International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 06, pp. 10-17, June

ABC Based Relational Graph Network for Personality Trait Prediction From Selfie-Derived Tabular Data

S. Prakash[3], A. Anish Rama[4]

[3]Assistant Professor, Department of Electrical and Electronics Engineering,
Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation,
Chennai, Tamil Nadu, India.
[4]Assistant Professor, Department of Electronics and Communication Engineering,
Arunachala College of Engineering for Women,
Tamil Nadu, India.

Corresponding Author E-mail:

ABSTRACT: Personality Trait (PT) predictions from facial images have gained increasing attention in intelligent human-computer interaction and behavioural analysis for analysis. In this paper an Artificial Bee Colony-based Relational Graph Network (ABC-RGN) is proposed for PT prediction using selfie-derived tabular data. The data pre-processing based on data cleaning, missing-value as well as duplicate removal, followed by outlier detection and data encoding to improve data quality. The processed features are normalized and analysed through feature correlation analysis to identify the most relevant personality-related attributes. Subsequently, the RGN models interactions among different feature types and captures complex relationships within the selfie-derived tabular data. The selected features are then optimized using the ABC algorithm, which improves the selection and representation of informative feature relationships. Experimental results is demonstrated in Python software using synthetic personality dataset the proposed structure attains an accuracy, recall as well as F1-score of 97.6% precision of 97.7% outperforming existing PT methods. The proposed ABC-RGN framework provides an effective approach for learning complex feature relationships and improving the reliability of PT prediction from selfie-based data.

Keywords: Personality Trait, Artificial Bee Colony, Relational Graph Network, Selfie-Derived Tabular Data.

1. Introduction

Personality is an important psychological dimension that helps to describe consistent patterns in human behaviour through a set of relatively stable and measurable characteristics [1]. It influences various aspects of human life, including preferences, emotions, motivation, lifestyle, thoughts, and overall well-being [2-3]. Consequently, PT supports numerous practical applications, such as recommendation systems, social network analysis, sentiment analysis, and early detection of psychological risks [4]. Personality assessment relied on questionnaires and interviews, which is time-consuming, costly, and sometimes subjective [5-6]. Furthermore, with the rapid growth of social media and online platforms, automatic personality recognition [7] from user-generated content and images.

2. Related Work

Several techniques are utilized for PT prediction. In [8] Barik et al (2025) have presented Enhanced Grid Search Optimization–Convolutional Neural Network (EGSO-CNN) technique for web phishing recognition. EGSO have search space and avoid unnecessary parameter combinations. However, evaluating multiple parameter combinations significantly increase optimization time. Consequently, in [9] Serrano-Guerrero et al (2024) have presented Stacked Ensemble Model for PT prediction. It combines multiple models to obtain more reliable predictions. Nevertheless, it is difficult to determine which individual model contributes most to the final personality prediction. Additionally, in [10] Naz et al (2024) have developed Bidirectional Long Short-Term Memory (Bi-LSTM) to predict the Extraversion PT. It captures information from both past and future contexts. Nonetheless, the dual-direction architecture increases training complexity. Furthermore, in [11] Ashawa et al (2025) have invented CNN–Random Forest (CNN-RF) for PT prediction. CNN automatically learns relevant features from the input data. However, CNN and RF make the system more difficult to design and tune. Moreover, in [12] Kalita et al (2025) have proposed Bi-LSTM to predict students second-term Grade Point Average (GPA) by learning patterns and dependencies from their academic performance data. Nevertheless, it overfit when the available student dataset is small.

2.1 Contribution of the work

- Data pre-processing by data cleaning, missing-value handling, duplicate removal, outlier detection, and data encoding are performed to improve data quality and consistency.
- Normalization based on feature correlation analysis is used to identify the most informative attributes associated with PT and reduce irrelevant information.
- Proposed RGN captures interactions and complex relationships among different types of selfie-derived features.
- The ABC algorithm is integrated to optimize the selected feature representation and identify an effective combination of informative features.

