

# Personality Analysis and Prediction from Bangla Handwritten Text Documents

Darpan Bhattacharya  
under the guidance of  
Dr. Chayan Halder

Department of Computer Science  
Ramakrishna Mission Vivekananda Centenary College

April 27, 2024

- 1 Introduction
- 2 Literature Review
- 3 The Big Five Model
- 4 Dataset
- 5 Methodology
- 6 Results
- 7 Conclusion
- 8 References

## 1 Introduction

## 2 Literature Review

## 3 The Big Five Model

## 4 Dataset

## 5 Methodology

## 6 Results

## 7 Conclusion

## 8 References

- To create an automated system to predict the personality traits of an individual by using their Bangla handwriting.
- Why?

- To create an automated system to predict the personality traits of an individual by using their Bangla handwriting.
- Why?

There are many regions in rural Bengal where Bangla is still the primary language used for both verbal and written communication. Having an automated system which can automatically analyze an individual's personality from their Bangla handwriting, may have practical utility in many fronts, medical and otherwise, especially in such places.

- Even today, there is extremely limited information available both in and about Bangla on digital platforms.
- The Bangla script contains features like conjunct characters, matras etc., which make techniques like line and character segmentation very difficult to execute.
- Besides that, there are the challenges of skew, slant, etc., which make processing handwritten texts, Bangla or otherwise, challenging.

- Even today, there is extremely limited information available both in and about Bangla on digital platforms.
- The Bangla script contains features like conjunct characters, matras etc., which make techniques like line and character segmentation very difficult to execute.

বাংলাকার শিল্পশৈলী অগ্রয়াচারের বক্তুন, গৃহস্থানের  
সুষম-সমাদৃত বুক্সুটি কেবি বল্মীচন, শুণালুর মহামূর্তি  
ক্ষিপণালুত ফরেরাঙ্গের শুলুম্বাস্তু, ফ্লুকদের বাচ্চুতি  
জ্ঞানচিহ্ন ধুম্বুর, কাজারকদের গোপঘো প্রেইপস্ট-নিম্বণ্যাস্তু  
অভ্যন্তরিয়াবামী ব্যাপ্তির নজরাস্তু, খন্দাষ্টু, খন্তুত  
চমুজয়ী লিজীলিকগ শ্রীমাতী অক্ষয়টা রূপ্যাষ্টু পৰ্যয়াস্তু,  
হস্তক্ষেত্র বিশারণ ইহোতাত্ত্বিক ব্যালুম্বন, বৃষক্ষুত্তু, ভৱ  
বাঁদরদেশ কাষেকাটী পুর্ণজী, দেখন-কেবি, হু, হলু, কিছু,

- Besides that, there are the challenges of skew, slant, etc., which make processing handwritten texts, Bangla or otherwise, challenging.

- Even today, there is extremely limited information available both in and about Bangla on digital platforms.
- The Bangla script contains features like conjunct characters, matras etc., which make techniques like line and character segmentation very difficult to execute.
- Besides that, there are the challenges of skew, slant, etc., which make processing handwritten texts, Bangla or otherwise, challenging.

1 Introduction

2 Literature Review

3 The Big Five Model

4 Dataset

5 Methodology

6 Results

7 Conclusion

8 References

There have been various studies showing the link between handwriting and neurological aspects of humans.

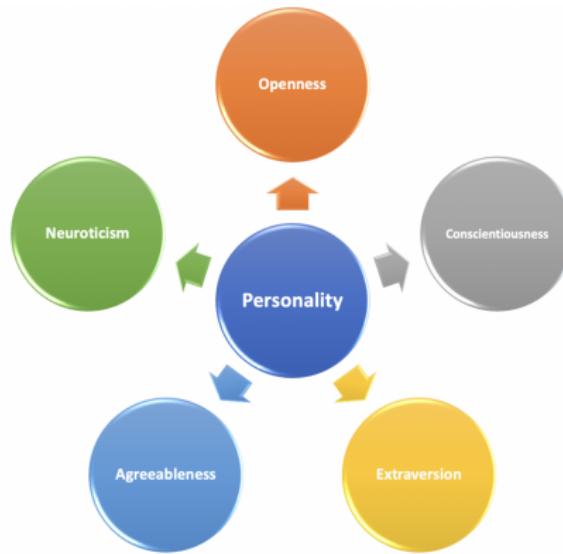
- Plamondon [Pla10] showed that the brain forms characters based on habits of writers and each neurological brain pattern forms a distinctive neuromuscular movement which is similar for individuals with the same type of personality.
- Gavrilescu and Vizireanu [GV18] proposed the first non-invasive three-layer architecture in literature based on neural networks that aimed to determine the Big Five personality traits of an individual by analyzing offline English handwriting.

