CS399 – Mini Project-II

Title - Predicting Big Five Personality Traits from Facial Features using Machine Learning.

Submitted by:

20bcs063 – Jaya Prakash

20bcs064 – Jayant Kumawat

20bcs092 – Neha Porwal

20bcs097 – Aryan Patel

Under the guidance of:

Dr. Chinmayananda A

Assistant Professor (ECE Department)

IIIT Dharwad

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY

Dharwad

ज्ञानेन विकासः

Certificate

This is to certify that the report entitled "**Predicting Big Five Personality Traits from Facial Features using Machine Learning**" is being submitted by **Jaya Prakash(20bcs063), Jayant Kumawat(20bcs064), Neha Porwal(20bcs092), Aryan Patel(20bcs097)** Indian Institute of Information Technology Dharwad, Karnataka, India. The report hasfulfilled all the requirements as per the regulations of the IndianInstitute of Information Technology Dharwad and in my opinion, has reached the standards needed for submission. The work, techniques, and results presented have not been submitted to any other University or Institute for the award of any other degree or diploma.

Acknowledgements

The world is renewing itself in every aspect. Every field is advancing itself in its way of approach. Practical and theoretical knowledge are both equally important and projects do complete the bridge between them. We hereby would thankour Assistant Professor **Dr. Chinmayananda A** sir for enlightening us, by giving us a chance to complete the knowledge bridge day by day.

ABSTRACT

Predicting an individual's personality traits from their facial features is a formidable challenge that hinges on the availability of a substantial dataset. In this study, over 138 volunteers, aged 18 to 25, were enlisted, and their facial features were meticulously analyzed to train a machine learning model. The model was fine-tuned using these facial features as inputs and demonstrated the ability to predict five distinct personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. However, the model's predictive accuracy for these traits was limited to 66.67%.

One noteworthy limitation of this research was the utilization of a relatively small and homogeneous dataset for training the model. To enhance the broader applicability of the findings, it is crucial that future research employs more expansive and diverse datasets. This should involve the inclusion of a wider spectrum of age groups, ethnicities, and cultural backgrounds to capture a more comprehensive array of facial features and personality traits, ultimately contributing to the development of a more robust and dependable model.

Despite its modest accuracy, the algorithm that was developed provides valuable insights into the potential connections between facial features and personality traits. The study underscores the necessity for further investigations to refine and expand the predictive capabilities of such models. Future research endeavors should delve into advanced machine learning techniques, such as deep learning architectures, to harness the potential of neural networks in uncovering subtle relationships between facial features and personality traits.

In conclusion, although the current model's accuracy in predicting personality traits from facial features is limited, it serves as a stepping stone for further exploration in this emerging field. Enhancing the size and diversity of the dataset, alongside the utilization of more sophisticated algorithms, are pivotal steps toward improving the accuracy and reliability of such predictions.

CONTENTS

1) Introduction... 01

2) Objectives 02

3) Literature Study.. 03

4) Methodology 06

   a) Categories of Facial Features Identified

   b) Data Collection

   c) Pre-Processing

   d) Post-Processing

5) Results 17

6) Conclusion 19

7) Challenges and Limitations 20

8) Future Scope 22

9) References.. 23

1) INTRODUCTION

Personality traits are an important factor in determining human behavior as they are relatively stable patterns of thought, feeling, and behavior that influence how individuals interact with others and respond to situations, personality traits have a significant impact on various aspects oflife, such as job performance, social interaction, and mental health. They also impact an individual's ability to work effectively with others, communicate effectively, and manage stress. Understanding personality traits can help identify individuals who may be at the risk of developingmental health problems and can inform interventions and treatments. There are broad number of categories in which we can categorize the personality of a person for our machine learning modelwe have used five broad categories of personality **(Openness, Conscientiousness, Extraversion,Agreeableness and Neuroticism)**. This ML model will not only predict the personality of a personfrom its facial image but it will also give the weightage of which facial part of that person contributes more to the personality. For training purpose, we took the facial picture of participantswithout spectacles and gave a questionnaire to fill in order to find their personality, as a result wefind that there is some percentage contribution of each facial feature in the personality of that person. This model helps us to better understand the behaviour of a person by predicting his personality.

2) OBJECTIVES

The primary objective of this project is to develop a machine learning model capable of predicting an individual's personality traits based on their facial image. To accomplish this objective, several specific goals were pursued.