3. Proposed Methodology

To predict PT from selfie-derived tabular data at first data pre-processing based on data cleaning the input data are cleaned by handling missing values, removing duplicate records, detecting outliers, and encoding unqualified attributes to improve data quality and consistency. Then feature normalization is normalized to bring different attributes into a consistent numerical range and facilitate effective model learning. Subsequently, normalization based on feature correlation analysis is done to identify the most relevant personality-related features and reduce the influence of irrelevant or redundant attributes. The proposed RGN are signified as interconnected nodes, allowing the RGN to learn relationships and interactions among different feature types. Furthermore to improve ABC algorithm is used to optimize the feature representation and identify an effective combination of informative features. The optimization improves the quality of the input provided to the RGN.

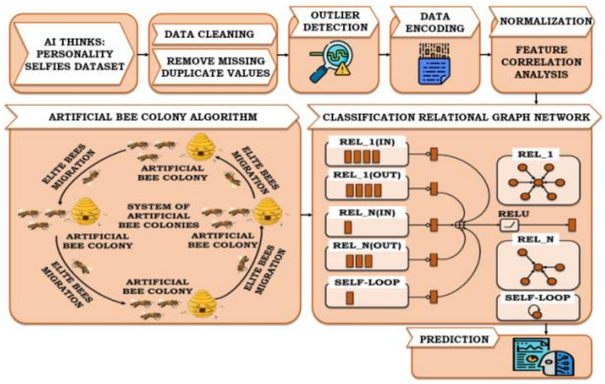

Figure 1: *Block diagram for proposed ABC-RGN*

3.1 Data pre-processing

For PT prediction the data pre-processing is done by data cleaning, removing errors, handling missing values, eliminating duplicate and abnormal records, and converting categorical information into numerical form selfie-derived data. Data cleaning is performed to identify and remove inconsistent records, whereas missing-value handling is used to maintain dataset completeness. Duplicate samples are removed to prevent biased model learning, and outlier detection is developed to identify abnormal observations that negatively affect feature relationships. This improves data quality, consistency, and model reliability during PT prediction.

3.2 Normalization

Normalization based on feature correlation analysis observes the relationships between the available selfie-derived attributes and PT. Features showing stronger relationships with PT are retained as informative attributes, whereas redundant features are removed. This helps reduce unnecessary information and allows the ABC-RGN model to focus on the features most relevant to personality prediction.

3.3 Relational Graph Network

For PT prediction the proposed RGN is used to model the interactions among different types of features extracted from selfie-derived tabular data. Figure 2 displays the architecture for RGN.

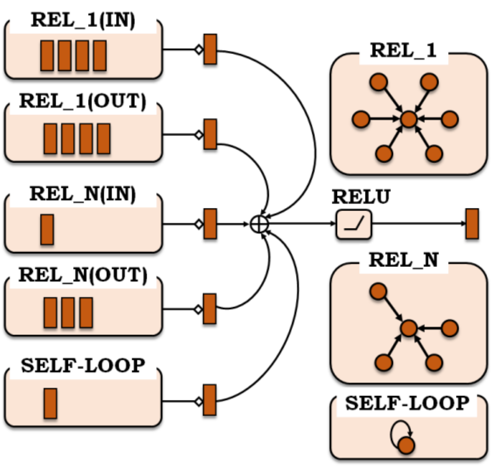

Figure 2: *Architecture for RGN*

The RGN represents the features as nodes in a graph, where the connections between nodes describe relationships between related attributes. This allows the model to learn both individual feature characteristics and their interactions, selfie-derived feature set be signified as:

$$G = (V, E, R) \qquad (1)$$

Where, V signifies the feature nodes, E denotes the connections between nodes, then R indicates different relationship types. Every node v_i have an initial feature designh_i.

$$h_i^{(l+1)} = \sigma(W_0^{(l)} h_i^{(l)} + \sum_{r\in R} \sum_{j\in N_i^T} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)}) \qquad (2)$$

Where, $h_i^{(l+1)}$indicates the representation of node i at layer l, N_i^T signifies the neighbours of node i joined through relation r, $W_r^{(l)}$indicates the learnable conversion matrix for relation r, $c_{i,r}$ denotes a normalization factor, and σ is the activation function. If attention is integrated into the RGN, the importance of a neighbouring feature is determined by:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\in N_i} \exp(e_{ik})} \qquad (3)$$