- Biswas et al. [BRHR23] created the only known study of personality analysis in Bangla so far using handwritten Bangla characters. While their work utilizes various structural and morphological features of handwriting at a character level, our work focuses doing so at document level to then compare the extracted features with the Big Five personality traits of the same individuals.

- 1 Introduction
- 2 Literature Review
- 3 The Big Five Model
- 4 Dataset
- 5 Methodology
- 6 Results
- 7 Conclusion
- 8 References

# The Big Five Model

The Big Five personality traits, sometimes known as the five-factor model of personality or OCEAN model, is a grouping of five unique characteristics used to study personality. It has been developed from the 1980s onward in psychological trait theory.



- **Openness:** Openness to experience refers to ones willingness to try new things as well as engage in imaginative and intellectual activities.
- **Conscientiousness:** Conscientiousness describes a persons ability to regulate impulse control to engage in goal-directed behaviors. It measures elements such as control, inhibition, and persistence of behavior.
- **Extraversion:** Extraversion reflects the tendency and intensity to which someone seeks interaction with their environment, particularly socially.

- **Agreeableness:** Agreeableness refers to how people tend to treat relationships with others.
- **Neuroticism:** Neuroticism describes the overall emotional stability of an individual through how they perceive the world. It takes into account how likely a person is to interpret events as threatening or difficult. It also includes ones propensity to experience negative emotions.

1 Introduction

2 Literature Review

3 The Big Five Model

4 Dataset

5 Methodology

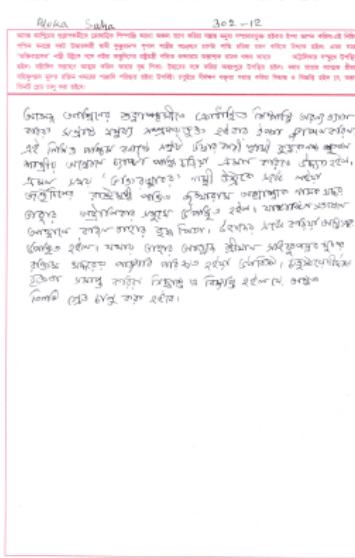
6 Results

7 Conclusion

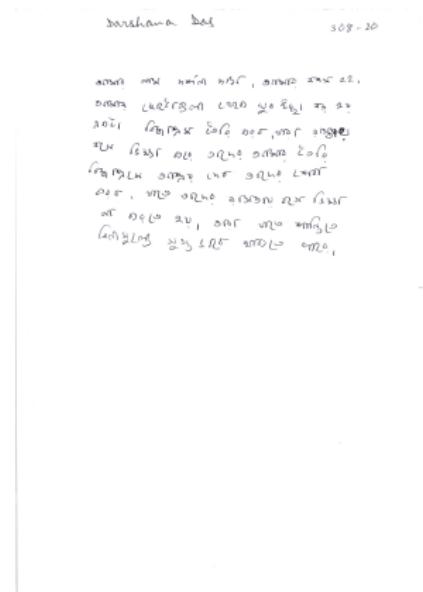
8 References

- For each individual, we utilized 13 data forms of guided handwriting where the individual had to hand-write the given text, and 1 data form of unrestricted handwriting where the individual was allowed to handwrite any text of their choice.
- We also utilized a Big Five personality trait of that individual which contained 44 close-ended and scaled questions, each of which could be answered in a scale from 1 to 5, 1 signifying strong disagreement and 5 signifying strong agreement, from which their Big-Five score was calculated.