The initial goal involved the collection of a comprehensive dataset of facial features. A diverse group of volunteers, encompassing individuals aged 18 to 25, was recruited for the study. Facial feature data, including measurements such as facial landmarks, symmetry, and proportions, were meticulously extracted from images or videos of the participants. It was imperative to ensure that the dataset was representative of various ethnicities, genders, and cultural backgrounds. This diversity was vital to account for potential variations in facial features across different populations and enhance the model's ability to generalize.

The subsequent goal was to train a machine learning model using the gathered facial feature data. The model was specifically designed to predict five key personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. Several machine learning algorithms, including decision trees, support vector machines, and neural networks, were explored to determine the most suitable approach for this task. The model was trained utilizing the facial feature data as inputs and the corresponding personality trait labels obtained through self-report questionnaires or other established assessment methods.

The third objective involved the assessment of the accuracy of the developed model in predicting personality traits. The trained model was put to the test using a separate set of data that was not part of the training phase. This evaluation aimed to gauge the model's performance in predicting each personality trait and to identify its strengths and limitations.

The final goal was to pinpoint opportunities for enhancing the model's predictive accuracy. Drawing insights from the evaluation results, the study aimed to identify factors that contributed to the model's performance. This process sought to uncover areas where the dataset or the machine learning algorithm could be enhanced to improve the accuracy of personality trait predictions.

3) LITERATURE STUDY

1) Analysis of personality traits' correlation to facial width-to-height ratio (fWHR) and mandibular line angle based on 16 personality factor in Chinese college students

Abstract

Researchers studied how facial features relate to personality traits. They examined the width-to-height ratio of faces (fWHR) and jaw shape. They found that these features had links to personality, with jaw shape being more important. In males, a wider fWHR was linked to lower sensitivity and self-reliance. In females, jaw shape was related to lower vigilance and higher anxiety. These links varied between genders. This suggests that facial features can tell us about personality, but it may depend on whether you're male or female. In the future, studying jaw shape could be a useful way to understand personality. The study had some limitations, like only including young adults and not using precise jaw measurements, which could be improved in future research.

Datasets

The study involved 904 participants, including 226 males and 678 females. They were all students at Beijing University of Chinese Medicine and were recruited between November 21, 2020, and January 9, 2021. The study was conducted in the Chinese language.

Inclusion criteria: Participants had to be students at Beijing University of Chinese Medicine (undergraduates, master's, or Ph.D. candidates) of Han nationality, aged 18 to 35. They should have had no noticeable facial changes, voluntarily joined the study, and signed informed consent forms. They also should have been free from any significant physical or mental illnesses or discomfort that might hinder their questionnaire completion, and they should not have taken a similar questionnaire in the past month.

Facial Feature

fWHR (Facial Width-to-Height Ratio): This is the ratio of the width of the face (distance between the right and left zygions - b1-b2) to the height of the face (distance between the lip peak - a2, and the brows - a1).

Bilateral Mandibular Line Angle: To measure this, 50 points on each side of the jaw were used to fit a line, and the mandibular line angle was the angle between the fitted line and horizontal lines.

This angle indicates the squareness or narrowness of the jaw, where a larger angle suggests a narrower jaw. The measurements were taken as shown in Figure 2. FaceDiag software developed by the Institute of Microelectronics of the Chinese Academy of Sciences was employed to process images, automatically identify location points, and measure these facial features.

2) **2.5D Facial Personality Prediction Based on Deep Learning** by Jia Xu, Weijian Tian, Guoyun Lv, Shiya Liu, and Yangyu Fan (2021)

Introduction:

Prominent facial areas, such as prominent landmarks of the forehead, nose, and chin, are related to a person's personality. Therefore, multiple perspectives (front, side, and 2.5D) facial images are more likely to describe a person's personality comprehensively and accurately. Herein, we use the term "2.5D" to refer to combinations of front and side views. The two main topics in existing face personality prediction research are the acquisition of datasets (face photos and personality data) and the design of computing networks.

Phases:

1. Construction of Face Database.
2. Selection of the Personality Evaluation Model.
3. Selection of the Prediction Network.