Where, e_{ij} signifies the relevance between feature nodes i as well as j. the aggregated representation is expressed as

$$h_i' = \sigma(\sum_{j\in N_i^T} \alpha_{ij} W h_j) \qquad (4)$$

The graph demonstration is finally passed to a prediction layer:

$$\hat{y} = softmax(W_o h_G + b_o) \qquad (5)$$

Where, W_o & $b_o$denotes the learnable parameters and $\hat{y}$ signifies the predicted PT probabilities. The RGN captures complex nonlinear dependencies among selfie-derived features that effectively represented when the attributes are considered independently.

Artificial Bee Colony

For PT prediction the ABC algorithm simulator the natural hunting behaviour of honey bees to resolve difficult optimization difficulties The ABC algorithm contains three main types of bees employed bees here each bee is assigned to a particular food source and evaluates its quality whereas searching for improved solutions

i) Scout bees search for new food sources when an existing source does not provide improvement, helping maintain diversity in the search process.

$$x_{i,j} = x_{min,j} + rand(0,1) \times (x_{max,j} - x_{min,j}) \quad (6)$$

Where, $x_{i,j}$ denotes location of bee i in measurement$j$, $x_{max,j}$ & $x_{min,j}$ indicates maximum as well as minimum restrictions in measurement$j$, $rand(0,1)$ signifies even casual numbers among0 & 1.

ii) Movements of employed and onlooker bees

Onlooker bees observe the information shared by employed bees and select promising food sources based on their quality

$$v_{i,j} = x_{i,j} + \varphi_{i,j} \times (x_{i,j} - x_{k,j}) \quad (7)$$

Where, $v_{i,j}$ denotes novel suggested location of bee i in measurement$j$, $x_{k,j}$ indicates casual tuning factors in array [-1, 1], $\varphi_{i,j}$signifies location of additional bee k casually designated in measurement j.The flowchart for ABC is shown in figure 3.

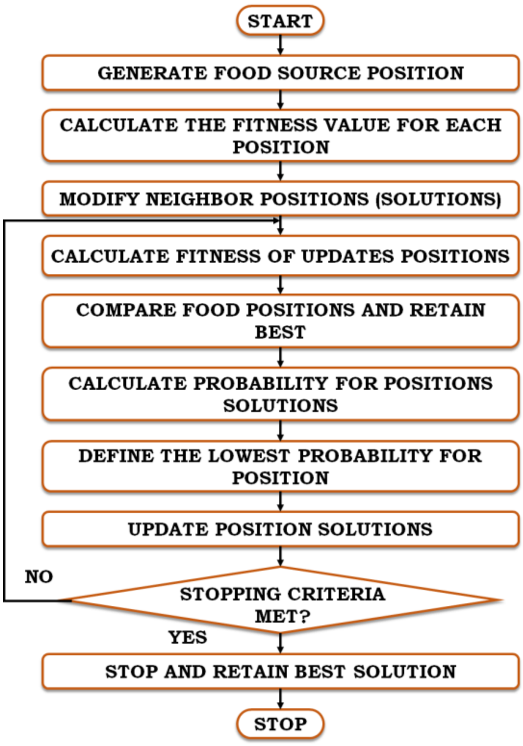

Figure 3: *Flowchart for ABC*

iii) The probability proportional is expressed as

$$P_i = \frac{f_i}{\sum_{i=1}^{n} f_i} \quad (8)$$

$$if\ f(v_i) < f(x_i)\ then\ x_i = v_i \quad (9)$$

$$f_0 = 0.9 \times \alpha + 0.5 \times \beta \quad (10)$$

Where, α denotes the inverse of the matrix and β indicates randomly generated values among 0 & 1. Thus the proposed ABC-RGN model used to reduce redundant information, and provides a more discriminative feature representation for accurate PT prediction.

4. Result and Discussion

For the proposed ABC-RGN based PT prediction it is implemented in Python software using synthetic personality dataset with 1,500 synthetic samples is used for model development and evaluation. The experimental results are interpreted as a proof-of-concept evaluation of the proposed framework rather than as evidence of generalization to real-world human personality assessment. The dataset columns are target X as image_id, user_selfie, extroversion score, confidence level, trustworthiness score, smile intensity, eye_contact_score, model_used, age range, gender, and target Y as emotion category.

