CS399 – Mini Project-1

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 has fulfilled all the requirements as per the regulations of the Indian Institute of Information Technology Dharwad and in my opinion, has reached the standards needed for submission. The work, techniques, and results presented have not beensubmitted 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 thank our Assistant Professor **Dr. Chinmayananda A** sir for enlightening us, by giving us a chance to complete the knowledge bridge day by day.

ABSTRACT

Predicting a person's personality traits from their facial features is a challenging task that requires a large amount of data. To train the machine learning model in this study, more than 138 volunteers between the ages of 18 and 25 were recruited, and their facial features were extracted for analysis. The model was trained using these features as inputs and can predict five different personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. However, the model's accuracy in predicting these traits was only 46 percent.

One limitation of this study was the relatively small and homogeneous dataset used for training the model. In future research, larger and more diverse datasets should be employed to enhance the generalizability of the findings. Incorporating a more extensive range of age groups, ethnicities, and cultural backgrounds would help capture a broader spectrum of facial features and personality traits, leading to a more robust and reliable model.

Despite its limited accuracy, the developed algorithm offers valuable insights into the potential connection between facial features and personality traits. The study highlights the need for additional investigations to refine and expand the predictive capabilities of such models. Future research efforts should explore advanced machine learning techniques, such as deep learning architectures, to leverage the power of neural networks in uncovering subtle relationships between facial features and personality traits.

In conclusion, while the current model's accuracy in predicting personality traits from facial features is modest, it lays the foundation for further exploration in this emerging field. Enhancing the dataset size and diversity, and utilizing more sophisticated algorithms, will be essential in improving the accuracy and reliability of such predictions.

CONTENTS

1) INTRODUCTION

Personality traits are an important factor in determining human behaviour as they are relatively stable patterns of thought, feeling, and behaviour that influence how individuals interact with others and respond to situations, personality traits have a significant impact on various aspects of life, 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 developing mental 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 model we have used five broad categories of personality **(Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism)**. This ML model will not only predict the personality of a person from 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 participants without spectacles and gave a questionnaire to fill in order to find their personality, as a result we find 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 objective of this project is to develop a machine learning model capable of predicting personality traits based on facial features. To achieve this objective, several specific goals were pursued.

The first goal was to collect a dataset of facial features. A sample of volunteers, consisting of individuals between the ages of 18 and 25, was recruited for the study. Facial feature data, including measurements such as facial landmarks, symmetry, and proportions, were extracted from images or videos of the participants. It was crucial to ensure a diverse dataset that encompassed individuals from different ethnicities, genders, and cultural backgrounds. This diversity would help account for potential variations in facial features across different populations and improve the generalizability of the model.

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

The third goal was to evaluate the accuracy of the developed model in predicting personality traits. The trained model was tested on a separate set of data, not used during the training phase, to assess its predictive performance. The accuracy of the model's predictions for each personality trait was measured and analyzed. The evaluation aimed to determine the strengths and limitations of the model and identify potential areas for improvement.

The fourth goal was to identify opportunities for enhancing the model's accuracy. Based on the evaluation results, insights were gathered to identify factors contributing to the model's performance. The study aimed to identify areas where the dataset or the machine learning algorithm could be improved to enhance the accuracy of personality trait predictions.

3) LITERATURE STUDY

1) **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.

2) Assessing the Big Five personality traits using real-life static facial images:

Method: The authors proposed a deep neural network model that takes static facial images as input and generates predictions for the Big Five personality traits (openness, conscientiousness, extraversion, agreeableness, and neuroticism). They used a dataset of 5,000 facial images from the internet, which were labelled with personality trait scores by human rates. The neural network model used a combination of convolutional and fully connected layers to learn features from the facial images and predict the personality trait scores.

Important Contribution: The paper's key contribution is the development of a deep learning-based model that can predict personality traits from facial images. The authors demonstrated that their model achieved state-of-the-art performance on this task and outperformed several baseline models. This work opens up new possibilities for using facial images as a non-intrusive and cost-effective way to assess personality traits.

Dataset Detail: The dataset used in the study contains 5,000 facial images of individuals, including both males and females, with different ethnicities, ages, and expressions. The images were collected from the internet, and each image was labeled with scores for the Big Five personality traits by human raters using the Ten Item Personality Inventory (TIPI) questionnaire.

