Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.067490

ARTICLE

An Efficient Deep Learning-Based Hybrid Framework for Personality Trait Prediction through Behavioral Analysis

Nareshkumar Raveendhran and Nimala Krishnan*

Department of Networking and Communications, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, 603203, Tamil Nadu, India

*Corresponding Author: Nimala Krishnan. Email: nimalak@srmist.edu.in

Received: 05 May 2025; Accepted: 14 July 2025; Published: 23 September 2025

ABSTRACT: Social media outlets deliver customers a medium for communication, exchange, and expression of their thoughts with others. The advent of social networks and the fast escalation of the quantity of data have created opportunities for textual evaluation. Utilising the user corpus, characteristics of social platform users, and other data, academic research may accurately discern the personality traits of users. This research examines the traits of consumer personalities. Usually, personality tests administered by psychological experts via interviews or self-report questionnaires are costly, time-consuming, complex, and labour-intensive. Currently, academics in computational linguistics are increasingly focused on predicting personality traits from social media data. An individual's personality comprises their traits and behavioral habits. To address this distinction, we propose a novel LSTM approach (BERT-LIWC-LSTM) that simultaneously incorporates users' enduring and immediate personality characteristics for textual personality recognition. Long-term Personality Encoding in the proposed paradigm captures and represents persisting personality traits. Short-term Personality Capturing records changing personality states. Experimental results demonstrate that the designed BERT-LIWC-LSTM model achieves an average improvement in accuracy of 3.41% on the Big Five dataset compared to current methods, thereby justifying the efficacy of encoding both stable and dynamic personality traits simultaneously through long- and short-term feature interaction.

KEYWORDS: Personality; deep learning; online social network; LSTM; big five

1 Introduction

The popularity of the Internet is increasing globally due to the rapid advancement in computer technology. Users have access to social media platforms that give them a venue for interaction, opinion sharing, and communication [1]. People's everyday lives have become completely reliant on social network services, and online communication has become the primary mode of communication. In terms of human differences in identifiable thought, emotion, and behaviors patterns, we speak of Personality. A person's Personality is a blending of emotions, attitudes, and general attributes [2]. Personality traits and qualities tremendously impact our lives; they influence our preferences for luxuries, culture, healthcare, and various other needs. The connection of individuals through online platforms has greatly influenced the progress of information technology [3]. These traces are primarily text-based, such as public posts and comments and other media types like movies and images. The study of personalities is the cornerstone of psychology, and personality recognition can be useful in a wide range of different fields, including social network analysis, recommendation engines, fraud detection, copyright attribution, sentiment analysis/opinion mining, and so on. Conventional personality tests are cumbersome and subjective, whereas social media provides a scalable option. Using NLP and deep learning makes automatic personality inference from texts feasible, although the complexity of language remains a significant limitation. The deep learning (DL) architecture and NLP tasks [4] are discussed in numerous research works. Detailed contextual meaning and the ever-changing character of personality have proven elusive in past studies. Language complexity, such as sarcasm and ambiguity, is often left unaddressed, and it frequently relies on superficial linguistic elements without incorporating psychological theory. These articles employ technical language, which can occasionally be difficult to comprehend. Due to the online documents' quick growth across numerous languages, there is now a large demand for their rating [5]. Recent advancements in social media research have uncovered a multitude of valuable observations on user behavior, especially on sites such as Facebook.

Our work generated several contributions.

- As a preliminary step, researchers offer Big Five model-based personality prediction for microblogging.
- Our method uses Social textual and microblog usage data to improve personality prediction. Given this, we extract textual features for each personality group using short-term and long-term encoding and apply LSTM algorithms to predict personality accurately.
- Experimental studies show our personality prediction technique outperforms traditional and advanced sentiment classifiers. We believe our study is the first to precisely evaluate user personality in social media analytics utilizing personality prediction.

In Section 2, we discussed the history of the related work and presented the methodology and model architecture design in Section 3. In Section 4, we clarified and analyzed the experimental setup and basic outcome. Section 5 outlines the findings and potential areas for further work.

2 Related Work

The study examines the substantial volume of user-generated information on social media platforms such as Facebook, Google, Twitter, and YouTube, emphasizing how this textual data might provide insights into people's personalities. It primarily seeks to find characteristics linked to psychopathy by classifying input text from both psychopaths and non-psychopaths. The research uses content analysis to identify unique personality traits that distinguish these two groups [6]. In recent years, ABSA has garnered significant interest from the scientific world, resulting in an explosion of unique and innovative methods and methodologies originating from the Sentiment Analysis domain [7].

The primary goal of this study is to investigate how utilizing Twitter (dataset) for personality assessments can improve user experience. Research proposes a method for predicting a user's Personality [8] using publicly available information on their Twitter account and the evaluation of the DISC (Dominance, Influence, Steadiness, and Conscientiousness) model. This research aimed to develop a method for identifying psychopaths using a convolutional neural network model [9,10] combined with an LSTM approach.

2.1 The Dark Triad: Beyond the Big Five Personality Traits

The dark triad, often known as antisocial personality, can be accurately predicted by machine learning and data mining methods [11] rather than psychologists. Based on the SD3 model, we surveyed a sample size of Facebook users for this study. People produce a lot of text information, and that text's grammatical and semantic information can reveal much about the author's personality [12]. The next work dataset comprises the personality profiles and information from 180,000 Facebook accounts [13,14]. Approximately 6000 tweets are posted to Twitter per second on average. The average daily time spent on Facebook is 35 min, and around 317,000 status updates are posted on the social media platform [15] every minute. Such social media data [16] is frequently used to forecast user personalities [17]. In [18], age, gender, personality traits, occupation, political inclination, and other user variables can all be predicted using prediction models [19] that have been developed effectively [20,21].