Dataset and Pre-processing:

- Samples and Procedure: The official language used in this study is Chinese. The data were based on a sample of 5,560 male and 8,547 female college students aged 18 to 25.
- Ethical Approval
- Establishment of the Personality Dataset (Big Five Personality Traits): The Big Five are openness, conscientiousness, extraversion, agreeableness, and neuroticism. A score of 0 to 60 is set for each dimension, such as agreeableness, where the higher individual's score, the more easy-going and pleasant the personality is [21, 22].
- Screening and Analysis of Image and Personality Data.

Neural Network for Personality Prediction Based on 2D Images.

Consequently, we employed a deep learning method to extract high-level features from face images for personality prediction. We used MobileNetV2 and residual network version 50 (ResNet50), two deep learning networks that are popular in academia, to classify personality traits.

To verify the experimental results, 5-fold cross-validation method was used. The data were randomly scrambled and divided into five pieces, and for each fold, one piece of data was further divided into equally sized test and validation sets, and the remaining four pieces as the training set. Take the average of the verification results from the five folds as the final result.

Results:

In this study, the data were scrambled and randomly divided into five sections, one of the sections was further divided into equally sized test and validation sets, and the remaining four sections served as the training set. Final prediction result is the average of the verification results from using each of the five parts as the verification set.

a) Firstly, all volunteers were Asian, who, due to cultural differences, emphasize their Asians place more emphasis on self-discipline and commitment, preciseness and meticulousness, resourcefulness and determination, and tenacity and steadiness.
b) Secondly, all volunteers were college students, who tend to have relatively little contact with society and do not take much responsibility. Therefore, their understanding of self-consciousness and agreeableness may not be comprehensive, affecting the corresponding score on the self-esteem scale and further affecting the prediction performance of these two dimensions.
c) Third, research is based on facial images, in which there are obvious differences between the features of Chinese and Western people. For example, Westerners have obvious facial contours with high noses, while Asians have relatively flat facial contours and soft lines. Therefore, the prediction results of Chinese and foreigners, especially Westerners, personalities based on facial features are bound to be different.

4) METHODOLOGY

a) Categories of Facial Features Identified

The ML model used in our project aims to predict a person's personality based on their facial features. We specifically focused on five facial features: eyebrows, forehead, face shape, cheeks, and nose. These features were selected based on their potential correlation with personality traits.

To analyze the facial features effectively, it is important to categorize them into various classifications. By categorizing these features, we can better understand their potential influence on personality prediction. Predicting the personality of a person from his picture using a Machine Learning model requires well known information about the facial features of that person, for a better understanding of facial features we have divided facial features into following categories –

I. **Eyebrows:** The role of eyebrows in personality prediction is an interesting aspect offacial feature analysis. Eyebrows have been recognized as a significant feature that canprovide insights into a person's personality traits. According to the article by BYRDIEwritten by Lindsey Metros (General Manager at BYRDIE) and reviewed by a group ofmakeup artists, eyebrows are categorized into twelve categories.

Round Eyebrows

Straight Eyebrows

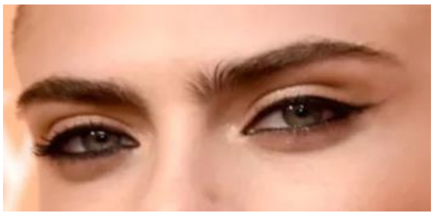
Peaked Eyebrows

Queen's Eyebrows

Long Eyebrows

Short Eyebrows

Thick Eyebrows



Thin Eyebrows

Diagonal Eyebrows

Tampered Eyebrows

“S”-Shaped Eyebrows

Hard-Arched Eyebrow

II. **Eye Shape:** Eye shape is an intriguing facial feature that has been explored in personality prediction studies. How to determine which type of eye shape you have? Everyone has a unique set of eyes, so based on different studies, we can classify the eyes into ten broad categories.

Almond Eyes:

Round Eyes:

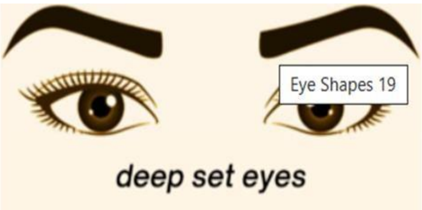
Monolid Eyes:

Protruding Eyes:

Downturned Eyes: **Upturned Eyes:**

Close- Set- Eyes: **Wideset Eyes:**

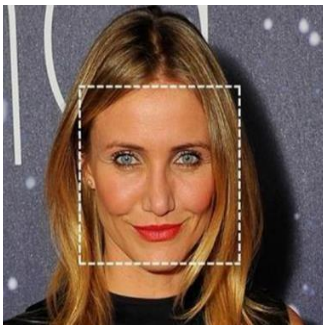
Hooded Eyes: **Deep-Set Eyes:**



III. **Face Shape:** Face shape is considered to have a role in predicting personality traits based on facial feature analysis. Different face shapes are believed to be associated with certain personality characteristics. Face can be categorized into nine broadcategories on the basis of its shape.