## Data Samples



**Figure 1:** An example of a data sample containing guided handwriting



**Figure 2:** An example of a data sample containing unrestricted handwriting

# Data Samples

SL NO : SET 2

## আমার সাধারণ বৈশিষ্ট্য

এখানে এমন অনেকগুলি বৈশিষ্ট্য রয়েছে যা নিজের ক্ষেত্রে প্রযোগ হতেও পারে বা নাও হতে পারে। যেমন, আমি কি এমন একজন যে অনেক সাথে সময় কাটাতে পছন্দ করে?

যে বক্তব্যাটির সাথে কটো সম্মত বা অসম্মত তা নির্ধারণ করতে দয়া করে প্রতিটি বিবৃতিতে নথর লিখুন।

১ একবারেই একমত নই	২ সম্ভৃত একমত নই	৩ সম্মত নয় অসম্মতও নয়	৪ খানিকটা একমত	৫ সম্পূর্ণ একমত
-------------------------	------------------------	----------------------------------	----------------------	-----------------------

আমি এমন একজন যে.....

- ১। \_\_\_\_4\_\_\_\_ অতিরিক্ত কাছে বলি (বাচাল)।
- ২। \_\_\_\_4\_\_\_\_ অন্যের খুঁত ধরি।
- ৩। \_\_\_\_4\_\_\_\_ পর্যাপ্ত কাজে সক্ষম।
- ৪। \_\_\_\_2\_\_\_\_ হাতাশাগ্রহ, বিষম।
- ৫। \_\_\_\_4\_\_\_\_ নতুন ধারণা, চিন্তা - ভাবনা করতে সক্ষম।
- ৬। \_\_\_\_1\_\_\_\_ সংযত বা চাপা বভাবের।
- ৭। \_\_\_\_3\_\_\_\_ উপকারী এবং হার্ষপূর নই।
- ৮। \_\_\_\_3\_\_\_\_ কথনোচ্চ অসাবধানী।
- ৯। \_\_\_\_1\_\_\_\_ সাধারণত আরামপ্রিয় হলেও চাপ নিতে পারে।
- ১০। \_\_\_\_4\_\_\_\_ বিভিন্ন বিষয় সম্পর্কে কোতুহলী।

Figure 3: A portion of the Big Five personality test questionnaire

## 1 Introduction

## 2 Literature Review

## 3 The Big Five Model

## 4 Dataset

## 5 Methodology

- Extracting the handwritten portion to be analyzed
- Feature Extraction
- Regression Analysis

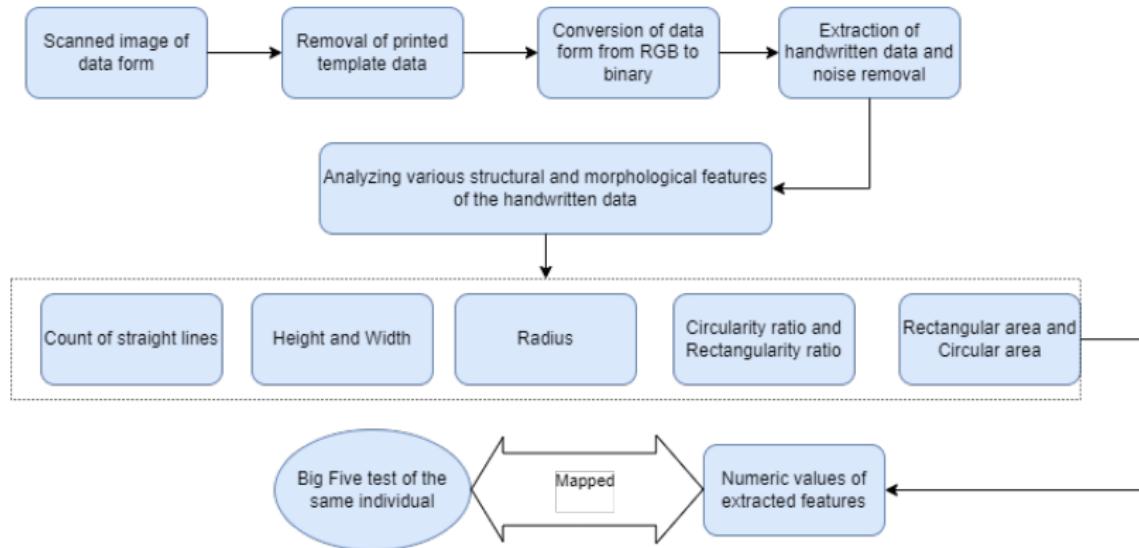
## 6 Results

## 7 Conclusion

In this study, we extracted various morphological and structural features of an individual's Bangla handwriting, and then used machine learning techniques to observe a correlation between the features obtained from their handwriting and their personality traits.