2.2 Personality Prediction Using Deep Learning

Li et al. [22] demonstrated that the HG-PerCon model improves personality prediction by incorporating historical semantic information and psychological knowledge through a transformer-based module and contrastive learning. This method extracts enduring personality characteristics from user postings and utilises a psychological knowledge network to shape user profiles based on language patterns and psychological principles. Lin [23] introduces a framework called DLP-Personality identification, which makes numerous important advances in personality identification. The meta-analysis offers a thorough investigation of the connections between obsessive-compulsive disorder (OCD) and the Big Five personality characteristics by combining data from 23 studies that included 30,138 people. The study establishes a clear link between OCD and neuroticism, with a significant effect size of 0.34. The study's strict adherence to PRISMA principles ensures a high level of methodological rigour, and the explicit recognition of the original data sources underscores the collaborative effort in advancing this specific area of research [24]. The MuPTA corpus integrates speech, video, and text for Big Five trait prediction, supporting richer multimodal personality analysis [25].

Taghvaei et al. [26] proposed a hybrid framework for Big Five personality recognition using FNN and DNN models. A two-stage decision fusion scheme combines multiple FNNs with a DNN to improve accuracy. Simulation results confirm the model's enhanced performance. In [27], the Multimodal Group Gated Score Fusion Network (GSFN) is a groundbreaking advancement in personality trait prediction. It successfully combines video, audio, and text modalities to achieve this integration. In [28], the suggested model utilizes a comprehensive approach to represent features across several modes and scales, including text, audio, and picture data. Studying brain function and disease processes increasingly relies on research into brain networks. However, current brain network design technologies are dependent on empirical users, inconsistent in repeated tests, and time-consuming. This allows for an effective capture of the subtle variations in emotion. The OCEAN-AI framework introduces an efficient EmoFormer model with a novel cross-hemiface attention fusion strategy for assessing the Big Five personality traits [29]. It also includes an open-source implementation and a mid-level emotional feature extractor to enhance multimodal personality inference.

3 Equations and Mathematical Expressions

DL algorithms, namely LSTM networks, efficiently tackle the task of personality prediction from social media content by using their capacity to catch and evaluate intricate sequential patterns. Utilizing specialized psycholinguistic knowledge, we have developed a comprehensive personality lexicon that encompasses LIWC (Linguistic Inquiry and Word Count) terms, as well as specific adjectives. Every term in this lexicon is associated with distinct personality attribute categories, effectively incorporating knowledge obtained via inadequate supervision. Expanding upon the existing personality dictionary and provide a method incorporating personality information into word embeddings. Fig. 1 provides a comprehensive depiction of the step-by-step procedure involved in this technique, which includes the Input Layer, the intermediate Personality Encoding Layer, and the Final Aggregation Layer. The proposed model has three fundamental layers: the Initial Layer, the Personality Encoding Layer, and the Aggregation Layer. Overall, the architecture is carefully crafted to handle and combine both consistent and changing personality characteristics, resulting in a strong and comprehensive depiction of user personalities.

