



ARTICLE

## Personality Trait Detection via Transfer Learning

Bashar Alshouha<sup>1</sup>, Jesus Serrano-Guerrero<sup>1,\*</sup>, Francisco Chiclana<sup>2</sup>, Francisco P. Romero<sup>1</sup> and Jose A. Olivas<sup>1</sup>

<sup>1</sup>Department of Information Technologies and Systems, University of Castilla-La Mancha, Ciudad Real, 13071, Spain

<sup>2</sup>Institute of Artificial Intelligence, School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester, LE1 9BH, UK

\*Corresponding Author: Jesus Serrano-Guerrero. Email: [jesus.serrano@uclm.es](mailto:jesus.serrano@uclm.es)

Received: 12 October 2023 Accepted: 18 December 2023 Published: 27 February 2024

### ABSTRACT

Personality recognition plays a pivotal role when developing user-centric solutions such as recommender systems or decision support systems across various domains, including education, e-commerce, or human resources. Traditional machine learning techniques have been broadly employed for personality trait identification; nevertheless, the development of new technologies based on deep learning has led to new opportunities to improve their performance. This study focuses on the capabilities of pre-trained language models such as BERT, RoBERTa, ALBERT, ELECTRA, ERNIE, or XLNet, to deal with the task of personality recognition. These models are able to capture structural features from textual content and comprehend a multitude of language facets and complex features such as hierarchical relationships or long-term dependencies. This makes them suitable to classify multi-label personality traits from reviews while mitigating computational costs. The focus of this approach centers on developing an architecture based on different layers able to capture the semantic context and structural features from texts. Moreover, it is able to fine-tune the previous models using the MyPersonality dataset, which comprises 9,917 status updates contributed by 250 Facebook users. These status updates are categorized according to the well-known Big Five personality model, setting the stage for a comprehensive exploration of personality traits. To test the proposal, a set of experiments have been performed using different metrics such as the exact match ratio, hamming loss, zero-one-loss, precision, recall, F1-score, and weighted averages. The results reveal ERNIE is the top-performing model, achieving an exact match ratio of 72.32%, an accuracy rate of 87.17%, and 84.41% of F1-score. The findings demonstrate that the tested models substantially outperform other state-of-the-art studies, enhancing the accuracy by at least 3% and confirming them as powerful tools for personality recognition. These findings represent substantial advancements in personality recognition, making them appropriate for the development of user-centric applications.

### KEYWORDS

Personality trait detection; pre-trained language model; big five model; transfer learning



## 1 Introduction

Personality trait detection is a research field whose main goal is to identify and analyze an individual's inherent psychological characteristics. These traits are typically categorized following personality models such as the Big Five [1] or the Myers-Briggs Type Indicator (MBTI) [2]. The analysis of personality traits provides immense value across diverse domains such as psychology, human resources, digital marketing, and personalization [3]. Moreover, it enables various applications such as user profiling [4], targeted content recommendations [5], and mental health assessments [6], among others.

The rapid development of Internet technologies has opened new opportunities for detecting personalities beyond the traditional costly mechanisms. Social networking sites have impelled individuals to share their opinions, ideas, and emotions with others, which reflects their behavior patterns, attitudes, and personality. There is a robust link between people's temperament and behavior and their comments posted on social networks [7]. For this reason, many researchers are currently interested in developing automatic personality classification systems from social networks.

Most of the current studies on personality trait prediction are based on classical machine learning techniques as well as new recent deep learning architectures, demonstrating their effectiveness in this domain. Most of these techniques are based on modeling lexical and linguistic features from text, and even adding extra features from social networks as it is described in detail in the literature section [8–12]. Nonetheless, the accuracy of the achieved results is still improvable. One of the reasons is the limited availability of datasets for training the state-of-the-art techniques; furthermore, the used datasets are usually very small. These techniques also present some limitations such as the inability to capture nuanced and context-aware semantics of text going beyond traditional word embeddings such as Word2Vec and GloVe. This is essential for addressing challenges such as polysemy, where words have multiple meanings in different contexts, a factor that can introduce ambiguity in personality trait detection. Another limitation is the inability to deal with out-of-vocabulary terms, that is, terms not used when training the models. For that reason, to overcome these limitations, it is proposed the use of pre-trained language models (PLMs) which have undergone extensive pre-training on vast amounts of textual data employing different types of tasks, which empowers them to outperform traditional methods. This intensive pre-training provides PLMs with the ability to understand language intricacies at a deep level, which can be particularly beneficial when fine-tuning personality detection models.