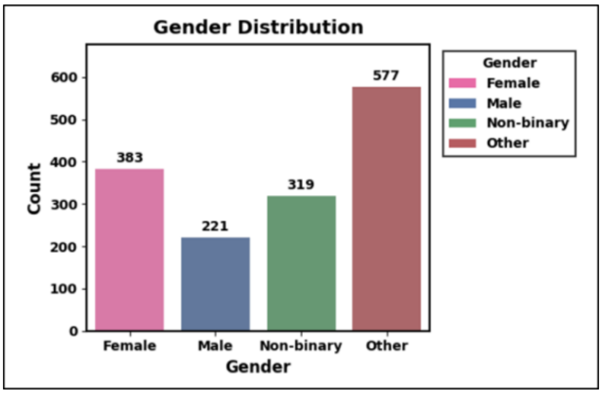

Figure 4: *Gender distribution*

Figure 4 displays the gender distribution. It have the count with categories for female of 383, male of 221, non-binary of 319, and other of 577.Highest Count: The other category have the highest count at 577 and lowest count is for male category have the lowest count at 221.

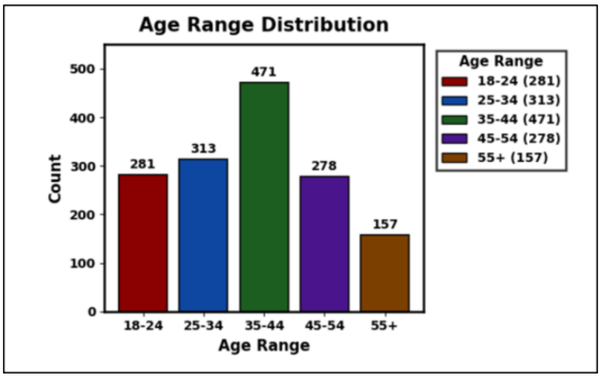

Figure 5: *Age range distribution*

Figure 5 shows the age range distribution. From age 18-24 it have the count of 281, from 25-34 it have the count of 313, 35-44 it have the count of 471, 45-54 it have the count of 278 and from age 50+ it have the count of 157 respectively.

Figure 6: *Smile intensity distribution*

Figure 6 displays the smile intensity distribution. The X-axis signifies smile intensity ranging from 0.0 to 1.0 and Y-axis denotes the frequency of occurrences for each intensity bin. The data features with three distinct peaks are around the 0.15, 0.58, and 0.85 intensity marks, indicated by the overlapped blue line.

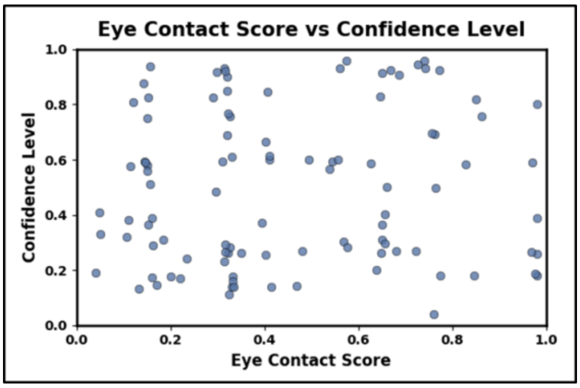

Figure 7: *Eye Contact score vs confidence level*

Figure 7 shows the eye Contact score vs confidence level. The data points are scattered evenly across the graph area without forming an upward or downward trend. And vertical clustering occurs at specific eye contact scores around 0.15, 0.33, 0.40, and 0.65, representing discrete measurements for those values.