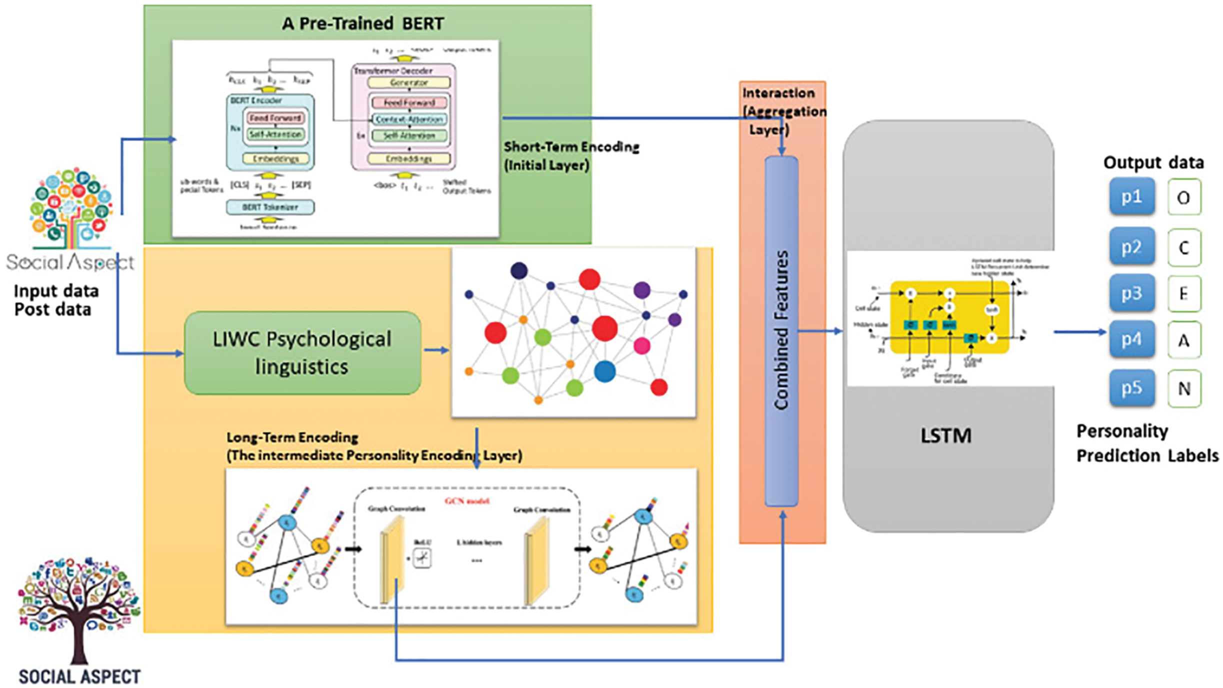


Figure 1: Proposal framework diagram

Linkage Value Computation

In textual personality identification, work might be characterized as a classification issue involving several documents and multiple labels. The formal collection of user-generated posts is represented as P = $\{p_1, p_2, \ldots, p_N\}$. In this collection, each post (p_i) $p_i = \{t_{1i}, t_{2i}, \ldots, t_{Li}\}$ represents the i-th post that contains L tokens. The objective is to make a prediction about the personality qualities, which are denoted by the Equation $Y = \{y_1, y_2, \ldots, y_T\}$, where each y_i refers to a particular personality feature.

The Input Layer is responsible for gathering contributions $P = \{p_1, p_2, \ldots, p_N\}$ and an external LIWC Psychological Knowledge Base. The network is composed of a matrix $M \in \mathbb{R}^{K\times d_g}$ that represents entity embeddings and a matrix $A \in RK \times K$ that captures the associations between entity nodes. In this situation, the symbol K indicates the number of entities, whereas d_g signifies the dimension of the features. To create the psychological graph $\mathcal{G}(M, A)$, it is necessary to acquire the characteristics of units for every user. The entities, represented as $E_u = \{w_1, w_2, \ldots, w_K\}$. These are terms the user specifies and are automatically retrieved from the user's post collections using the LIWC Linguistics extraction. Subsequently, we use GloVe to transform these entities into node characteristics in Eq. (1).

$$\tilde{e}_i = \text{GloVe}(w_i) \in \mathbb{R}^{1\times d_g}, \tag{1}$$

with this architecture, w_i represents the user's i-th entity node, and GloVe(·) corresponds to the GloVe word vector format. The result of this operation is the node feature matrix M, which consists of $M = \{e_1, e_2, \ldots, e_K\}$.

However, it is necessary to connect the obtained entity nodes using the external LIWC Psychological Knowledge Base. More precisely, an edge is formed between two things when they fall under the same LIWC category in Eq. (2).

$$A_{ij} = \begin{cases} 1, C(e_i) \cap C(e_j) \neq \varnothing \\ 0, \text{ otherwise} \end{cases}. \tag{2}$$

The notation $A_{ij} = 1$ represents an edge between node i and j, where as $A_{ij} = 0$ indicates the absence of an edge.

The function $C(\cdot)$ collects LIWC categories for a psychological entity. To determine category membership, each word from the input text is checked against the LIWC dictionary using the following matching Formulas (3) and (4):

$$M(w_i, C_k) = \begin{cases} 1, & \text{if } w_i \in D(C_k) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$M(w_i, C_k)$ indicates if the word w_i matches LIWC category C_k.

$D(C_k)$ is the dictionary of words belonging to the category C_k.

The connection strength between entities is calculated using the frequency of matched words within the same category:

$$F_{C_k} = \frac{\sum_{i=1}^{N} M(w_i, C_k)}{N} \tag{4}$$

F_{C_k} is the frequency of words in the LIWC category C_k.

N, this is the total number of words analyzed.

This module analyses users' long-term personality trait representation by examining their use patterns of psychological entities using the psychological graph. To capture use trends, we utilize a Graph Convolutional Neural Network (GCN) to propagate entity information inside the psychological graph $\mathcal{G}(M, A)$. In each layer of the GCN, the output M^{l-1} from the previous layer is used as the input to learn the representation W^l of the current layer. Eq. (5) presents the learning formula for each GCN layer.

$$\begin{aligned} M^l &= \sigma\left(\hat{A} M^{l-1} W^l\right) \\ \hat{A} &= D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \end{aligned} \tag{5}$$

The symbol $\sigma(\cdot)$ denotes the LeakyReLU activation function, and $W^l \in \mathbb{R}^{d \times d}$ indicates the transformation matrix that may be learned. The degree matrix, D, is defined as having elements. $D_{ij} = \sum_j A_{ij}$. The normalized adjacency matrix is denoted as $\hat{A}$.

The Short-term Personality Encoding module represents the user's short-term personality states using the semantic representations of each of the user's postings. Explicit and implicit aspect extraction in ABSA, highlighting contemporary methodologies such as pre-trained models like BERT, which adeptly recognize nuances of implicit sentiment in informal language, especially within social media environments.

In order to acquire the contextual description of each post, we implement the pre-trained language model BERT. Eq. (6) presents the method for obtaining the contextual representation of each post, indicated as p_i, as follows:

$$h_i = \text{BERT}(p_i) \in \mathbb{R}^{L \times d} \tag{6}$$

Short-term Personality Encoding successfully captures the user's relevant post depictions, resulting in enhanced and informative short-term state representations.