Therefore, this study presents a new framework able to integrate and adapt different PLMs to detect multiple personality traits. Additionally, this research tests the proposal by introducing a robust set of evaluation metrics (exact match ratio, hamming loss, and the weighted average of precision, recall, and F1-score) to measure the efficiency and effectiveness of personality trait detection based on PLMs. This fact allows offering a more comprehensive and insightful evaluation of its results.

To sum up, the primary contributions of this study are:

- To study the capabilities of transfer learning applied to personality detection with the aim of improving the effectiveness of the current methods, especially, when the available datasets are small.
- To propose a general framework for fine-tuning different PLMs to detect multiple personality traits accurately.
- To study the effectiveness of the proposal against other recent state-of-the-art studies based on traditional machine learning techniques and deep learning architectures.
- To perform a deep analysis of the obtained results comparing the effectiveness of the different PLMs.

The remainder of the paper is organized as follows: [Section 2](#) describes the most relevant studies on personality recognition and transfer learning whereas [Section 3](#) explains in detail the proposed framework integrating the different PLMs. [Section 4](#) describes the experimental setup and [Section 5](#) presents the results of the performed experiments analyzing the achieved results. Finally, the conclusions reached are mentioned in the last Section.

## 2 Literature Review

### 2.1 *Transfer Learning in Natural Language Processing*

Transfer learning has emerged as a powerful technique for machine learning [13], particularly in the field of natural language processing (NLP), where it has demonstrated superior performance in text classification tasks regarding previous traditional techniques [14]. Unlike traditional machine learning approaches, transfer learning allows the transmission of knowledge from a source domain to a target domain, even when the data in these domains may differ [15].

One of the key benefits of transfer learning in NLP is its ability to handle situations where the target domain has insufficient training data. By leveraging information from a different but related source domain, transfer learning enables predictions to be made on unseen instances in the target domain. This capability is especially valuable when there is limited data available for a specific problem, but there are many for a related problem [16]. Transfer learning serves as a technique for improving the performance of a learner, such as a classifier, by transferring knowledge between two domains [17].

In NLP, transfer learning encompasses various types of learning such as domain adaptation [18], cross-lingual learning [19], or multi-task learning [20]. The adaptation of pre-trained models to downstream tasks exemplifies a sequential transfer learning task, where tasks are learned sequentially, and the target task benefits from labeled data. Through this sequential transfer learning process, the knowledge acquired from the source task or domain is effectively adapted and applied to the target task or domain [21].

PLMs have transformed the landscape of NLP, particularly in the realm of text classification. Text classification relies on the accurate interpretation and categorization of textual data. In the context of text classification, PLMs provide several advantages. First, through pre-training, PLMs acquire universal language representations, enabling them to capture nuances and semantics across diverse contexts. This comprehensive understanding greatly enhances their performance in downstream tasks. Second, pre-training serves as a strong initialization for the models, expediting convergence and improving generalization. Lastly, pre-training acts as a form of regularization, mitigating overfitting risks in scenarios with limited training data.

One of the most well-known examples of the PLMs is BERT. BERT has garnered widespread adoption and utilization across diverse NLP tasks such as sentiment analysis [22], question answering [23] or text summarization [24]. Nonetheless, BERT's large model size presents challenges when it comes to training the model from scratch. To address this concern, researchers have devised optimized variants and compression techniques such as DistilBERT [25], ALBERT [26], or RoBERTa [27]. These models offer efficient alternatives, efficiently reducing the model size and computational demands while preserving or even enhancing its performance.

Furthermore, self-supervised learning helps improve pre-trained models in different tasks for NLP as it is applied in [28]. These methods are also becoming increasingly popular in computer vision. One example of is the Masked Autoencoders (MAE) framework [29]. MAE was initially developed for image processing but has since evolved into a versatile tool that can handle a variety of data types

such as video, audio, or temporal predictions [30,31]. SemMAE [32] is a significant advancement in this field, introducing a new concept called Semantic-Guided Masking.

## 2.2 Personality Detection

Artificial intelligence has gained visibility across various domains due to its capability to address complex problems through advanced data analysis and predictive capabilities. Some recent applications showcasing its significance can be, for instance, automated language translation [33], image recognition [34], groundwater storage change forecasting [35], climate change forecasting [36], or environmental change analysis [37], among many others.