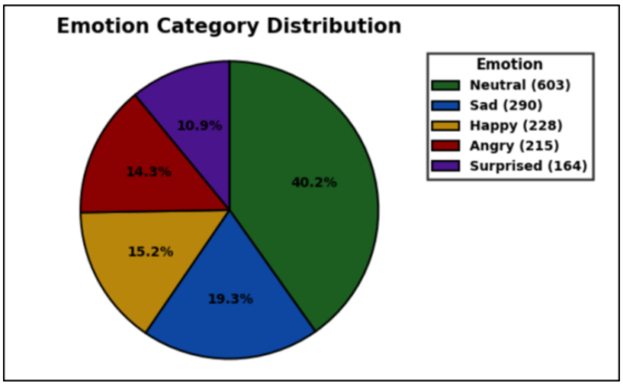

Figure 8: *Emotion category distribution*

Figure 8 displays the emotion category distribution. The neutral have the emotion value of 603 with 40.2%, sad have the emotion value of 290 with 19.3%, Happy have the emotion value of 228 with 15.2%, angry have the emotion value of 215 with 14.3%, and surprised have the emotion value of 164 with 10.9% respectively.

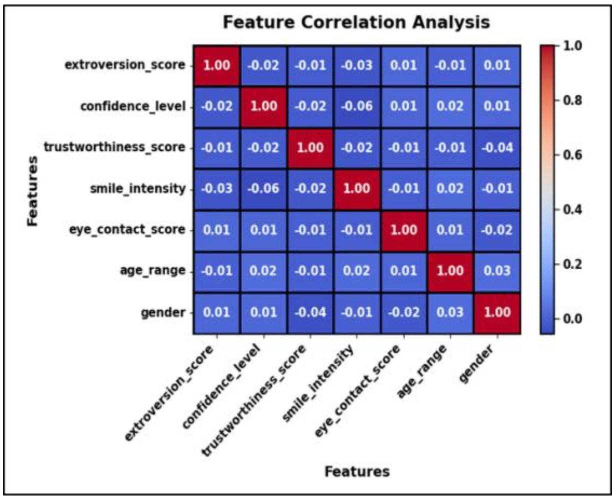

Figure 9: *Feature Correlation analysis*

Figure 9 shows the feature correlation analysis. All features have very weak correlations approximately −0.06 to 0.03 with each other, demonstrating minimal multicollinearity and largely independent information for PT prediction.

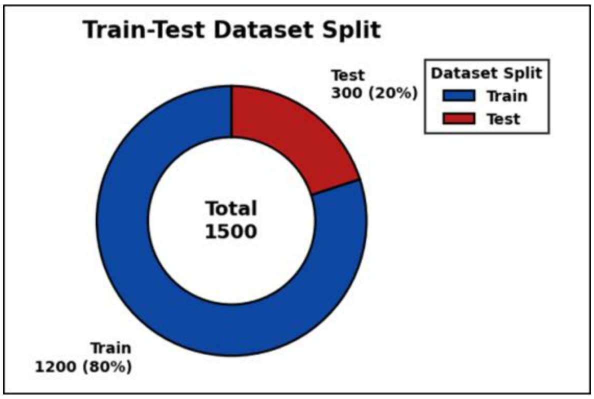

Figure 10: *Train-Test dataset split*

Figure 10 displays the train-test dataset split. The total data is 1500 in that tests have the value of 300 with 20% and train have the value of 1200 with 80% for PT prediction.

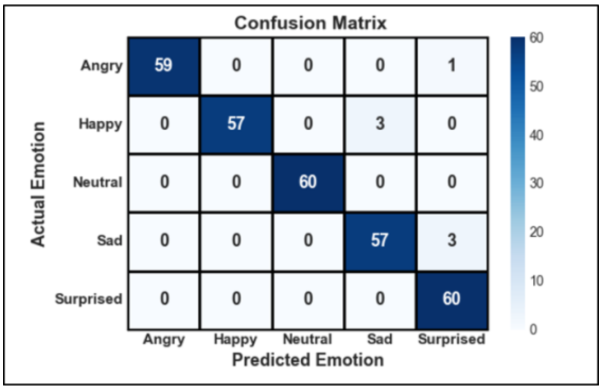

Figure 11: *Confusion matrix*

Figure 11 displays the confusion matrix using proposed ABC-RGN for PT prediction. The angry have the value of 59, happy and sad have the value of 57 and neutral and surprised have the value of 60 respectively.