First, we create a bipartite graph called $\tilde{\mathcal{G}}(\tilde{M}, \tilde{A})$. The node matrix $\tilde{M}$, which belongs to the set of real numbers $\tilde{M} \in \mathbb{R}^{(K+N)\times d}$, is formed by combining the long-term personality representations M^l and the short-term representations H^l. The adjacency matrix $\hat{A}$ shows the relationship between user posts and entities. Cap$\tilde{A}_{ij}$ is equal to 1 if entity e_i exists in post p_i, and 0 otherwise. Using the bipartite graph $\tilde{\mathcal{G}}(\tilde{M}, \tilde{A})$ as a basis, we use a different Graph Convolutional Network (GCN) to enhance the user's long-term and short-term aspects in Eq. (7). This process is defined as follows:

$$\tilde{M}^l = \sigma\left(\hat{\tilde{A}}\tilde{M}^{l-1}W^l\right)$$
$$\hat{\tilde{A}} = D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}} \tag{7}$$

Ultimately, we acquire the revised long-term personality representations. $\tilde{M}^l = \left\{\tilde{e}_1^l, \tilde{e}_2^l, \ldots, \tilde{e}_K^l\right\}$ and the revised short-term personality representations $\tilde{H}^l = \left\{\tilde{h}_1^l, \tilde{h}_2^l, \ldots, \tilde{h}_N^l\right\}$. This reciprocal contact enhances their bond, enabling us to accurately record evolving changes while acquiring a rather consistent personality profile.

In order to do this, we first use a simple but efficient method of average pooling inspired by previous works. This method allows us to enhance the long-term and short-term representations by transforming them into two low-dimensional user vectors, $U_l \in \mathbb{R}^{1\times d}$, and $U_s \in \mathbb{R}^{1\times d}$ which are sparse in Eq. (8).

$$U_l = \frac{1}{N}\sum_i \tilde{e}_i^l$$
$$U_s = \frac{1}{K}\sum_i \tilde{h}_i^l \tag{8}$$

We use a controlled integration technique to balance the impact of long-term trait and short-term state representations on personality prediction. This controlled integration technique enables flexible learning from both representations and ultimate personality prediction in Eqs. (9) and (10):

$$U = \alpha U_l + (1-\alpha)\, U_s$$
$$\alpha = \sigma\left(\left[U_l; U_s\right] W + b\right) \tag{9}$$

$W \in \mathbb{R}^{2d\times 1}$ is a trainable weight matrix, whereas $b \in R$ is a bias term. A linear transformation and softmax function forecast each personality feature using the final user representation U.

$$\hat{y} = \mathrm{softmax}\left(UW_u + b_u\right) \tag{10}$$

$\hat{y}$ is the predicted output.

4 Results and Discussion

Two personality models are used most often in personality identification. These are the Big Five. Since the Big Five model is consistent with the psychological theoretical framework that we have developed, our study is centered on it. Furthermore, the widespread availability of Big Five labels makes it especially well-suited for developing personality recognition models specifically customized to social media. This paper utilizes authentic Facebook data to build the personality recognition model. The experimental data is obtained from two primary sources: the corpus, which comprises microblog information and users' Big Five personality ratings, and the findings retrieved using the LIWC tool, followed by Physiological linguistics to create a knowledge graph. The GloVe approach, a variant of word2vec, is used to produce word vectors. These vectors are then included as input for the Long-term Personality Encoding. The outcome of this method is the input matrix used for the LSTM model.

4.1 Evaluation Metrics

The input data for the Short-term Personality Encoding layer consists of information that a pre-trained BERT model has processed. This model has been trained on the entire Facebook corpus and numerous social media postings. As a result, it can successfully capture the subtle aspects of short-term personality characteristics. The experimental data is partitioned into training and test sets, following a 7:3 ratio. The training set has 6893 data points, whereas the test set has 3024. The experiments in this study are performed using Python, with Python Notebook as the development tool. Performance assesses the model's ability to accurately identify all relevant instances within a certain category, as outlined in Eqs. (11)–(14).

$$Overall\ Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{11}$$

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{12}$$

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{13}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

Table 1 provides an overview of the findings obtained by the baseline models and the BERT-LIWC-LSTM algorithm. The results show that our suggested BERT-LIWC-LSTM model beats the other models, as it achieved the highest average F1 score on the MyPersonality dataset. The efficiency of BERT-LIWC-LSTM in personality recognition is shown by the fact that there is a relative improvement of 5% compared to the prior state-of-the-art models.

Table 1: Comparison of previous results and proposed model for personality prediction

Models	Dataset	Accuracy	Precision	Recall	F1
UMLFiT [24]	Essay dataset	58.85	58.88	57.90	57.85
Deep learning Text [24]	Facebook	64.80	63.50	58.50	57.98
ensemble RoBERTa [30]	Youtube	63.2	60.2	80.4	74.1
ensemble BERT [30]	Twitter	66.0	64.0	80.7	73.0
BPT + TF-IGM	Twitter dataset	70.89	69	71	71.5
BILSTM [31]	Essay dataset	71.48	68.81	78.75	73.43
BPT + TF-IGM	Instagram	86.06	85.05	87.2	86.48
BERT-LIWC-LSTM (Our proposed model)	**myPersonality dataset**	**91.67**	**93**	**92**	**91**