The following slide shows a flowchart demonstrating the workflow of our framework.

# Workflow



## 1 Introduction

## 2 Literature Review

## 3 The Big Five Model

## 4 Dataset

## 5 Methodology

- Extracting the handwritten portion to be analyzed
- Feature Extraction
- Regression Analysis

## 6 Results

## 7 Conclusion

- Firstly, we had to separate the handwritten portion from the data sample.
- Since the template data was in red, we could remove that by removing all segments of red colour.
- We then used erosion and dilation techniques to merge the components to filter out the noise and extract the handwritten data.
- Finally, we converted the handwritten data into binary for feature extraction.

- Firstly, we had to separate the handwritten portion from the data sample.
- Since the template data was in red, we could remove that by removing all segments of red colour.
- We then used erosion and dilation techniques to merge the components to filter out the noise and extract the handwritten data.
- Finally, we converted the handwritten data into binary for feature extraction.

- Firstly, we had to separate the handwritten portion from the data sample.
- Since the template data was in red, we could remove that by removing all segments of red colour.
- We then used erosion and dilation techniques to merge the components to filter out the noise and extract the handwritten data.
- Finally, we converted the handwritten data into binary for feature extraction.

## Extracting the handwritten portion to be analyzed

- Firstly, we had to separate the handwritten portion from the data sample.
- Since the template data was in red, we could remove that by removing all segments of red colour.
- We then used erosion and dilation techniques to merge the components to filter out the noise and extract the handwritten data.
- Finally, we converted the handwritten data into binary for feature extraction.

## A visual representation

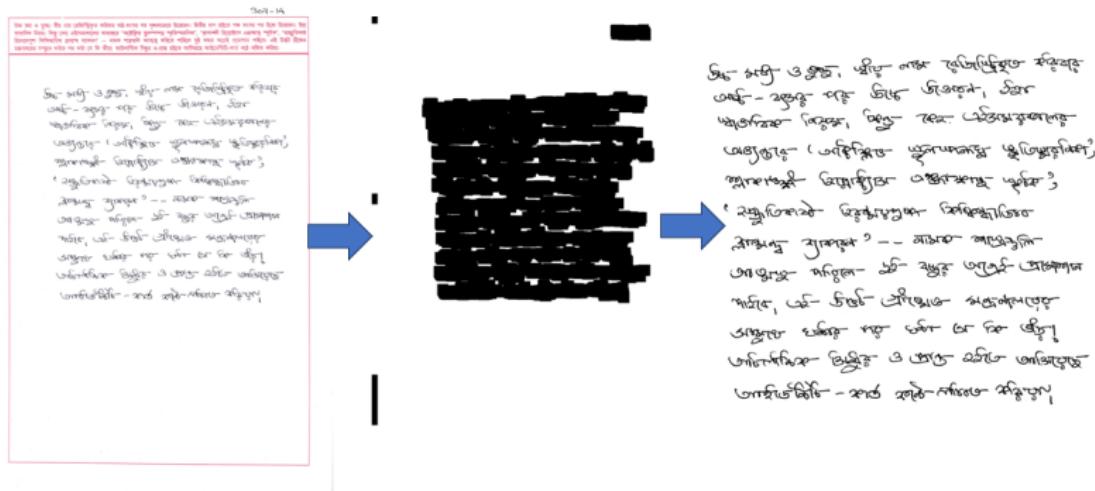


Figure 4: Data Sample to eroded and dilated diagram to binary image

## 1 Introduction

## 2 Literature Review

## 3 The Big Five Model

## 4 Dataset

## 5 Methodology

- Extracting the handwritten portion to be analyzed
- **Feature Extraction**
- Regression Analysis

## 6 Results

## 7 Conclusion

We extracted the following structural and morphological features from the data:

- **Straightness (S):** Here we tried to analyze the horizontal straightness of the handwriting. We did that by running a small horizontal mask over the entire handwritten data, and counting the number of significant overlaps with the mask in the data.
- **Height (H) and Width (W):** We calculated the median height and width of the handwriting in each data sample. We did that by computing the height and width of each connected component of the handwriting and then computing their median. Extremely small connected components were discarded as noise.