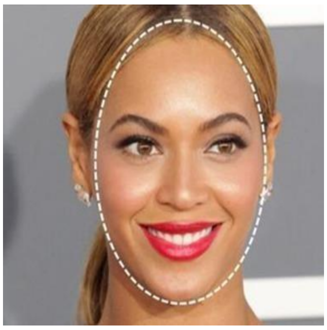
Round Face Shape:

Square Shape Face:



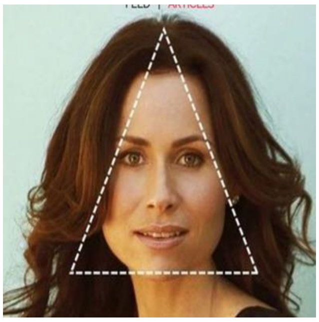
Heart Face Shape:

Oval Face Shape:

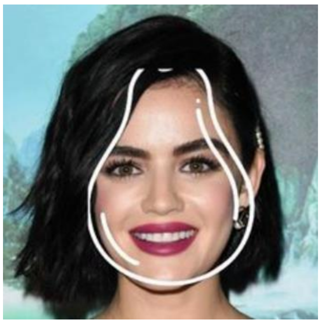


Rectangle Face Shape:

Triangle shape Face:

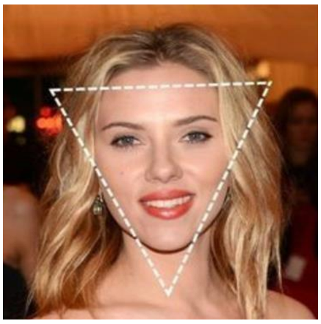


Pear Face Shape:



Diamond Shape:

Inverted Triangle Face Shape:



IV. **Nose Shape:** Nose shape is considered to have a role in predicting personality traitsbased on facial feature analysis. Different nose shapes are believed to be associated with certain personality characteristics. Nose can be categorized into below categorieson the basis of its shape.

Greek nose:

Straight Nose:

Crooked Nose:

Button Nose:

Concave Nose:

Nubian Nose:

Hook Nose:

Arched Nose:

b) Data Collection

The data collection process for our project involved gathering primary data from 138 volunteers who were college students aged between 18 and 25 years. Out of these volunteers, 7 were female, and the remaining were male. To collect the data, we utilized Google Forms, which included a total of 38 questions related to the five different personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism).

In addition to the questionnaire, the Google Form also included an image section where the volunteers were asked to upload a front-facing image of their face without wearing spectacles. Furthermore, the volunteers were required to rate each question on a scale of 0 to 5. This approach allowed us to collect a significant amount of data regarding both facial features andpersonality traits of the volunteers.

All of the collected data was securely stored on a Google Drive, with restricted access limitedto the team members involved in the project. This ensured the privacy and confidentiality of the volunteers' information, preventing unauthorized access to the data. However, it is important to note that since the Google Form was filled out without the presence of an invigilator, there is a possibility of potential biases in the data collection process. These biasesmay have influenced the quality and accuracy of the collected data, which could ultimately impact the reliability of our project

C) Pre-Processing:

In the pre-processing stage of the project, a significant amount of manual work is involved. After collecting the responses to the questionnaire, the data is categorized into five different personality trait categories. This categorization process determines which questions belong toeach category, such as Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

To calculate the overall score for each person, a formula is applied. The formula takes into account the responses provided for each question within a category and assigns a numerical value to represent the individual's score for that particular personality trait category. The specific details of the formula and scoring methodology are as shown below:

A	B
Personality Type	Formula to get personality score
Openness (F3)	(Q3+Q10+Q14+Q16+Q22+Q24+Q38)/7
Extraversion (F1)	(Q5+Q7+Q12+Q18+Q23+Q26+(6-Q31)+Q35)/8
Agreeableness (F4)	(Q2+Q9+(6-Q11)+(6-Q17)+Q21+Q29+Q32+Q34+Q36)/9
Neuroticism (F2)	(Q4+Q6+Q13+Q20+Q25+(6-Q28)+Q30+Q37)/8
Conscientiousness (F5)	(Q1+Q8+Q15+Q19+(6-Q27)+Q33)/6
	Qi is the score that you get for ith question from the questionnaire

.