Personality prediction is another area that is gaining momentum. Numerous studies have been recently conducted, especially, using traditional machine learning algorithms and applying different feature selection techniques to construct feature vectors [38–41]. Many studies indicate that there is a high correlation between textual features and personality traits. For instance, Tadesse et al. [11] studied the relationship between linguistic features and personality behavior using different classifiers. According to the results, the XGBoost classifier using social network analysis (SNA) features outperformed the other models. In another study, Amirhosseini et al. [42] developed a new method also based on XGBoost to predict personality according to the MBTI types. Azucar et al. [43] investigated the correlation between the Big Five personality traits and negative textual information. Han et al. [44] proposed an interpretable personality detection model based on a personality prior-knowledge lexicon, that studied the relationship between term semantic categories and personality traits. Kumar et al. [45] extracted features using the global vectors for word representation (GloVe) model and TF-IDF to feed an ensemble model integrating SVM and extreme gradient boosting (XGBoost) as classifiers. In this case, the absence of demographic data, such as age and gender, limits its ability to provide nuanced and context-aware personality predictions.

Apart from traditional machine learning techniques, the development of new deep learning architectures has allowed researchers to improve the results of the task of personality trait categorization. For instance, Majumder et al. [46] utilized a convolutional neural network (CNN) to extract semantic features integrated with document-level stylistic features to classify personality traits. Sun et al. [47] proposed the concept of latent sentence to present the abstract feature integration, which was integrated into a bidirectional long short-term memory (Bi-LSTM)-CNN architecture. In [48], a hybrid deep learning model integrating CNN with the LSTM was developed to identify eight personality traits. Tandra et al. [9] tested different machine learning and deep learning algorithms to classify personality traits based on the Big Five personality model. Furthermore, they utilized different feature selection and resampling techniques to improve their performance.

Rahman et al. [49] investigated the impacts of various activation functions in CNNs. Xue et al. [50] proposed a personality classification model based on an attention mechanism applied to a recurrent conventional neural network (RCNN) combined with a CNN, which is eligible for learning intricate and hidden semantic features from textual content. Other research studied the correlation between personality traits and emotion [12].

Zhao et al. [51] introduced an innovative method using an attention-based LSTM model, leveraging users' theme preferences and text sentiment features as attention information to improve the accuracy of the results. Moreover, Wang et al. [52] presented a hierarchical hybrid model, HMAtn-ECBiL, which integrates self-attention, CNN, and BiLSTM modules to capture word-level and post-level contributions to personality information and dependencies between scattered posts.

In other recent studies, researchers have harnessed the power of machine learning and deep learning models to predict personality traits from a variety of data sources. Ren et al. [53] explored a novel multi-label personality detection model, incorporating emotional and semantic features from user-generated texts. Anari et al. [54] proposed a lightweight deep convolutional neural network for personality trait prediction based on Persian handwriting, demonstrating reasonable results for various traits. Furthermore, William et al. [55] provided insights into the connection between personality traits and entrepreneurial success, emphasizing the significant impact of the different personalities detected. Kamalesh et al. [56] delved into predicting personality traits based on social media interactions, employing a Binary-Partitioning Transformer with the Term Frequency and Inverse Gravity Moment approach, which enhances trait prediction. Ramezani et al. [57] proposed five new methods, including deep learning-based approaches, to enhance the accuracy through ensemble modeling and hierarchical attention networks. Lastly, Suman et al. [58] combined data from various modalities such as text, facial features, and audio, revealing the potential of deep learning-based personality prediction across multiple dimensions.

Focusing on different applications of transfer learning, El-Demerdash et al. [59] proposed the use of the Universal Language Model Fine-tuning (ULMFiT) to detect personality traits. The results indicate that the use of ULMFiT improves accuracy by about 1% in comparison with the most recent methods. Aslan et al. [60] also addressed the problem of categorizing personality traits from videos by leveraging embeddings from language models to construct a multimodal framework to classify personality traits. Mehta et al. created a model using contextualized embeddings and psycholinguistic features [61]. The results demonstrate that the language model features from BERT [62] combined with an MLP outperform traditional psycholinguistics features. In another study, El-Demerdash et al. [10] proposed a new deep learning model that takes advantage of PLMs such as Elmo, ULMFiT, and BERT, using both classifiers and data-level fusion. Among the limitations found, the incorporation of emotion and slang-based features, as well as the recognition of personality traits in multimedia content can be mentioned.