4.2 Comparison with Baseline Methods

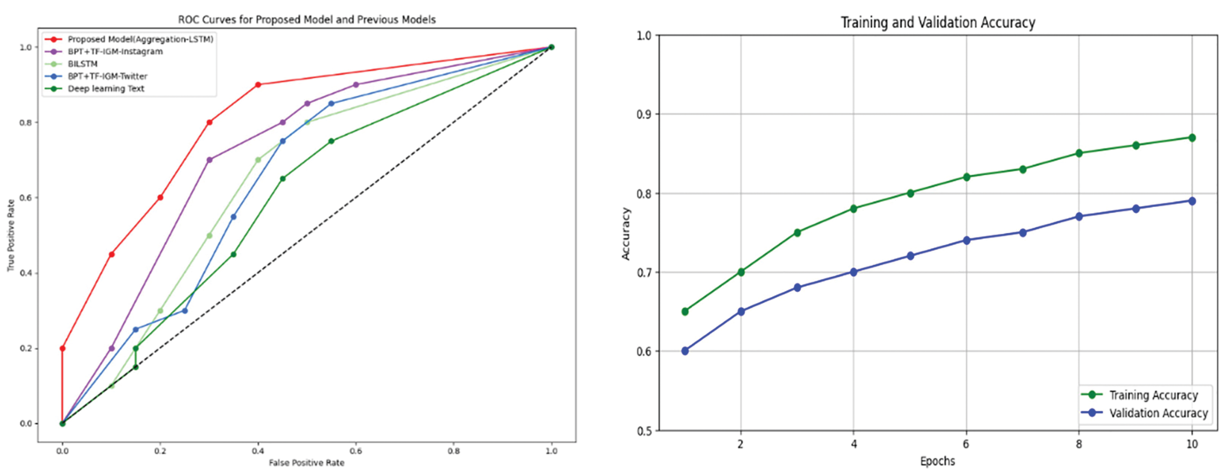
In the personality recognition task, we evaluate BERT-LIWC-LSTM in comparison to the following models, which are considered to be state-of-the-art.

Table 2 provides the performance assessment of the proposed BERT-LIWC-LSTM model. Experimental findings on four datasets—Facebook, Twitter, Essay, and myPersonality—are shown to illustrate the flexibility of the recommended BERT-LIWC-LSTM model. It produced the best accuracy of 91.67% on the myPersonality dataset and outperformed other methods on all platforms consistently. These results verify the model's strength to analyze varied social media content and user activities. It gets 91.67% of the answers right on the myPersonality dataset, which is a big improvement over earlier models. For example, the BPT + TF-IGM model on the Instagram dataset achieves an accuracy of 86.06%, which is more than 5.6% better than the previous result. In the same way, the BERT-LIWC-LSTM model does much better than older models, such as BiLSTM (71.48% accuracy on the Essay dataset) and ensemble BERT (66.0% accuracy on Twitter). This improvement can also be observed in other metrics, such as accuracy (93% for BERT-LIWC-LSTM vs. 85.05% for BPT + TF-IGM), memory (92% vs. 87.2%), and F1 score (91% vs. 86.48%).

Table 2: Comparison of different dataset results with the proposed model for personality prediction

Models	Dataset	Accuracy	Precision	Recall	F1
BERT-LIWC-LSTM (Our proposed model)	Facebook dataset	89.80	87.50	85.0	86.8
	Twitter dataset	88.89	88	87	87.5
	Essay dataset	90.12	89.7	88.1	89.15
	myPersonality dataset	91.67	93	92	91

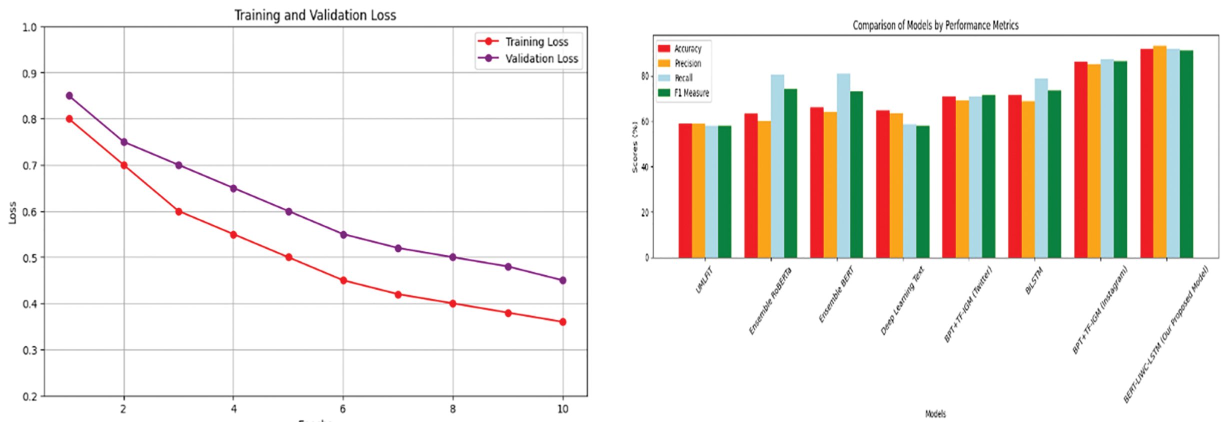
Fig. 2a shows the Receiver Operating Characteristic (ROC) curve for the proposed model beside alternative models, allowing for a direct comparison. The graph illustrates the superior performance of the suggested BERT-LIWC-LSTM model compared to the other models being evaluated. The ROC curve proves that the proposed model performs better or equivalent in differentiating personality characteristics, confirming its accuracy and reliability in personality identification tests.



(a) (b)

Figure 2: (**a**): ROC curve for proposed and comparison models; (**b**): Accuracy performance

Figs. 2b and 3a present the accuracy and loss metrics for the suggested model. The graphs provide valuable insights into the model's temporal performance, illustrating its accuracy in predicting personality characteristics and its proficiency in minimizing training mistakes. Fig. 3b displays a bar chart comparing various models.