Conclusions: The authors found that their deep learning model achieved high accuracy in predicting personality traits from facial images. They also found that certain facial regions, such as the eyes and mouth, were more informative for predicting certain personality traits. The authors suggested that their model could be used in various applications, such as improving human-robot interactions and personalizing marketing strategies.

Limitations: One limitation of the study is that the dataset used in the study may not be representative of the general population, as it was collected from the internet and labeled by a specific group of raters. Additionally, the study only considered the Big Five personality traits and did not evaluate the model's performance on other personality trait models. The authors noted that the model's performance may be limited by the quality and diversity of the facial images used as input.

3) **Facial Emotion Recognition Using Conventional Machine Learning and Deep Learning Methods: Current Achievements, Analysis and Remaining Challenges by Amjad Rehman Kha**n

Introduction: The paper presents an overview of facial emotion recognition (FER) using conventional machine learning and deep learning methods, highlighting the importance of FER in various fields such as psychology, medicine, security, and human-computer interaction. The paper provides a brief history of FER and describes the challenges in accurately recognizing emotions from facial expressions.

Methods: The paper describes the methodology used to conduct the review, which includes a comprehensive literature search of relevant databases and a detailed analysis of recent studies in the field of FER. The paper also presents an overview of conventional machine learning methods such as support vector machines (SVMs) and decision trees, and deep learning methods such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), and their application in FER.

Important contribution: The paper provides an in-depth analysis of the achievements and limitations of conventional machine learning and deep learning methods in FER. The paper highlights the advantages of deep learning methods over conventional machine learning methods, such as their ability to learn high-level features from raw data, which leads to better performance in FER tasks.

Dataset details: The paper describes several popular datasets used in FER research, including the CK+, Oulu-CASIA, Affect Net, and FER2013 datasets. The paper provides a brief overview of these datasets, including the number of images, emotions, and subjects, and their use in evaluating FER algorithms.

Conclusions: The paper concludes that deep learning methods have shown superior performance over conventional machine learning methods in FER tasks, but there are still many challenges that need to be addressed, such as the need for more diverse and challenging datasets, the need for interpretability and explain ability of FER models, and the need to address ethical and privacy concerns. The paper highlights the potential future directions of research in FER, such as the integration of FER with other modalities, and the development of more robust and efficient FER models.

Limitations: The paper acknowledges some limitations of the study, including the possibility of missing relevant studies in the literature search, the lack of a standardized methodology for evaluating FER algorithms, and the limitations of the datasets used in FER research, such as their small size and limited diversity.

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 of facial feature analysis. Eyebrows have been recognized as a significant feature that can provide insights into a person's personality traits. According to the article by BYRDIE written by Lindsey Metros (General Manager at BYRDIE) and reviewed by a group of makeup artists, eyebrows are categorized into twelve categories.

Round Eyebrows

Straight Eyebrows

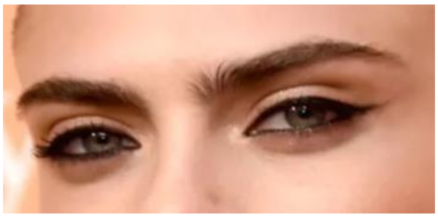
Peaked Eyebrows

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:

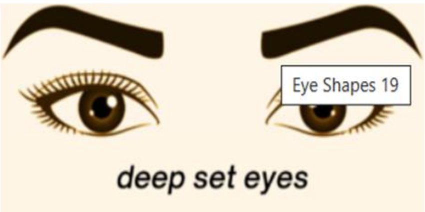
Upturned Eyes:

Close- Set- Eyes:

Wideset Eyes:

Hooded Eyes:

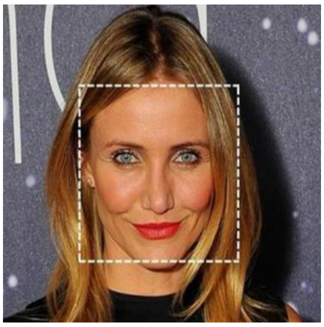
Deep-Set Eyes:



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 broad categories on the basis of its shape.