### 3 Proposal

The goal of this study is to detect multiple personality traits over texts taking advantage of diverse transfer learning technologies. For that reason, the proposed architecture has been designed to integrate different PLMs which will have been fine-tuned to recognize the personality traits following the Big Five model. The effectiveness of the selected models has been demonstrated in NLP tasks:

- **BERT (Bidirectional Encoder Representations from Transformers)** is a language representation model that was first presented in [62]. It is a bidirectional transformer model pre-trained on a huge unannotated dataset, which can be adapted for a large variety of downstream NLP tasks. BERT uses the masked language model (MLM) and next sentence prediction (NSP) to achieve language comprehension. The “bert-base-uncased”<sup>1</sup> version has been used in the experimental section, which makes no distinction between lowercase and uppercase terms.
- **RoBERTa (Robustly Optimized BERT-Pretraining Approach)** is an enhanced version of the BERT model. It is derived from BERT’s MLM strategy adjusting its critical hyperparameters. It eliminates the NSP task, provides dynamic token masking during the training phase, and can enhance the performance on many downstream tasks because it is trained with additional batch size and learning rate [27]. For the experiments, the “roberta-base”<sup>2</sup> version has been utilized.

---

<sup>1</sup><https://huggingface.co/bert-base-uncased>

<sup>2</sup><https://huggingface.co/roberta-base>

- **ALBERT (A Lite BERT)** is a transformer model that has been self-supervised and trained on a significant corpus of English data. It is a lite BERT that minimizes BERT's parameters integrating two-parameter reduction techniques, which enables the model to share parameters between layers, operates as a stabilizing factor when training, and assists in generalization. It uses factorized embedding parameterization and cross-layer parameter [26]. In the experimental section, the “albert-base-v1”<sup>3</sup> uncased version has been used.
- **XLNet** is a PLM based on transformers. Instead of MLM and NSP, it employs an autoregressive approach to learning bidirectional contexts by predicting tokens in a given sequence in a random order, which helps the model capture a more accurate relation between tokens in a particular sequence [63]. The “xlnet-base-cased”<sup>4</sup> version has been selected for the experiments.
- **ELECTRA** is another pre-training method whose fundamental goal is to detect replaced tokens in input sentences using two transformer models: the generator and the discriminator. The generator function is used to create tokens to replace some of the original tokens, whereas the discriminator attempts to determine which tokens in the sequence were replaced by the generator [64]. The “google/electra-small-discriminator”<sup>5</sup> version has been assessed in the experimental setup.
- **ERNIE (Enhanced Representation through kNowledge IntEgration)** acquires language representations utilizing knowledge masking techniques [65]. It is a continuous pre-training system that incorporates lexical, syntactic, and semantic information through massive data via multi-task learning, consequently boosting its existing knowledge base. In the experimental setup, the “nghuyong/ernie-2.0-en”<sup>6</sup> version has been used.

### 3.1 Architecture of the Proposal

The workflow diagram depicted in Fig. 1 consists of several crucial functional blocks, each one playing a unique role in the architecture of PLMs. First, the “Masking Techniques for Input” block preprocesses the input text data by masking specific tokens within the sequence, creating a partially observed context for training token prediction. Next, the “Embedding Layer” translates tokens into high-dimensional vectors for semantic understanding. The heart of the model lies in the “Transformer Layers (1–12)”, where the multi-head attention and feedforward networks enable context representation. The “Pooler Layer (CLS)” aggregates sequence information into a single [CLS] vector for downstream tasks. A “Dense Layer” enhances feature extraction, while “Normalization” and “Dropout” layers ensure stable training and regularization. Finally, the “Tanh Function” captures nonlinear relationships.

The proposed framework includes two primary phases whose goal is to exploit and integrate the advantages of all previous models to detect personality traits following the Big Five model:

#### *Phase 1. Data preprocessing*

First, a tokenizer divides the text into different tokens following a limited set of rules. Then, the tokens are transformed into numbers, which are fed into the model to construct tensors. One of the limitations is that the input lengths for PLMs are often constrained; for that reason, padding and truncation strategies have been utilized to address this problem, inserting special padding tokens into the sequences having fewer tokens or truncating the longest sequences. This process ensures the tensors have fixed length vectors representing the sentences. In the experiments, all PLMs have been utilized

<sup>3</sup><https://huggingface.co/albert-base-v1>

<sup>4</sup><https://huggingface.co/xlnet-base-cased>

<sup>5</sup><https://huggingface.co/google/electra-small-discriminator>

<sup>6</sup><https://huggingface.co/nghuyong/ernie-2.0-en>

with their associated pre-trained tokenizer, which assures that the text is divided in the same manner as a pre-training corpus and that the same vocabulary is used when pre-training.