Once the scores are calculated, each person is assigned a personality type based on their scorein each category.

All the gathered information, including the assigned personality types and the values assignedto facial features, is compiled and uploaded to a CSV file, likely for further analysis and modeltraining. Overall, the pre-processing stage involves manual efforts to categorize the questionnaire responses, calculate scores and assign personality types. This process lays the foundation for subsequent analysis and model development.

c) Post Processing

In the post-processing stage, after calculating the scores for each person and categorizing their personality types, the focus shifts to developing the machine learning (ML) model. Several Python libraries are utilized for this purpose, particularly we used haarcascade function of CV2 library which helped the model to mark some important parts of an image like face shape, eye shape, nose shape and lips shape after getting the marking of these special parts of image we used an inbuilt function crop to crop this special part from the image and that we put these crop part from image to train our model. Additionally, various ML algorithms are employed, including Random Forest, Gradient Boosting, Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine.

First step in model development involves importing the CSV file containing information about the facial features and corresponding personality types.

Next, the dataset is split into a 1:4 ratio, with one portion reserved for testing purposes andthe remaining four portions used for training the model. This approach ensures that enoughdata is available for training while still allowing for evaluation on unseen data.

of kernel functions.

Each of these algorithms offers its own strengths and characteristics, and their selection depends on the specific requirements and nature of the data being used in the ML model. Once the model is trained, it is ready for testing. Approximately 20% of the overall data set is used for this purpose. The testing phase evaluates the model's performance on unseen data to assess its accuracy and generalization capabilities.

Overall, the post-processing stage involves importing the data, splitting it into training and testing sets, training the ML model using Random Forest, and evaluating the model's performance on the testing data. This stage is crucial for validating the model and assessing its effectiveness in predicting personality types based on facial features.

Random Forest algorithm is applied to train the ML model. Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It is known for its robustness and ability to handle complex relationships between features and labels.

Gradient Boosting: Gradient Boosting is an ensemble learning method that combines multiple weak predictive models, typically decision trees, to create a strong predictive model. It works by sequentially adding models to correct the mistakes made by previous models, gradually improving the overall predictions.

Decision Tree: Decision Tree is a popular algorithm that uses a tree-like structure to make decisions based on features. It partitions the data into smaller subsets based on the values of different features, creating a tree of decision rules. Each internal node represents a feature, each branch represents a decision rule, and each leaf node represents an outcome or prediction.