We extracted the following structural and morphological features from the data:

- **Straightness (S):** Here we tried to analyze the horizontal straightness of the handwriting. We did that by running a small horizontal mask over the entire handwritten data, and counting the number of significant overlaps with the mask in the data.
- **Height (H) and Width (W):** We calculated the median height and width of the handwriting in each data sample. We did that by computing the height and width of each connected component of the handwriting and then computing their median. Extremely small connected components were discarded as noise.



- **Radius (R)**: We calculated the median radius of the handwriting in each data sample. We did that by computing the minimum radius that could enclose a connected component for each connected component and then computed their median.

- **Rectangularity Ratio (RR):**

$$RR = \frac{\text{Contour area of a particular component}}{\text{Area of the smallest rectangle enclosing the component}}$$

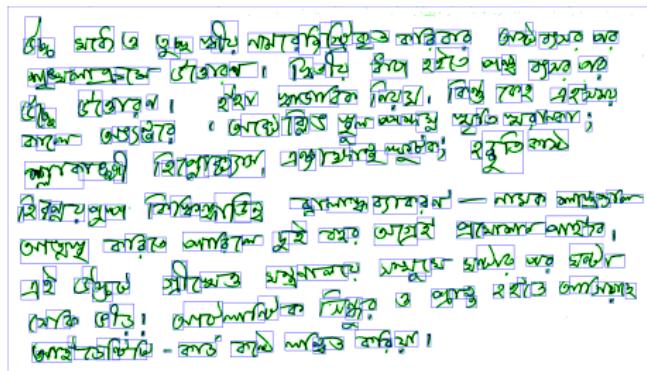


Figure 5: An image demonstrating calculation of rectangularity ratio (Contours are in green, rectangles are in blue)

- **Circularity Ratio (RR):**

$$CR = \frac{\text{Contour area of a particular component}}{\text{Area of the smallest circle enclosing the component}}$$

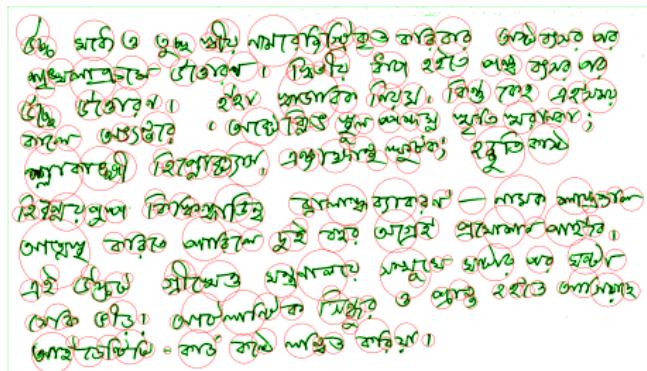


Figure 6: An image demonstrating calculation of circularity ratio (Contours are in green, circles are in red)

We extracted the following structural and morphological features from the data:

- **Rectangular Area (RA):** We calculated the rectangular area (RA) by computing the area of the rectangle formed by the median height (H) and the median width (W) of the components.

$$RA = H \cdot W$$

- **Circular Area (CA):** We calculated the circular area (CA) by computing the area of the circle formed by the median radius (R) of the components.

$$CA = \pi \cdot R^2$$

We extracted the following structural and morphological features from the data:

- **Rectangular Area (RA):** We calculated the rectangular area (RA) by computing the area of the rectangle formed by the median height (H) and the median width (W) of the components.

$$RA = H \cdot W$$

- **Circular Area (CA):** We calculated the circular area (CA) by computing the area of the circle formed by the median radius (R) of the components.

$$CA = \pi \cdot R^2$$

## 1 Introduction

## 2 Literature Review

## 3 The Big Five Model

## 4 Dataset

## 5 Methodology

- Extracting the handwritten portion to be analyzed
- Feature Extraction
- Regression Analysis

## 6 Results

## 7 Conclusion

We used the RandomForestRegressor and MultiOutputRegressor from the `skicit-learn` module [PVG<sup>+11</sup>] for our purpose. We split our dataset into training and testing sets in a 70:30 ratio.

