,0]=8
else:
    if array[i][1]==1:
        if array[i][2]==1:
            if array[i][3]==1:
                label[i,0]=9
            else:
                label[i,0]=10
        else:
            if array[i][3]==1:
                label[i,0]=11
            else:
                label[i,0]=12
```

```
        else:
            if array[i][2]==1:
                if array[i][3]==1:
                    label[i,0]=13
                else:
                    label[i,0]=14
            else:
                if array[i][3]==1:
                    label[i,0]=15
                else:
                    label[i,0]=16
    return label


def label_type_binary(array):
    label = np.zeros((len(array),1))
    for i in range(len(array)):
        if array[i][0]==1:
            if array[i][1]==1:
                label[i,0]=1
            else:
                label[i,0]=2
        else:
            if array[i][1]==1:
                label[i,0]=3
            else:
                label[i,0]=4
    return label
```

```
male = pd.read_excel(io='./face-male-0426.xlsx')
female = pd.read_excel(io='./face-female-0426.xlsx')
```

Male statistics

```
value_male = male.values
male_ind, female_ind = [], []
for i in range(len(value_male)):
    male_ind.append(int(re.sub('\D','',value_male[i,0])))

# male_label = label_type(value_male[:,1:])
# male_label_four = label_type_binary(value_male[:,[1,3]])
# np.savetxt('male_label_four.txt',male_label_four)
```

Female statistics

```
value_female = female.values
for i in range(len(value_female)):
    female_ind.append(int(re.sub('\D','',value_female[i,0])))

# female_label = label_type(value_female[:,1:])
# female_label_four = label_type_binary(value_female[:,[1,3]])
# np.savetxt('female_label_four.txt',female_label_four)
```

Image number extraction

```
glob = lambda k: sorted(gl(k))
pjoin = os.path.join
path = '.\img'

vit_ind = []
for pt in glob(pjoin(path, '*.jpg')):
```

```
    vit_ind.append(int(re.sub('\D','',pt)))
```

Loading dataset

```
vit_data = np.loadtxt('face_data_vit.txt',delimiter=' ')
vit_label = np.zeros((max(vit_ind),1))
# for i in range(len(female_ind)):
#     vit_label[female_ind[i]-1,0] = female_label_four[i]
# for i in range(len(male_ind)):
#     vit_label[male_ind[i]-1,0] = male_label_four[i]
c=4
for i in range(len(female_ind)):
    vit_label[female_ind[i]-1,0] = value_female[i,c]
for i in range(len(male_ind)):
    vit_label[male_ind[i]-1,0] = value_male[i,c]
vit_label = vit_label[np.nonzero(vit_label)[0]]

# np.savetxt('img_label_c1.txt',vit_label)
```

Train the classifier

```
from sklearn.ensemble import AdaBoostClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.model_selection import StratifiedKFold
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
import numpy as np
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
classifiers = [
    DecisionTreeClassifier(),
```

```
    RandomForestClassifier(),
    AdaBoostClassifier(),
    SVC(gamma=2, C=1),
    KNeighborsClassifier(3)]
acc = np.zeros((10,len(classifiers)))
data, label = vit_data, vit_label
np.where(label==10)
```

Code for Cross Validation

```
skf = StratifiedKFold(n_splits=10)
k=0
for train, test in skf.split(data, label):
    print(k)
    train_data, train_label = data[train,:], label[train]
    test_data, test_label = data[test,:], label[test]
    l=0
    for clf in classifiers:
        clf.fit(train_data, train_label)
        acc[k,l] = clf.score(test_data, test_label)
        l+=1
        k+=1