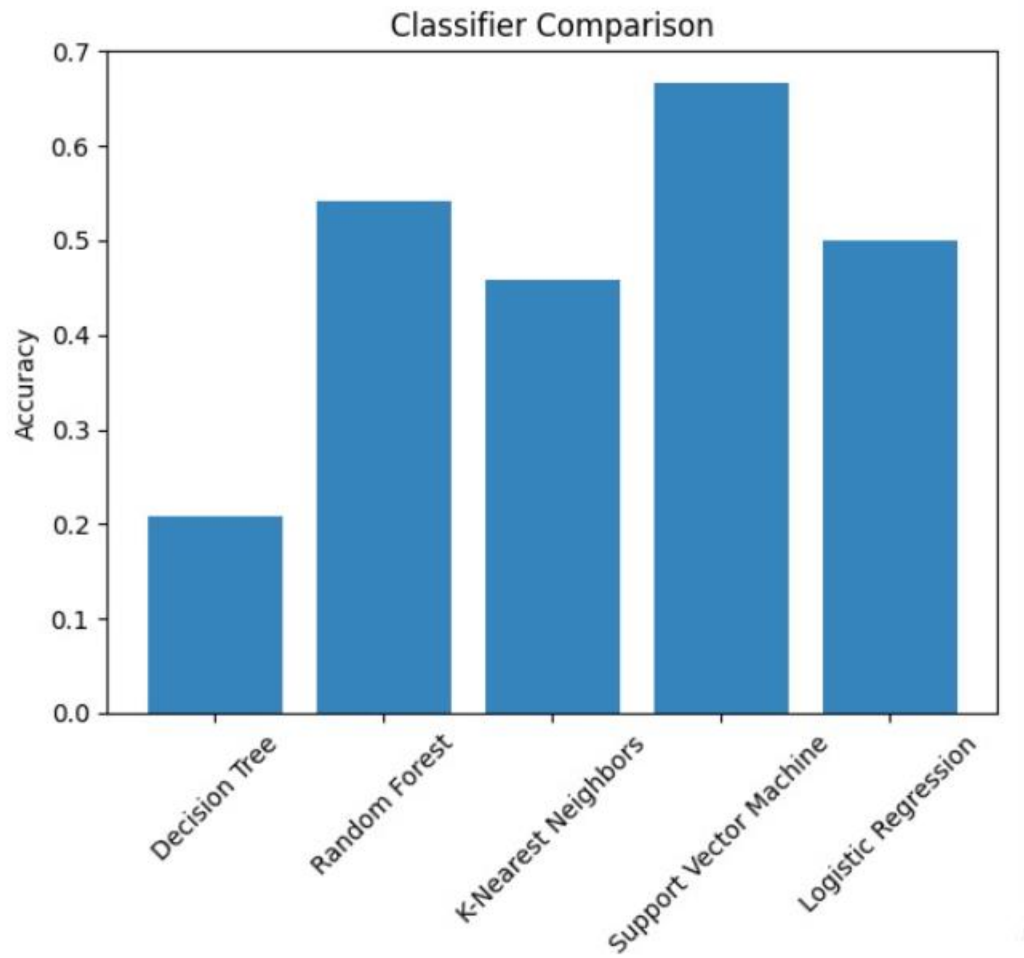
K-Nearest Neighbors (KNN): K-Nearest Neighbors is a simple yet effective algorithm for classification tasks. It classifies data points by finding the k nearest neighbors in the training set and assigning the most common class label among those neighbors to the test data point. KNN is based on the assumption that similar instances are likely to belong to the same class.

Support Vector Machine (SVM): Support Vector Machine is a powerful algorithm used for both classification and regression tasks. It aims to find the optimal hyperplane that separates the data points of different classes with the maximum margin. SVM can handle high-dimensional data and is effective in handling non-linear relationships through the use

5) RESULTS

In our study, we explored the performance of several machine learning models, including Support Vector Machines (SVM), Decision Tree, Random Forest, Logistic Regression, and k-Nearest Neighbors (KNN), for personality prediction. Each model yielded different accuracy levels in predicting personality traits.

The classification reports for each model provide comprehensive insights into their performance. They include metrics such as precision, recall, and F1-score for each personality trait, offering an understanding of the strengths and weaknesses of each model in predicting specific traits.



	Model	Accuracy
0	Logistic Regression	0.500000
1	Random Forest	0.541667
2	Support Vector Machine	0.666667
3	K-Nearest Neighbors	0.458333
4	Decision Tree	0.208333
5	GRADIENT BOOSTING	0.208333

6) CONCLUSION

The accuracy of the model developed in this study is around 66.67%, which is not very impressive. This can be due to several reasons, including the relatively small dataset used in this study, as well as the complexity of the task of predicting personality traits from facial features. It is worth noting that predicting personality traits from facial features is a challenging task, as facial features alone may not provide a complete picture of a person's personality.

To improve the accuracy of the model, future research can focus on increasing the dataset size and diversity. This can involve recruiting a larger number of participants with a wider range of ages, ethnicities, and cultural backgrounds. In addition, an expert in facial feature categorization can be consulted to ensure accurate identification and classification of each facial feature in the dataset.

In conclusion, while the model's accuracy in predicting personality traits from facial features may be limited, the findings of this study provide a valuable starting point for further research in this field. With larger and more diverse datasets, as well as the assistance of experts in facial feature categorization, it is possible to develop more accurate models for predicting personality traits from facial features.

7) CHALLENGES AN LIMITATIONS

The challenges in our project can be elaborated as follows:

Limited Dataset Size: The accuracy rate of 50 percent can be attributed to the relatively small size of the dataset. With only 157 volunteers, the dataset may not provide enough diversity and variability in facial features and personality traits. A larger dataset would allow for better generalization and more accurate predictions.

Lack of Expertise in Facial Feature Extraction: The precision of feature extraction may be affected by the lack of expertise in accurately identifying and extracting facial features. Facial feature analysis requires a deep understanding of the nuances and variations in different features. Without sufficient expertise, there is a higher chance of misclassification or inaccurate extraction of features, which can impact the overall accuracy of the model.