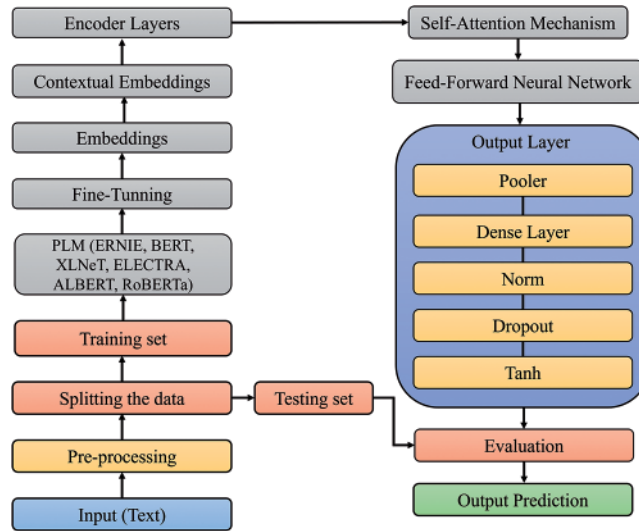


Figure 1: Workflow diagram for PLMs

**Phase 2. Fine-tuning framework**

Transfer learning consists in adapting the knowledge gained from a basic task to a target task, even for new domains. In this research, the PLMs have been adapted to classify multi-label personality traits from reviews. PLMs significantly reduce computation costs, capture structural features from texts, and apprehend plentiful facets of language relevant to downstream tasks, such as hierarchical relations, long-term dependencies, and sentiments. Therefore, the previous knowledge from the PLMs has been fine-tuned for the task of personality trait categorization. To do so, the last layer of all PLMs has been selected as an adaptation layer due to its simplicity and effectiveness.

In this approach, a deviation from the traditional use of the [CLS] token solely for fine-tuning and probability distribution prediction was implemented. Instead, it was integrated with additional classification head layers, namely a dense layer, normalization layer, dropout layer, and *tanh* activation function. This customized classification head facilitated accurate predictions of the five personality traits. Through this adaptation, the sequence classification capabilities inherent in all PLMs were utilized, thus tailoring the model to effectively capture the necessary information for personality trait prediction. In this manner, the linear layer weights from various PLMs have been reconfigured following the personality trait classification task. Fig. 2 illustrates the necessary phases to identify different personality traits by fine-tuning the mentioned PLMs.