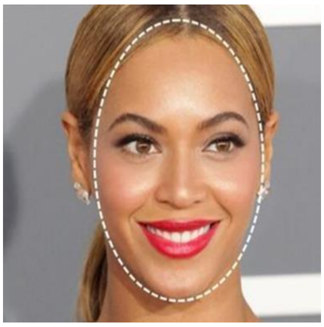
Round Face Shape:

Square Shape Face:



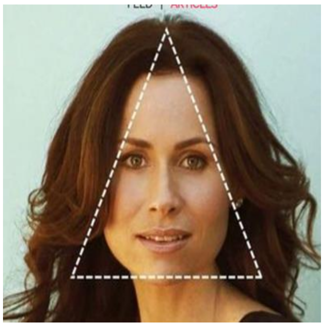
Heart Face Shape:

Oval Face Shape:

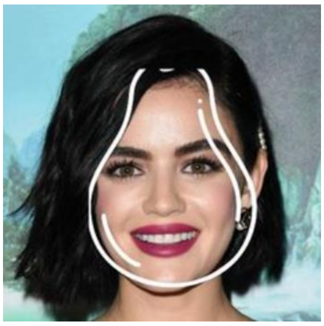


Rectangle Face Shape:

Triangle shape Face:

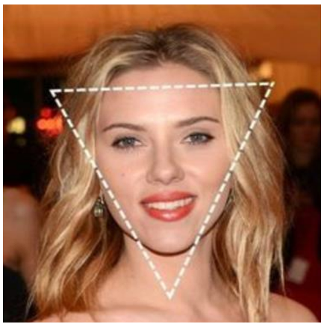


Pear Face Shape:



Diamond Shape:

Inverted Triangle Face Shape:



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

All of the collected data was securely stored on a Google Drive, with restricted access limited to 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 biases may 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 to each 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 score in each category.

To facilitate this process, the facial features are also manually assessed and categorized. The features considered are eyebrows, forehead, face shape, cheeks, and nose. Each feature is divided into distinct categories based on their characteristics, such as long or short eyebrows, broad or tiny forehead, and long or wide face shape, among others. For each feature, a value of 1 is assigned if the person possesses a particular characteristic associated with a category, and 0 is assigned if they do not possess that characteristic.

Features	Categories
Eyebrows	Long, Short, Thick, Thin
Forehead	Broad, Tiny
Face	Long, Wide
Cheeks	Fleshy, Normal
Nose	Long , Short, Sharp, and Blunt

All the gathered information, including the assigned personality types and the values assigned to facial features, is compiled and uploaded to a CSV file, likely for further analysis and model training.

Overall, the pre-processing stage involves manual efforts to categorize the questionnaire responses, calculate scores, assign personality types, and categorize facial features. 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, with pandas and scikit-learn being particularly important. Additionally, various ML algorithms are employed, including Random Forest, Gradient Boosting, Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM).

The first step in model development involves importing the CSV file containing information about the facial features and corresponding personality types. The data is then prepared for training the model. This typically involves creating two NumPy matrices: 'X' for features and 'y' for labels (personality types).

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

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 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.

5) RESULTS

In our study, we explored the performance of several machine learning models, including Support Vector Machines (SVM), Decision Tree, Random Forest, Gradient Boosting, 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.

1) Random Forest Model

	precision	recall	f1-score	support
Openness	0.55	0.63	0.59	19
Conscientiousness	0.42	0.50	0.45	10
Extraversion	0.47	0.47	0.47	17
Agreeableness	0.00	0.00	0.00	4
Neuroticism	0.00	0.00	0.00	3
accuracy			0.47	53
macro avg	0.29	0.32	0.30	53
weighted avg	0.43	0.47	0.45	53

2) Kth Nearest Neighbor Model

```
print(classification_report(y_test, knn_y_pred, target_names=target_names))
```

	precision	recall	f1-score	support
Openness	0.41	0.63	0.50	19
Conscientiousness	0.38	0.50	0.43	10
Extraversion	0.36	0.24	0.29	17
Agreeableness	0.00	0.00	0.00	4
Neuroticism	0.00	0.00	0.00	3
accuracy			0.40	53
macro avg	0.23	0.27	0.24	53
weighted avg	0.34	0.40	0.35	53