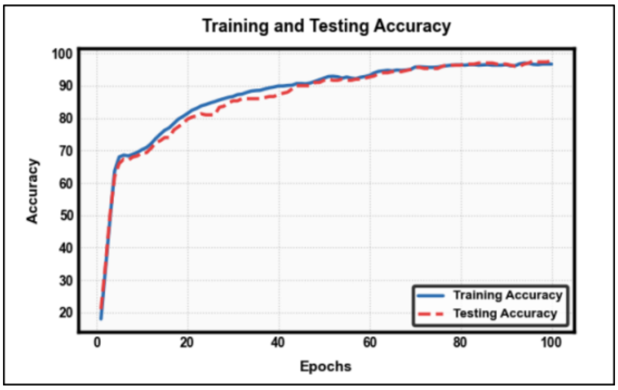

Figure 12: *Testing and training accuracy*

Figure 12 shows the training and testing accuracy. It increases gradually from approximately 18–19% to 97–98% over 100 epochs, with the two curves remaining close, signifying stable learning and good generalization with minimal overfitting.

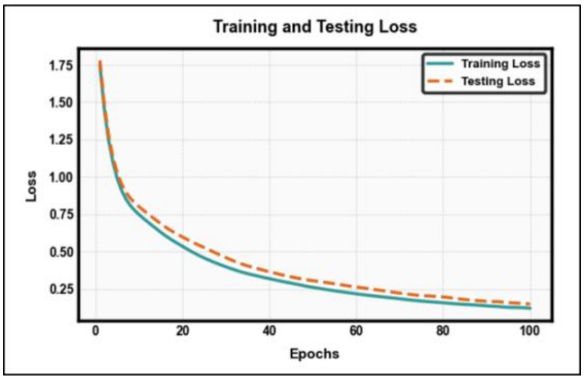

Figure 13: *Training and testing loss*

Figure 13 displays training as well as testing loss. The testing loss excesses only slightly higher than the training loss throughout the entire training process. Testing loss does not start to increase early whereas the training loss continues drop to 0.25, which indicate overfitting.

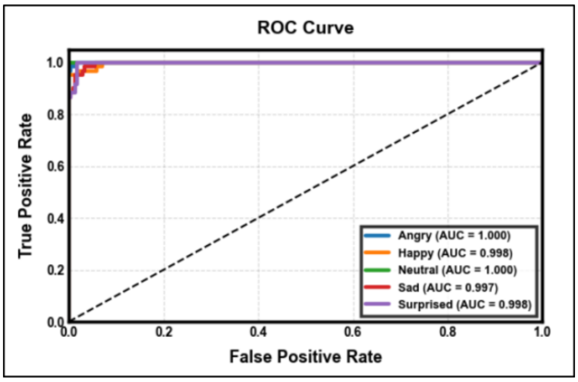

Figure 14: *ROC curve*

Figure 14 displays the ROC curvature for PT prediction using ABC-RGN. The angry and neutral have the AUC of 1 and happy, sad and surprised have the AUC value of 0.99 respectively.

4.1 Comparison with the Proposed Technique

In the evaluation with the proposed technique the proposed technique is compared with the existing techniques as presented in table 1. Liu et al [7] have the accuracy of 86.7%, precision of 90.5%, recall of 86.7 % and F1-score of 87% Naz et al [10] have the accuracy of 92.52, precision of 91.67%, recall of 93.44%, and F1-score of 92.49%, Fitrah et al [3] have the accuracy of 82.23%, precision of 92.02%, recall of 92.11%

and F1-score of 91.98, Jayaprakash et al [5] have the accuracy of 97%, precision of 93%, recall of 92% and F1- score of 92% and the Proposed Work have the higher value with precision of 97.7% and accuracy, recall and F1-score of 97.6% using proposed ABC-RGN for PT prediction.