for features, targets in train_loader:
        Break
```

Previous dataset and its training result

This data has problem so the train result was worse, and the data had been check and modified.

Male (10009)

	1,1,1,1	1,1,1,-1	1,1,-1,1	1,1,-1,-1	1,-1,1,1	1,-1,1,-1	1,-1,-1,1	1,-1,-1,-1
Number	**108**	**1053**	**142**	**1462**	**222**	**440**	**242**	**879**
Percent	0.011	0.105	0.014	0.146	0.022	0.044	0.024	0.088
	-1,1,1,1	-1,1,1,-1	-1,1,-1,1	-1,1,-1,-1	-1,-1,1,1	-1,-1,1,-1	-1,-1,-1,1	-1,-1,-1,-1
Number	**203**	**2651**	**47**	**1930**	**71**	**225**	**59**	**275**
Percent	0.020	0.265	0.005	0.193	0.007	0.022	0.006	0.027

Female (10207)								
	1,1,1,1	1,1,1,-1	1,1,-1,1	1,1,-1,-1	1,-1,1,1	1,-1,1,-1	1,-1,-1,1	1,-1,-1,-1
Number	**481**	**1339**	**298**	**2392**	**575**	**433**	**156**	**388**
Percent	0.047	0.131	0.029	0.234	0.056	0.042	0.015	0.038
	-1,1,1,1	-1,1,1,-1	-1,1,-1,1	-1,1,-1,-1	-1,-1,1,1	-1,-1,1,-1	-1,-1,-1,1	-1,-1,-1,-1
Number	**497**	**1050**	**260**	**1171**	**388**	**346**	**183**	**250**
Percent	0.049	0.103	0.025	0.115	0.038	0.034	0.018	0.024

Male + Female (20216)								
	1,1,1,1	1,1,1,-1	1,1,-1,1	1,1,-1,-1	1,-1,1,1	1,-1,1,-1	1,-1,-1,1	1,-1,-1,-1

Number	**589**	**2392**	**440**	**3854**	**797**	**873**	**398**	**1267**
Percent	0.029	0.118	0.022	0.191	0.039	0.043	0.020	0.063
	-1,1,1,1	-1,1,1,-1	-1,1,-1,1	-1,1,-1,-1	-1,-1,1,1	-1,-1,1,-1	-1,-1,-1,1	-1,-1,-1,-1
Number	**700**	**3701**	**307**	**3101**	**459**	**571**	**242**	**525**
Percent	0.035	0.183	0.015	0.153	0.023	0.028	0.012	0.026

The following is its results.

Male and Female			
CV	DT	RF	Adaboost
1	0.143707	0.285431	0.254708
2	0.166501	0.331021	0.315659
3	0.163033	0.335976	0.31665
4	0.175421	**0.339941**	0.326561
5	0.159564	0.320119	0.323092
6	0.170961	0.301288	0.27106
7	0.147175	0.251734	0.214073
8	0.145689	0.228444	0.203171
9	0.143707	0.210109	0.200198
10	0.14668	0.226957	0.216056
Final	0.1562438	**0.283102**	0.2641228

Female

CV	DT	RF	Adaboost
1	0.134446	0.271835	0.251227
2	0.147203	0.279686	0.266928
3	0.154224	0.273084	0.25835
4	0.137525	0.289784	0.272102
5	0.162083	0.273084	0.280943
6	0.145383	**0.29666**	0.27112
7	0.146365	0.297642	0.277014
8	0.166994	0.295678	0.277996
9	0.139489	0.256385	0.24558
10	0.111984	0.225933	0.224951
Final	0.1445696	**0.2759771**	0.2626211

Previous code for training used Resnet model

The training result was worse so didn't be shown in the main text.

Start training

```
NUM_EPOCHS = 3

model = resnet18(num_classes=4)

model = model.to(DEVICE)

#optimizer: Adam
optimizer = torch.optim.Adam(model.parameters(), lr=0.001)
valid_loader = test_loader


def compute_accuracy_and_loss(model, data_loader, device):
    correct_pred, num_examples = 0, 0
    cross_entropy = 0.
    for i, (features, targets) in enumerate(data_loader):

        features = features.to(device)
        targets = targets.to(device)

        logits, probas = model(features)
        cross_entropy += F.cross_entropy(logits, targets).item()
        _, predicted_labels = torch.max(probas, 1)
        num_examples += targets.size(0)
        correct_pred += (predicted_labels == targets).sum()
    return correct_pred.float()/num_examples * 100, cross_entropy/num_examples
```

```
start_time = time.time()
train_acc_lst, valid_acc_lst = [], []
train_loss_lst, valid_loss_lst = [], []

for epoch in range(NUM_EPOCHS):

    model.train()

    for batch_idx, (features, targets) in enumerate(train_loader):

        ### PREPARE MINIBATCH
        features = features.to(DEVICE)
        targets = targets.to(DEVICE)

        ### FORWARD AND BACK PROP
        logits, probas = model(features)
        cost = F.cross_entropy(logits, targets)
        optimizer.zero_grad()

        cost.backward()

        ### UPDATE MODEL PARAMETERS
        optimizer.step()

        ### LOGGING
        if not batch_idx % 500:
            print (f'Epoch: {epoch+1:03d}/{NUM_EPOCHS:03d} | '
                   f'Batch {batch_idx:04d}/{len(train_loader):04d} |'
                   f' Cost: {cost:.4f}')
```

```
    # no need to build the computation graph for backprop when computing accuracy
    model.eval()
    with torch.set_grad_enabled(False):
        train_acc, train_loss = compute_accuracy_and_loss(model, train_loader,
device=DEVICE)
        valid_acc, valid_loss = compute_accuracy_and_loss(model, valid_loader,
device=DEVICE)
        train_acc_lst.append(train_acc)
        valid_acc_lst.append(valid_acc)
        train_loss_lst.append(train_loss)
        valid_loss_lst.append(valid_loss)
        print(f'Epoch: {epoch+1:03d}/{NUM_EPOCHS:03d} Train Acc.:
{train_acc:.2f}%'
              f' | Validation Acc.: {valid_acc:.2f}%')

    elapsed = (time.time() - start_time)/60
    print(f'Time elapsed: {elapsed:.2f} min')

elapsed = (time.time() - start_time)/60
print(f'Total Training Time: {elapsed:.2f} min')
```

Test phase

```
model.eval()
with torch.set_grad_enabled(False): # save memory during inference
    test_acc, test_loss = compute_accuracy_and_loss(model, test_loader, DEVICE)
        print(f'Test accuracy: {test_acc:.2f}%')
```

Prediction

```
_, predictions = model.forward(features[:8].to(DEVICE))
predictions = torch.argmax(predictions, dim=1)
print(predictions)

features = features[:7]
fig = plt.figure()

# print(features[i].size())
for i in range(6):
    plt.subplot(2,3,i+1)
    plt.tight_layout()
    tmp = features[i]
    plt.imshow(np.transpose(tmp, (1, 2, 0)))
    plt.title("Actual value: {}".format(targets[i])+'\n'+"Prediction value:
{}".format(predictions[i]),size = 10)

#     plt.title("Prediction value: {}".format(tname[targets[i]]))
plt.show()
```