3) Support Vector Machine Model

```
print(classification_report(y_test, svm_y_pred, target_names=target_names))
```

	precision	recall	f1-score	support
Openness	0.36	0.74	0.48	19
Conscientiousness	0.22	0.20	0.21	10
Extraversion	0.40	0.12	0.18	17
Agreeableness	0.00	0.00	0.00	4
Neuroticism	0.00	0.00	0.00	3
accuracy			0.34	53
macro avg	0.20	0.21	0.18	53
weighted avg	0.30	0.34	0.27	53

4) Gradient Boosting Model

```
print(classification_report(y_test, gradient_y_pred, target_names=target_names))
```

	precision	recall	f1-score	support
Openness	0.52	0.58	0.55	19
Conscientiousness	0.18	0.20	0.19	10
Extraversion	0.50	0.41	0.45	17
Agreeableness	0.33	0.25	0.29	4
Neuroticism	0.25	0.33	0.29	3
accuracy			0.42	53
macro avg	0.36	0.35	0.35	53
weighted avg	0.42	0.42	0.42	53

5) Decision Tree Model

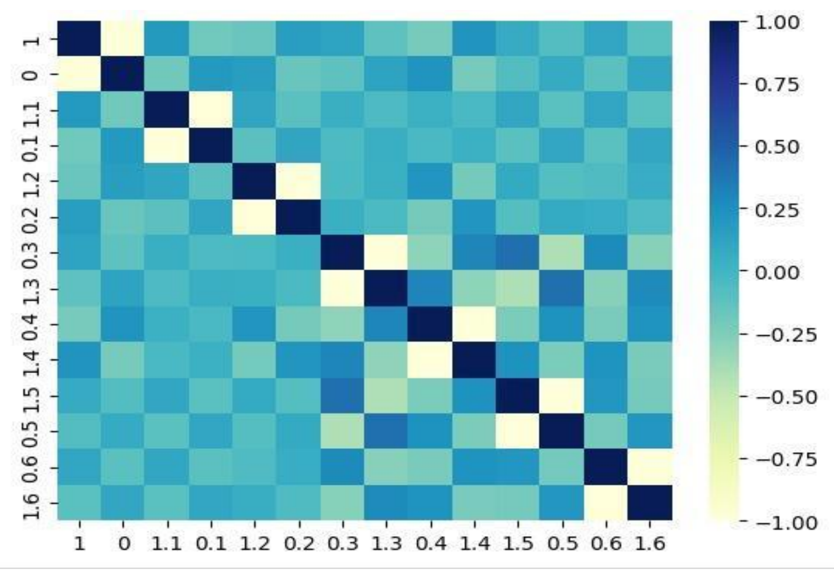
```
print(classification_report(y_test, decision_y_pred, target_names=target_names))
```

	precision	recall	f1-score	support
Openness	0.46	0.58	0.51	19
Conscientiousness	0.29	0.20	0.24	10
Extraversion	0.53	0.53	0.53	17
Agreeableness	0.00	0.00	0.00	4
Neuroticism	0.00	0.00	0.00	3
accuracy			0.42	53
macro avg	0.25	0.26	0.26	53
weighted avg	0.39	0.42	0.40	53

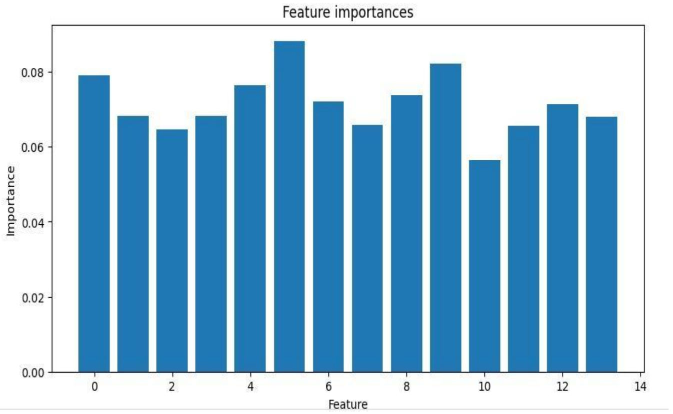
In the below table, each facial feature's contribution to predicting different personality traits is presented visually. The graph allows for the identification of the relative importance of each facial feature in determining an individual's personality. By analyzing the graph, it is possible to determine the facial features with the strongest impact on predicting certain personality traits, as well as those with a weaker impact.