Table 1: *Comparison with the Proposed Technique*

Study	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Liu et al [7]	86.7	90.5	86.7	87
Naz et al [10]	92.52	91.67	93.44	92.49
Fitrah et al [3]	82.23	92.02	92.11	91.98
Jayaprakash et al [5]	97	93	92	92
Proposed Work	97.6	97.7	97.6	97.6

5. Conclusion

In this paper the proposed ABC-RGN framework is used for PT prediction using selfie-derived tabular data. The input data quality is cleaned, missing values are handled, duplicate data's are removed, outliers are detection, encoding process is utilized, most relevant personality-related attributes are analysed using feature correlation, ABC-based feature optimization, and RGN-based relational learning identify informative features as well as capture complex relationships among personality-related attributes. Experimental evaluation using a synthetic personality dataset in Python demonstrates that the proposed framework achieves 97.6% accuracy, 97.7% precision, 97.6% recall, as well as 97.6% F1-score, outperforming the considered existing PT prediction methods. The proposed ABC-RGN technique provides reliable and accurate personality prediction whereas effectively utilizing relationships among selfie-derived features. The future work focus on validating the framework using larger real-world facial datasets and extending it to real-time personality assessment

References

1. Wentao Hu; Zichen Shao, Year: 2025, "Design and evaluation of a GenAI-based personalized educational content system tailored to personality traits and emotional responses for adaptive learning", Computers in Human Behavior Reports, Vol: 19, pp. 100735.
2. Yan Cheng; Huiling Chen; Xuanfeng Guo; Bo Ding; Zhipeng Pei, Year: 2026, "Personality-aware emotion recognition in conversation with large language models", Pattern Recognition, Vol: 178, pp. 113460.
3. Muh Fitrah; Anastasia Sofroniou; Novi Yarmanetti; Indriani H. Ismail; Hetty Anggraini; Ita Chairun Nissa; Bakti Widyaningrum; Irul Khotijah; Prabowo Dwi Kurniawan; Dedi Setiawan, Year: 2025, "Are teachers ready to adopt deep learning pedagogy? The role of technology and 21st-century competencies amid educational policy reform", Education sciences, Vol: 15, No: 10, pp. 1344.
4. Hans JG Hassell; Gary E. Hollibaugh Jr; Matthew R. Miles, Year: 2026, "Not just elections: personality traits and ambition for political office", Political Behavior, Vol: 48, No: 1, pp. 97-117.
5. D. Jayaprakash; C.S. Kanimozhiselvi, Year: 2024, "Multinomial logistic regression method for early detection of autism spectrum disorders", Measurement: Sensors, Vol: 33, pp. 101125.
6. Pranavi Reddy, Year: 2024, "Diagnosis of autism in children using deep learning techniques by analyzing facial features", Engineering Proceedings, Vol: 59, No: 1, pp. 198.
7. Yu Liu; Tangzhi Liu; Xingliang Liu; Song Wang; Wei Wu; Changjun Wang; Tong Liu, Year: 2026,

"Driver risk-level identification incorporating personality traits, demographic characteristics, and driving behaviors", Traffic Injury Prevention, Vol: 27, No: 5, pp. 535-545.

8. Kousik Barik; Sanjay Misra; Raghini Mohan, Year: 2025, "Web-based phishing URL detection model using deep learning optimization techniques", International Journal of Data Science and Analytics, Vol: 20, No: 5, pp. 4449-4471.
9. Jesus Serrano-Guerrero; Bashar Alshouha; Mohammad Bani-Doumi; Francisco Chiclana; Francisco P. Romero; Jose A. Olivas, Year: 2024, "Combining machine learning algorithms for personality trait prediction", Egyptian Informatics Journal, Vol: 25, pp. 100439.
10. Anam Naz; Hikmat Ullah Khan; Sami Alesawi; Omar Ibrahim Abouola; Ali Daud; Muhammad Ramzan, Year: 2024, "AI knows you: deep learning model for prediction of extroversion personality trait", IEEE Access, Vol: 12, pp. 159152-159175.
11. Moses Ashawa; Joshua David Bryan; Nsikak Owoh, Year: 2025, "Personality prediction model: An enhanced machine learning approach", Electronics, Vol: 14, No: 13, pp. 2558.
12. Emi Kalita; Abdullah Mana Alfarwan; Houssam El Aouifi; Ashima Kukkar; Sadiq Hussain; Tazid Ali; Silvia Gaftandzhieva, Year: 2025, "Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation", In Frontiers in Education, Vol: 10, pp. 1581247.