In the Big Five personality test that was conducted, each individual obtained a score out of 44 for each personality trait. For our regressor models, we normalized the score for each personality trait.

1 Introduction

2 Literature Review

3 The Big Five Model

4 Dataset

5 Methodology

6 Results

7 Conclusion

8 References

We obtained a Root Mean Squared Error (RMSE) of approximately  $0.15(\pm 0.005)$  for prediction on the entire data set, both for the RandomForestRegressor and MultiOutputRegressor models. Table 1 gives the RMSE distributions for the different personality traits, with an adjustment of approximately  $\pm 0.005$  for different executions.

## Result Table

Trait	RandomForestRegressor	MultiOutputRegressor
Openness	0.1132	0.1227
Conscientiousness	0.1490	0.1550
Extraversion	0.1334	0.1548
Agreeableness	0.1894	0.1921
Neuroticism	0.1678	0.1665
<b>Average</b>	<b>0.1506</b>	<b>0.1582</b>

Table 1: RMSE distributions of the different personality traits

The following are charts which depict the comparison between the original values and the predicted values. For the sake of visual clarity, comparisons have been provided for 15 randomly picked data samples from the dataset.

# Openness Chart

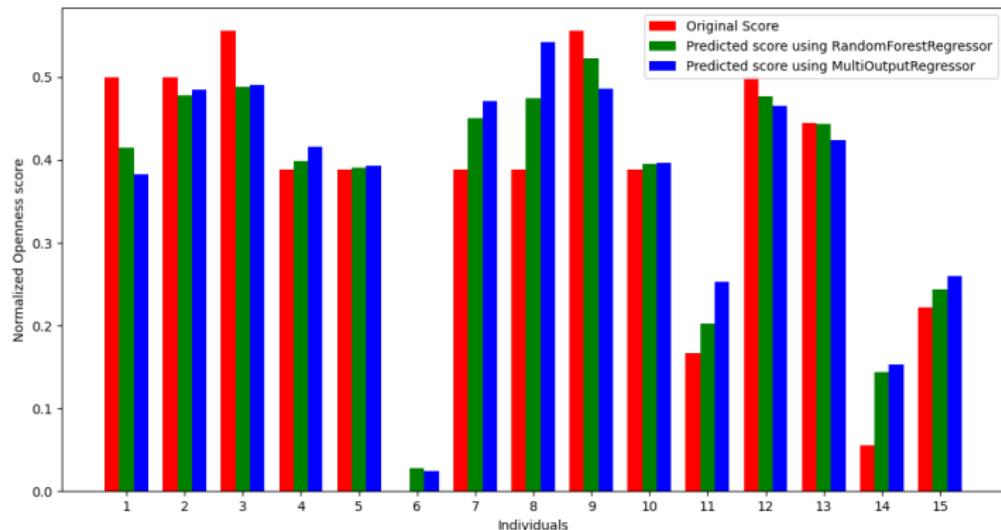