	long(Eyebrows)	thick(Eyebrows)	broad(Forehead)	Long(Face_shape)	fleshy(Cheek)	Long(Nose)	Sharp(Nose)
openness	0.543478	0.565217	0.630435	0.500000	0.543478	0.608696	0.391304
onscientiousness	0.461538	0.653846	0.615385	0.538462	0.461538	0.730769	0.230769
extraversion	0.658537	0.512195	0.634146	0.487805	0.585366	0.609756	0.536585
agreeableness	0.363636	0.454545	0.545455	0.363636	0.545455	0.545455	0.363636
neuroticism	0.375000	0.875000	0.875000	0.500000	0.500000	0.375000	0.375000

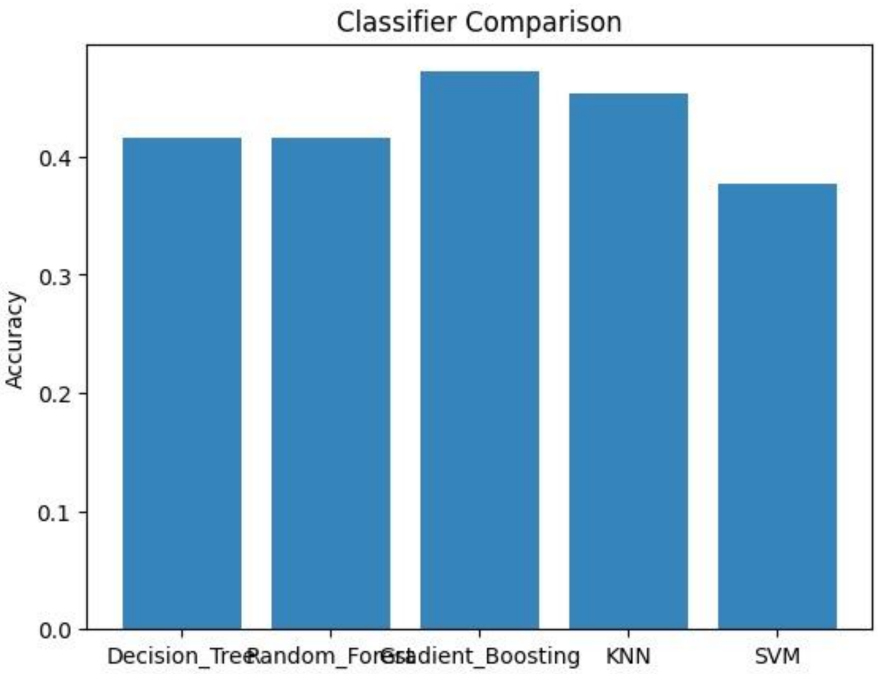
To analyze the relationships between facial features, we employed a heatmap visualization. This heatmap helped us uncover correlations and dependencies among the features, aiding in feature selection and understanding their significance in predicting personality traits. It provides a visual representation of how each personality trait is related to every other personality trait. By analyzing the graph, one can determine the strength of the relationships between different personality traits. The heat map graph is useful in identifying which personality traits are positively or negatively correlated and helps in gaining insights into the complex interrelationships between personality traits.



Furthermore, we developed a graph illustrating the percentage contribution of each feature towards predicting the different personality traits. This graph provides valuable insights into the relative importance of each feature type, shedding light on the facial cues that have the most significant impact on predicting specific personality traits. It displays the weightage assigned to each feature in determining an individual's personality. By examining the graph, it is possible to identify which features have the most significant impact on predicting specific personality traits and which features have a less significant impact.



To evaluate and compare the effectiveness of the models, we created a classifier comparison graph. This graph visually represents the accuracy scores achieved by each model, enabling a clear comparison of their performance and highlighting their respective strengths and weaknesses in personality prediction.



6) CONCLUSION

The accuracy of the model developed in this study is around 43%, 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.

The model developed in this study has also revealed that each category of facial feature contributes a certain amount of weightage to each personality trait. For example, the shape and size of a person's eyes may contribute more to their level of openness, while the shape and size of their nose may contribute more to their level of conscientiousness. This information can be useful for future research to explore the relationship between specific facial features and personality traits.

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 43 percent can be attributed to the relatively small size of the dataset. With only 138 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 Invigilator 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 may 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 accurately predict personality traits based on facial features.

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)
- 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