Absence of a Personality Expert: The absence of a personality expert during the questionnaire design phase can impact the accuracy of the questionnaire itself. Designing effective and reliable personality assessment questions requires expertise in understanding the intricacies of different personality traits and ensuring the questions capture the relevant aspects accurately. Without a personality expert's guidance, the questionnaire may lack the necessary precision and validity, leading to lower accuracy in personality predictions.

Lack of Invigilators during Questionnaire Filling: The questionnaire not being filled in the presence of an invigilator introduces a potential source of error. Without proper supervision, there is a risk of inconsistent or inaccurate responses from the volunteers. Some individuals may provide hasty or careless answers, while others deliberately manipulate their responses. These errors can introduce noise and bias into the dataset, compromising the reliability and accuracy of the model.

Missing Facial Pictures and Reduced Sample Size: The absence of facial pictures from some volunteers and the resulting reduced sample size can introduce bias and affect the representativeness of the dataset. If certain demographics or individuals with specific facial features are disproportionately missing, it can lead to an unbalanced dataset and impact the model's ability to predict personality traits based on facial features accurately.

Addressing these challenges would involve expanding the dataset, acquiring expertise in facial feature extraction, involving a personality expert in the questionnaire design, implementing measures to ensure accurate responses during data collection, and minimizing missing data to improve the reliability and accuracy of the model.

8) FUTURE SCOPE

The future scope of the project involves addressing the challenges identified and implementing improvements to enhance the accuracy and reliability of the model.

Firstly, increasing the dataset size is crucial to improve the accuracy rate. By collecting data from a larger and more diverse group of volunteers, we can ensure better representation of facial features and personality traits. This will help mitigate the limitations of a small dataset and improve the generalizability of the model's predictions.

Secondly, it is important to acquire expertise in facial feature extraction. By involving professionals with knowledge and experience in accurately identifying and extracting facial features, we can enhance the precision of feature extraction. This will minimize errors and inconsistencies, improving the overall reliability of the data.

Furthermore, incorporating additional facial features and expanding the categories within each feature can lead to more comprehensive personality predictions. Including features such as eyes, lips, chin, and jawline, and introducing more specific categories for each feature, will capture a wider range of facial characteristics. This will provide a more nuanced understanding of the relationship between facial features and personality traits.

To improve the reliability of the data, measures should be taken to ensure accurate responses during the questionnaire filling process. This could involve introducing an invigilator to oversee the questionnaire filling, which would minimize errors and deliberate manipulation. Additionally, having a personality expert involved in the questionnaire design phase will enhance the accuracy and validity of the questions, improving the overall quality of the data collected.

In summary, the future direction of the project involves increasing the dataset size, acquiring expertise in facial feature extraction, incorporating additional facial features and more categories within each feature, and implementing measures to improve the reliability of the data. These improvements will contribute to a more accurate and reliable machine learning model for predicting personality based on facial features.

9) REFERENCES

The following references were consulted and referred to in our project:

- 12 Types of Eyebrow Shapes and What They Reveal About You (byrdie.com)
- Analysis of personality traits' correlation to facial width-to-height ratio (fWHR) and mandibular line angle based on 16 personality factor in Chinese college students | PLOS ONE
- Understanding Different Eye Shapes: Which Do You Have? | NVISION Eye Centers (nvisioncenters.com) Understanding Different Eye Shapes: Which Do You Have? | NVISION Eye Centers (nvisioncenters.com)
- Types of Eyelids: 9 Different Shapes of Eyes and Their Names (stylesatlife.com)
- What's My Eye Shape? (Learn How to Tell Here) (visioncenter.org)
- Understanding Different Eye Shapes: Which Do You Have? | NVISION Eye Centers (nvisioncenters.com) Understanding Different Eye Shapes: Which Do You Have? | NVISION Eye Centers (nvisioncenters.com)
- Types of Eyelids: 9 Different Shapes of Eyes and Their Names (stylesatlife.com)
- What's My Eye Shape? (Learn How to Tell Here) (visioncenter.org)
- .https://winkgo.com/shape-nose-reveals-personality/
- https://www.cosmopolitan.in/life/news/a11432/10-types-noses-and-what-secrets-they-reveal-about-your-personality