Figure 7: Original v/s predicted values for Openness personality trait

# Conscientiousness Chart

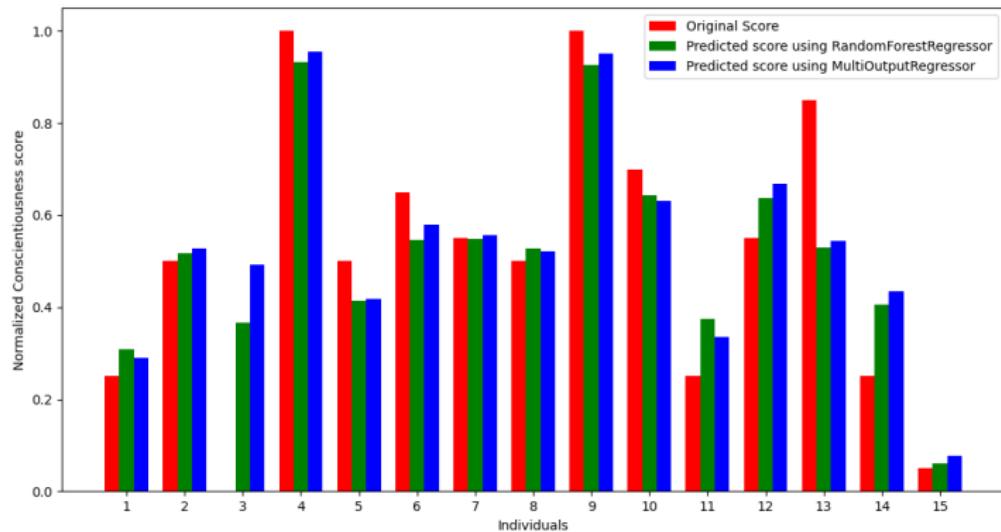


Figure 8: Original v/s predicted values for Conscientiousness personality trait

# Extraversion Chart

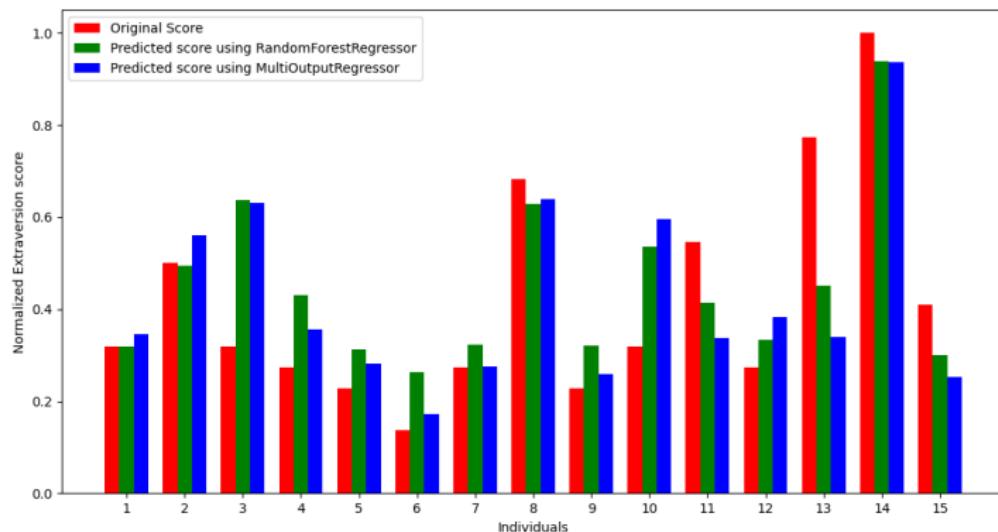


Figure 9: Original v/s predicted values for Extraversion personality trait

# Agreeableness Chart

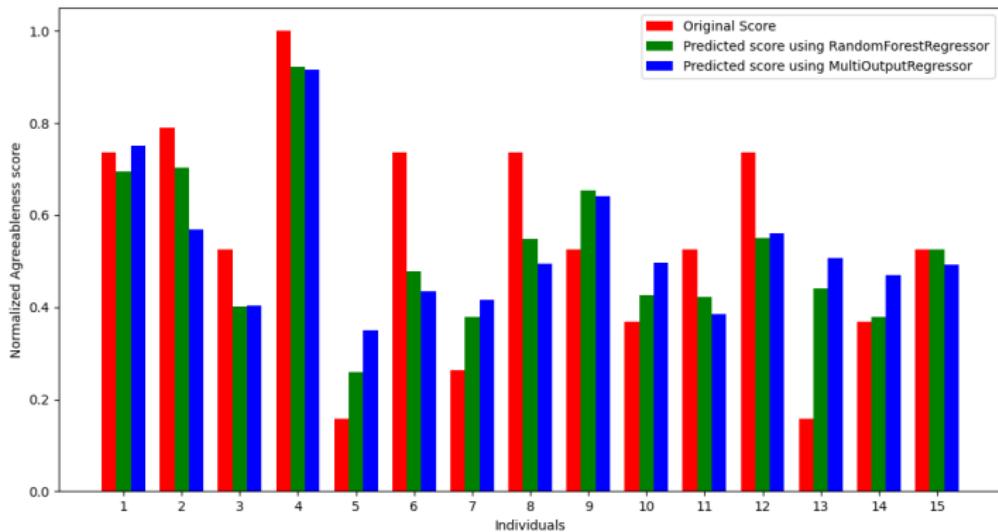


Figure 10: Original v/s predicted values for Agreeableness personality trait

# Neuroticism Chart

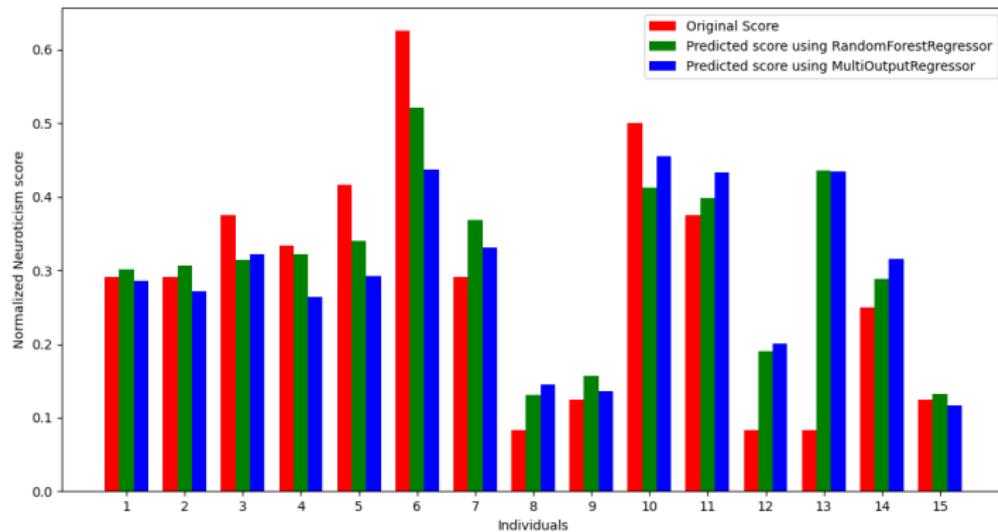


Figure 11: Original v/s predicted values for Neuroticism personality trait

1 Introduction

2 Literature Review

3 The Big Five Model

4 Dataset

5 Methodology

6 Results

7 Conclusion

8 References

- In future, the amount of data samples that are analyzed should be increased for providing a much more accurate picture of the correlation between handwriting and Big Five personality traits.
- Line and character segmentation can be done on the handwritten data to provide better and more precise morphological features for analysis. The data can be tested on various other regressor models to obtain more precise results. Our study can also be extended to various other languages as well.