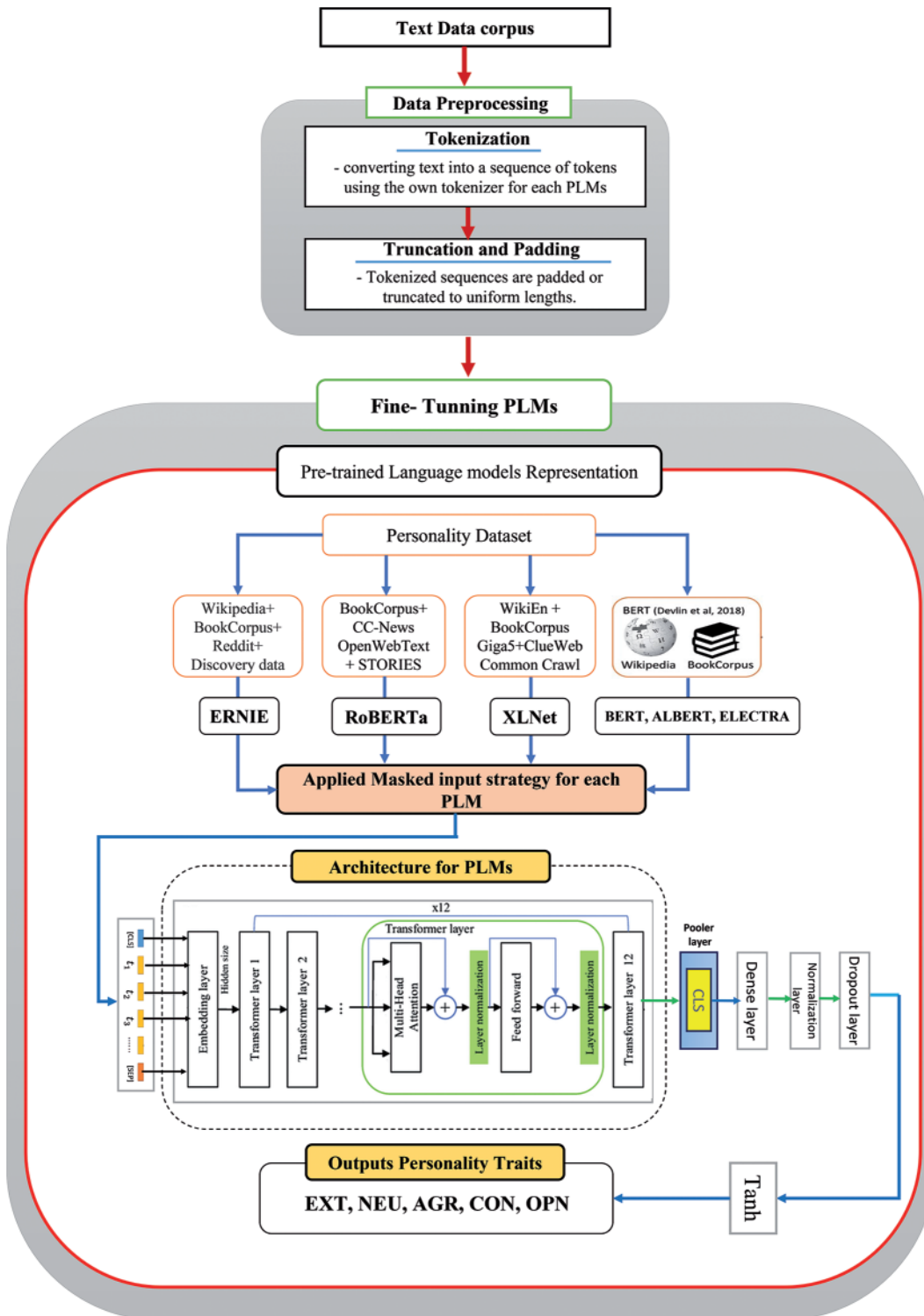


Figure 2: Framework followed to fine-tune the PLMs



### 3.2 Detailed Implementation of the Architecture

To adapt PLMs such as BERT, RoBERTa, ALBERT, ELECTRA, ERNIE, and XLNet for personality trait classification, the following steps were performed. These steps highlight the main differences and contributions made in comparison to other architectures [66–70].

First, the PLM and its specific tokenizer were loaded along with the dataset to ensure compatibility between the selected model architecture and the tokenized dataset. Then, the only used dataset, MyPersonality, was preprocessed, involving tasks such as tokenization, truncation, and padding. Tokenization was performed using the respective tokenizer, breaking the input texts into individual tokens and incorporating special tokens such as [CLS] and [SEP]. In those cases where the input length exceeded the maximum token limit of the model, truncation was applied to trim the text, and padding was added to match the maximum length.

In the tokenization step, a percentage of tokens in the input text were randomly selected and replaced with a special [MASK] token, implementing the masked input strategy. This strategy enables the model to learn how to predict the masked tokens during training and capture bidirectional contextual information. This masked input strategy is an essential part of the pre-training process for these adapted models.

After tokenizing, the tokens were converted into numerical representations using the tokenizer's vocabulary, which assigns a unique numerical ID to each token. The tokenizer also handled the conversion of the special tokens into their corresponding IDs. Finally, the tokenizer returned tensors representing the tokenized input texts that would be fed into the model for further processing.

Next, the PLM was instantiated, maintaining the base layers responsible for capturing linguistic features and the transformer layers. The weights in the base layers, including the embedding layer and transformer layers, were kept retaining the already learned representations and ensuring they were not modified during the fine-tuning process.

The architecture of the classification head layers was then customized to match the requirements of the personality classification task. This customization involved the addition of several layers, namely a dense layer, a normalization layer, a dropout layer, and a *tanh* activation function. The dense layer introduces non-linearity, the normalization layer enhances stability, the dropout layer prevents overfitting, and the *tanh* activation function captures nuanced characteristics. These layers collectively contribute to improving the model's performance and adaptability. The inclusion of these layers enhances the model's capability to capture intricate patterns and relationships relevant to personality trait prediction.

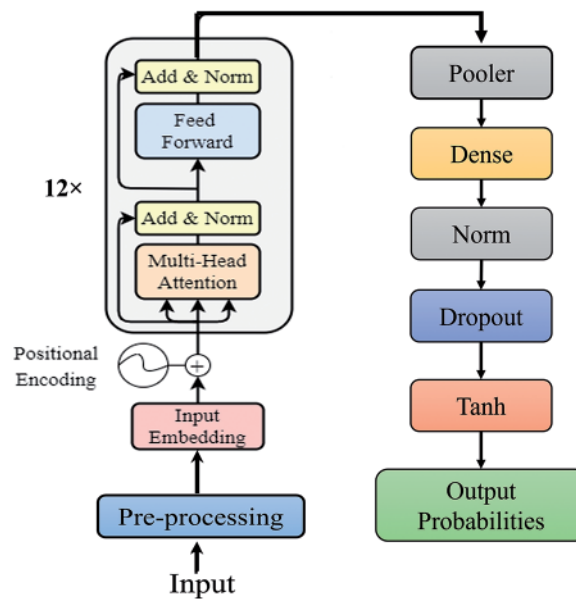
In addition to the customization of the classification head layers, an output layer was added to the model to classify the input text into the Big Five personality traits: neuroticism (NEU), extraversion (EXT), conscientiousness (CON), agreeableness (AGR), and openness (OPN). This specialized output layer is responsible for making predictions.