(a) (b)

Figure 3: (**a**): Loss performance for proposed model; (**b**) Comparison across various models with other datasets

A batch size of 32 compromises computing efficiency and generalization, whilst a learning rate of 0.0001 facilitates consistent convergence without overshooting. A dropout rate of 0.3 efficiently mitigates overfitting by promoting improved generalization, providing robust performance across varied data inputs, as shown in Table 3.

Table 3: Efficiently parameter overfitting by promoting an improved posed model

Parameter	Best values (Selected)	Alternative 1	Alternative 2	Alternative 3
Batch size	32	16	64	128
Learning rate	0.0001	0.001	0.0005	0.00005
Dropout rate	0.3	0.2	0.5	0.4
Epochs	30	20	50	40
Optimizer	Adam	SGD	RMSProp	AdamW
Hidden layer size	256	128	512	1024
Activation function	LeakyReLU	ReLU	Tanh	ELU
Accuracy	91.67%	88.24%	90.10%	89.75%
Precision	93%	89.56%	91.25%	90.45%
Recall	92%	87.91%	89.12%	88.76%
F1-score	91%	88.73%	90.15%	89.10%

Experimental validation employing the BERT-LIWC-LSTM model on the myPersonality dataset validates the proposed framework.

The worst-case time complexity of the proposed BERT-LIWC-LSTM framework is:

$$O\left(N \cdot L \cdot C + E^2 + T \cdot E^2 \cdot H^2 + N \cdot L^2 \cdot d + N \cdot L \cdot H^2\right)$$

where N: Number of user posts, L: Average length of each post, C: Number of LIWC categories, E: Number of entity nodes, T: Number of GCN layers, H: Hidden units in LSTM and d: Embedding dimension.

Performance dropped significantly with alternate parameter values, proving the configuration's efficiency. These consistent findings across folds and parameter settings prove the model's practicality and dependability.

Table 4 shows that the model achieves a regular accuracy of 77.64% across various personality traits. Openness has the highest accuracy at 94.8%, making it the easiest trait to predict. Conscientiousness has the lowest accuracy at 71.2%, indicating room for improvement. Extraversion, agreeableness, and neuroticism fall between these two, with scores of 76.7%, 71.8%, and 73.7%, respectively. Overall, the model performs well in predicting different traits, with some variation in accuracy across them. Ablation studies were utilized to estimate component contributions. Eliminating the LIWC module resulted in the greatest performance loss, underscoring its importance in absorbing psychological information. The highest performance was achieved by integrating BERT, LIWC, and LSTM, proving that semantic, psychological, and sequential modelling are compatible.

Table 4: Comparison of previous results and proposed model for personality prediction using mypersonality dataset

Models	**Accuracy**					
	Openness	**Conscientiousness**	**Extraversion**	**Agreeableness**	**Neuroticism**	**Average**
CNN + LSTM [26]	76.8	59.68	73.45	59.0	60.9	65.98
SNA, LIWC SPLICE [32]	73.1	63.2	59.0	62.4	74.2	66.74
HMAttn-ECBiL [33]	84.57	68.65	73.94	70.7	62.1	72.0
HPMN GloVe [34]	80.2	81.6	81.1	70.7	81.3	78.9
HPMN BERT [34]	82.3	82.6	82.5	72.7	82.7	80.5
BPT + TF-IGM	92.1	69.2	75.8	70.6	72.9	76.12
BERT-LIWC-LSTM (Our proposed model)	94.8	71.2	76.7	71.8	73.7	77.64

The BERT-LIWC-LSTM model is the most accurate overall, at 77.64%. It does especially well at identifying Openness, at 94.8%. With an accuracy rate of 80.5%, HPMN BERT does better than it, doing especially well in Openness and Neuroticism. HPMN GloVe and BPT + TF-IGM, on the other hand, do well for some traits but not for others consistently. The average accuracy of the SNA, LIWC SPLICE, and CNN + LSTM models is lower, which shows that they cannot handle multiple traits. Overall, BERT-LIWC-LSTM and HPMN BERT are the best at identifying personality traits. This shows how important it is to pick models based on your goals.

The suggested BERT-LIWC-LSTM model has considerable efficacy on modern datasets compared to conventional benchmark datasets. It is evident from its assessment of the myPersonality dataset, a modern and varied dataset that reflects authentic social media interactions. Compared to that, various models assessed on platforms such as Twitter, YouTube, Instagram, and Facebook demonstrate worse performance, with accuracies varying from 58.85% to 86.06%.

Fig. 4 compares the suggested model's Big Five personality characteristics to those of other models. The suggested BERT-LIWC-LSTM model either equals or slightly surpasses the different models (HMAttn-ECBiL and BPT + TF-IGM). The BERT-LIWC-LSTM model demonstrates significant enhancements in qualities like Openness and Conscientiousness, indicating its efficacy in precisely capturing important aspects of personality. The findings confirm that the suggested model provides a distinct advantage in identifying personality traits, thereby confirming its method of combining enduring personality characteristics with immediate user actions.