From our results, we can conclude that analyzing Bangla handwriting on a document level can be quite satisfactorily compared with their Big Five personality traits. Our results show that our methods are quite promising and have a very high potential of being used in real-life situations to identify an individual's personality traits from their handwriting.

## 1 Introduction

## 2 Literature Review

## 3 The Big Five Model

## 4 Dataset

## 5 Methodology

## 6 Results

## 7 Conclusion

## 8 References

# References

- [big] <https://www.simplypsychology.org/big-five-personality.html>.  
[Accessed 26-04-2024].
- [BRHR23] Suparna Saha Biswas, Arpan Roy, Chayan Halder, and Kaushik Roy.  
Personality analysis for handwritten documents: a case study with bangla characters.  
*Innovations in Systems and Software Engineering*, Oct 2023.
- [GV18] Mihai Gavrilescu and Nicolae Vizireanu.  
Predicting the big five personality traits from handwriting.  
*EURASIP Journal on Image and Video Processing*, 2018, 07 2018.
- [Pla10] Réjean Plamondon.  
Neuromuscular studies of handwriting generation and representation.  
pages 261 – 261, 12 2010.
- [PVG<sup>+</sup>11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay.  
Scikit-learn: Machine learning in Python.  
*Journal of Machine Learning Research*, 12:2825–2830, 2011.

# Thank You