The adapted model was subsequently trained on the preprocessed training set, with only the weights in the classification head layers updated. This fine-tuning process allowed the model to learn task-specific features while leveraging the pre-trained base layers' language understanding capabilities. The Trainer API<sup>7</sup>, with its various training options and features, including logging, gradient accumulation, and mixed precision, was utilized to simplify the training process and fine-tune hyperparameters such as the learning rate, batch size, and regularization techniques to optimize the model performance.

---

<sup>7</sup>[https://huggingface.co/docs/transformers/main\\_classes/trainer](https://huggingface.co/docs/transformers/main_classes/trainer)

To sum up, most of the conventional model architectures in the field of personality trait prediction are employed for inference and transfer learning; nonetheless, this approach involves fine-tuning the model and adapting it specifically for the classification of multi-label personality traits from the reviews. To do so, the goal of this architecture was to fine-tune the model classification head, which includes a dense layer, normalization layer, dropout layer, and *tanh* activation function. These additional head layers were added to improve the model's ability to capture complex patterns in personality trait prediction. Customizing the architecture with these extra head layers, along with the utilization of specialized tokens such as the [CLS] token, the model was optimized to effectively capture and leverage essential information for precise personality trait prediction. The network topology diagram (see Fig. 3) provides an insightful visualization of the architecture shared by the PLMs used in this study.



**Figure 3:** Network topology diagram of the PLMs

#### 4 Experimental Setup

To assess the proposed framework, the MyPersonality<sup>8</sup> dataset has been selected. This dataset comprises 9,917 textual statuses authored by 250 users, categorized according to the Big Five personality model [71]. The distribution of the Big Five personality traits, i.e., neuroticism (NEU), extraversion (EXT), conscientiousness (CON), agreeableness (AGR), and openness (OPN) is presented in Table 1.

**Table 1:** Dataset distribution according to the personality traits

Value	EXT	NEU	AGR	CON	OPN
YES	4210	3717	5268	4556	7370
NO	5707	6200	4649	5361	2547

<sup>8</sup>[https://web.archive.org/web/20170313202822/http://mypersonality.org/wiki/lib/exe/fetch.php?media=wiki:mypersonality\\_final.zip](https://web.archive.org/web/20170313202822/http://mypersonality.org/wiki/lib/exe/fetch.php?media=wiki:mypersonality_final.zip)

To ensure the data quality and address ethical concerns, the MyPersonality dataset underwent a careful curation process. During the data selection, any external information was removed preserving the user's privacy. The original data collection was collected via a Facebook application that explicitly sought the users' consent for research purposes. Although inherent biases are common in social media-derived datasets, measures were taken to ensure diversity and representation, leveraging the diversity of Facebook users [71].

To properly evaluate the performance of PLMs integrated into the proposed architecture, the dataset has been split into a training and testing set. 80% of the status posts have been utilized for training the PLMs, whereas 20% have been utilized to evaluate the performance of customized models. A 10-fold cross-validation strategy has been followed.

In this research, the hyperparameter tuning strategy for all the models was conducted following a meticulous and systematic process to optimize the accuracy of personality trait detection. A series of experiments were executed to explore various hyperparameter combinations, especially considering critical factors such as the number of training epochs, batch sizes, optimizer choices, learning rates, and dropout rates. The approach involved searching through a predefined hyperparameter space, examining different values for key parameters such as the number of epochs (e.g., 3, 4, 5, 6, 7), batch sizes (e.g., 8, 16, 24, 32, 64), optimizer choices (e.g., Adam, SGD), learning rates (e.g.,  $1e-5$ ,  $2e-5$ ,  $1e-3$ ), and various dropout rates (e.g., 0.1, 0.3, 0.5). After numerous experiments, the optimal combination of hyperparameters for each model was identified. These best hyperparameters included the use of the Adam optimizer with a learning rate of  $1e-5$ , a batch size of 8, and a training process spanning 5 epochs, with a dropout rate of 0.3.