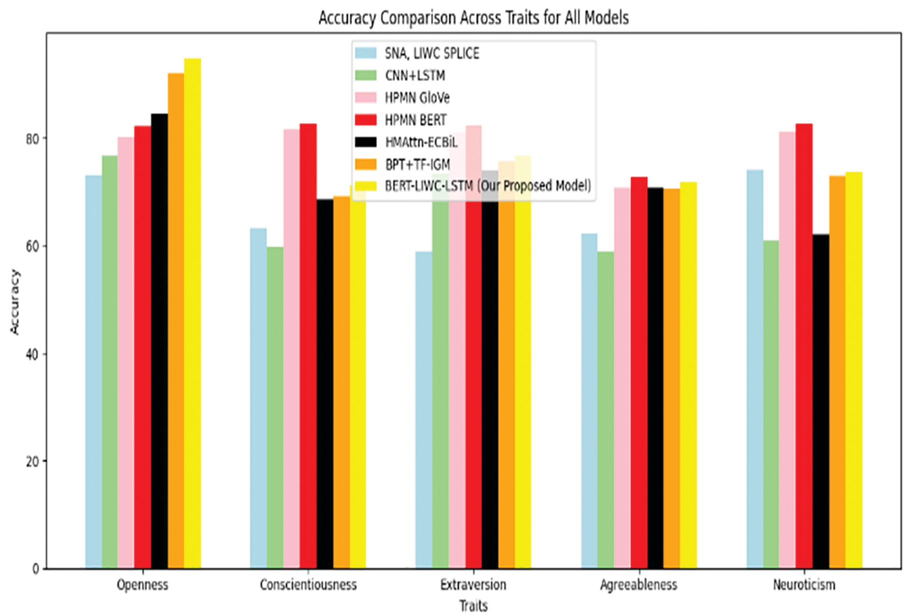


Figure 4: Comparison across various models with mypersonality dataset

5 Conclusion

This research presents BERT-LIWC-LSTM, a model that successfully combines short-term behavioural cues with long-term personality traits to enhance the accuracy of personality prediction. The model took advantage of the stability and dynamic nature of personality through the incorporation of contextual embeddings (BERT), psychological dictionaries (LIWC), and sequential learning (LSTM). It effectively identified user-specific language patterns and dependencies between phrases. Experimental testing on the MyPersonality dataset showed that the model attained a 3.41% improvement in accuracy compared to state-of-the-art baselines, verifying its efficacy in personality recognition tasks. The model also helped to advance computational linguistics by enhancing lexical-based psycholinguistic hypotheses.

In future research, we intend to generalise the use of the BERT-LIWC-LSTM framework to other personality frameworks other than the Big Five, like the Myers-Briggs Type Indicator (MBTI), MuPTA and ChaLearn First Impressions V2 in order to test its robustness and generalizability across different theoretical frameworks. Future studies may also investigate incorporating multimodal input data, e.g., images and video, along with text, in order to create a richer and more detailed understanding of user personalities within real-world online settings.

Acknowledgement: The authors gratefully acknowledge the anonymous reviewers and editor for their insightful comments and valuable suggestions.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: The authors confirm their contribution to the paper as follows: Study conception and design: Nareshkumar Raveendhran; data collection: Nareshkumar Raveendhran; analysis and interpretation of results: Nareshkumar Raveendhran and Nimala Krishnan; draft manuscript preparation: Nareshkumar Raveendhran and Nimala Krishnan. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data and materials used in this research are available upon reasonable request to the corresponding author.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. El-Demerdash K, El-Khoribi RA, Ismail Shoman MA, Abdou S. Deep learning based fusion strategies for personality prediction. Egypt Inform J. 2022;23(1):47–53. doi:10.1016/j.eij.2021.05.004.
2. Raveendhran N, Krishnan N. A novel hybrid SMOTE oversampling approach for balancing class distribution on social media text. Bull Electr Eng Inform. 2025;14(1):638–46. doi:10.11591/eei.v14i1.8380.
3. Hashemi Motlagh SM, Rezvani MH, Khounsiavash M. AI methods for personality traits recognition: a systematic review. Neurocomputing. 2025;640:130301. doi:10.1016/j.neucom.2025.130301.
4. Xue D, Wu L, Hong Z, Guo S, Gao L, Wu Z, et al. Deep learning-based personality recognition from text posts of online social networks. Appl Intell. 2018;48(11):4232–46. doi:10.1007/s10489-018-1212-4.
5. Feher A, Vernon PA. Looking beyond the big five: a selective review of alternatives to the big five model of personality. Personal Individ Differ. 2021;169(1):110002. doi:10.1016/j.paid.2020.110002.
6. Ryumina E, Markitantov M, Ryumin D, Karpov A. OCEAN-AI framework with EmoFormer cross-hemiface attention approach for personality traits assessment. Expert Syst Appl. 2024;239(2):122441. doi:10.1016/j.eswa.2023.122441.
7. Souri A, Hosseinpour S, Rahmani AM. Personality classification based on profiles of social networks' users and the five-factor model of personality. Hum Centric Comput Inf Sci. 2018;8(1):24. doi:10.1186/s13673-018-0147-4.
8. Alshouha B, Serrano-Guerrero J, Chiclana F, Romero FP, Olivas JA. Personality trait detection via transfer learning. Comput Mater Contin. 2024;78(2):1933–56. doi:10.32604/cmc.2023.046711.
9. Liu D, Campbell WK. The big five personality traits, big two metatraits and social media: a meta-analysis. J Res Personal. 2017;70(1):229–40. doi:10.1016/j.jrp.2017.08.004.
10. Ren Z, Shen Q, Diao X, Xu H. A sentiment-aware deep learning approach for personality detection from text. Inf Process Manag. 2021;58(3):102532. doi:10.1016/j.ipm.2021.102532.
11. Buettner R. Predicting user behavior in electronic markets based on personality-mining in large online social networks. Electron Mark. 2017;27(3):247–65. doi:10.1007/s12525-016-0228-z.
12. Landers RN, Lounsbury JW. An investigation of big five and narrow personality traits in relation to internet usage. Comput Hum Behav. 2006;22(2):283–93. doi:10.1016/j.chb.2004.06.001.
13. Dandannavar PS, Mangalwede SR, Kulkarni PM. Social media text—a source for personality prediction. In: 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS); 2018 Dec 21–22; Belgaum, India. p. 62–5. doi:10.1109/CTEMS.2018.8769304.
14. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A. 2013;110(15):5802–5. doi:10.1073/pnas.1218772110.
15. Hernández Y, Martínez A, Estrada H, Ortiz J, Acevedo C. Machine learning approach for personality recognition in Spanish texts. Appl Sci. 2022;12(6):2985. doi:10.3390/app12062985.
16. Ramezani M, Feizi-Derakhshi MR, Balafar MA, Asgari-Chenaghlu M, Feizi-Derakhshi AR, Nikzad-Khasmakhi N, et al. Automatic personality prediction: an enhanced method using ensemble modeling. Neural Comput Appl. 2022;34(21):18369–89. doi:10.1007/s00521-022-07444-6.

17. Ellouze M, Hadrich Belguith L. A hybrid approach for the detection and monitoring of people having personality disorders on social networks. Soc Netw Anal Min. 2022;12(1):67. doi:10.1007/s13278-022-00884-x.
18. Kamalesh MD, Bharathi B. Personality prediction model for social media using machine learning technique. Comput Electr Eng. 2022;100(3/2021):107852. doi:10.1016/j.compeleceng.2022.107852.
19. Lin CK, Lee YY, Yu CH, Chen HH. Exploring ensemble of models in taxonomy-based cross-domain sentiment classification, information and knowledge management. In: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management; 2014 Nov 3–7; Shanghai, China.
20. Cao X, Yu J, Zhuang Y. Injecting user identity into pretrained language models for document-level sentiment classification. IEEE Access. 2022;10:30157–67. doi:10.1109/access.2022.3158975.
21. Shum KM, Ptaszynski M, Masui F. Big five personality trait prediction based on user comments. Information. 2025;16(5):418. doi:10.3390/info16050418.
22. Li M, Zhu Y, Li S, Wu B. HG-PerCon: cross-view contrastive learning for personality prediction. Neural Netw. 2024;169(2):542–54. doi:10.1016/j.neunet.2023.10.042.
23. Lin H. DLP-personality detection: a text-based personality detection framework with psycholinguistic features and pre-trained features. Multimed Tools Appl. 2024;83(13):37275–94. doi:10.1007/s11042-023-17015-z.
24. Mukta MSH, Ahmad J, Zaman A, Islam S. Attention and meta-heuristic based general self-efficacy prediction model from multimodal social media dataset. IEEE Access. 2024;12(1):36853–73. doi:10.1109/access.2024.3373558.
25. Ren G, Diao L, Guo F, Hong T. A co-attention based multi-modal fusion network for review helpfulness prediction. Inf Process Manag. 2024;61(1):103573. doi:10.1016/j.ipm.2023.103573.
26. Taghvaei N, Masoumi B, Keyvanpour MR. A hybrid framework for personality prediction based on fuzzy neural networks and deep neural networks. J AI Data Min. 2021 Jul 1;9(3):283–94.
27. Ryumina E, Markitantov M, Ryumin D, Karpov A. Gated Siamese Fusion Network based on multimodal deep and hand-crafted features for personality traits assessment. Pattern Recognit Lett. Sep 2024;185:45–51. doi:10.1016/j.patrec.2024.07.004.
28. Wang X, Lyu J, Kim BG, Parameshachari BD, Li K, Li Q. Exploring multimodal multiscale features for sentiment analysis using fuzzy-deep neural network learning. IEEE Trans Fuzzy Syst. 2025;35(1):28–42. doi:10.1109/tfuzz.2024.3419140.
29. Liu S, Luo Z, Fu W. Fcdnet: fuzzy cognition-based dynamic fusion network for multimodal sentiment analysis. IEEE Trans Fuzzy Syst. 2025;33(1):3–14. doi:10.1109/tfuzz.2024.3407739.
30. Tsani EF, Suhartono D. Personality identification from social media using ensemble bert and roberta. Informatica. 2023 Nov;47(4):537–44. doi:10.31449/inf.v47i4.4771.
31. Ramezani M, Feizi-Derakhshi MR, Balafar MA. Knowledge graph-enabled text-based automatic personality prediction. Comput Intell Neurosci. 2022;2022:3732351. doi:10.1155/2022/3732351.
32. Sonawale O. Hybrid deep learning framework for personality prediction in E-recruitment. Int J Sci Res Eng Manag. 2025;9(6):1–9. doi:10.55041/ijsrem49831.
33. Wang X, Sui Y, Zheng K, Shi Y, Cao S. Personality classification of social users based on feature fusion. Sensors. 2021;21(20):6758. doi:10.3390/s21206758.
34. Zhu Y, Hu L, Ning N, Zhang W, Wu B. A lexical psycholinguistic knowledge-guided graph neural network for interpretable personality detection. Knowl Based Syst. 2022;249(1):108952. doi:10.1016/j.knosys.2022.108952.