The implementation of the PLMs was carried out using the Huggingface library [72]. In the context of architectural parameters, the choice of the model versions and configurations was driven by their effectiveness in NLP tasks, considering factors such as architectural complexity and representation capabilities. This aspect underscored the parameter selection process, which encompassed tokenization, embedding dimension, and sequence length standardization to maintain a consistent setup across all models. Furthermore, model-specific tokenizers from Huggingface were employed to ensure model-specific adaptability. The loss function, Binary Cross-Entropy with Logits (BCEWithLogitsLoss), was also selected to perform the classification task.

The experiments were conducted using Google Colab, a cloud-based platform that provided access to T4 GPUs, essential for the effective training of deep learning models. The training times for each model reflected computational complexity. Notably, the RoBERTa model needed the highest training time, requiring approximately 16 h to complete the process. The XLNet model took around 14 h, while both ERNIE and ELECTRA needed around 12 and 11 h, respectively. In contrast, the BERT and ALBERT models had relatively shorter training times, each one taking approximately 8 h. The longest training times for some specific models were primarily due to their more extensive network architectures, which demanded additional computational resources for optimization.

#### **4.1 Evaluation Metrics**

To assess the results, the measures precision, recall, F1-score, weighted averages (P-weighted and R-weighted and F1-weighted), and accuracy, have been implemented as described in [73]:

- **Exact Match Ratio (EMR)** assesses the proportion of cases where all of their labels are classified correctly:

$$EMR = \frac{1}{n} \sum_{i=1}^n I(y_i = \hat{y}_i) \quad (1)$$

where  $n$  is the number of cases,  $I$  is the indicator function,  $y_i$  is the number of true labels for the  $i$ -th cases, and  $\hat{y}_i$  the number of correctly predicted labels for the  $i$ -th case.

- **Zero-One Loss:** This metric measures the proportion of cases where the predicted value differs from the actual value:

$$Zero - one - loss = \frac{1}{n} \sum_{i=1}^n I(y_i \neq \hat{y}_i) \quad (2)$$

- **Hamming Loss (HL)** calculates the ratio of incorrectly predicted traits to the total number of traits:

$$HL = \frac{1}{nL} \sum_{i=1}^n \sum_{j=1}^L I(y_j^i \neq \hat{y}_j^i) \quad (3)$$

- **Precision (P)** computes the proportion of correctly recognized labels over the total number of predicted labels, averaged over all cases:

$$Precision = \frac{1}{n} \sum_{i=1}^n \frac{y_i \cap \hat{y}_i}{y_i} \quad (4)$$

where  $n$  is the number of cases/status,  $y_i$  is the number of actual labels/traits for  $i$ -th case, and  $\hat{y}_i$  is the number of successfully predicted labels for  $i$ -th case.

- **Recall (R)** is the proportion of correctly recognized labels to the total number of expected labels, averaged over all cases:

$$Recall = \frac{1}{n} \sum_{i=1}^n \frac{y_i \cap \hat{y}_i}{\hat{y}_i} \quad (5)$$

- **F1-score (F1)** is the harmonic mean of precision and recall:

$$F1 - score = \frac{1}{n} \sum_{i=1}^n \frac{2(y_i \cap \hat{y}_i)}{y_i + \hat{y}_i} \quad (6)$$

- **Weighted average** considers the number of cases in each label and can yield a value of F-weighted that is not between P-weighted and R-weighted. The following equations can be used to compute the P-weighted, R-weighted, and F-weighted:

$$R_{weighted} = \sum_{i=1}^L (R(i) \times weight(i)) \quad (7)$$

$$P_{weighted} = \sum_{i=1}^L (P(i) \times weight(i)) \quad (8)$$

$$F1_{weighted} = \sum_{i=1}^L (F1(i) \times weight(i)) \quad (9)$$

- **Accuracy (ACC)** computes the proportion of correctly classified traits over the total number of traits:

$$Accuracy = \frac{1}{n} \sum_{i=1}^n \frac{y_i \cap \hat{y}_i}{y_i \cup \hat{y}_i} \quad (10)$$

## 5 Results and Discussion

This section summarizes the findings and contributions made after assessing the effectiveness of the different PLMs tuned for detecting multiple personality traits in the dataset.

As presented in [Table 2](#), ERNIE demonstrated outstanding results, achieving an accuracy of 87.17%, an F1-score of 84.41%, and an EMR of 73.32%. ERNIE's performance surpasses the other models, underlining its accuracy in predicting personality traits. The high EMR value is particularly noteworthy, as it indicates that ERNIE consistently classifies all personality traits correctly for a significant proportion of cases. This consistency underscores ERNIE's robustness and reliability for this task.

**Table 2:** Performance results for the PLMs

Models/Metrics	BERT	RoBERTa	ALBERT	XLNet	ELECTRA	ERNIE
EMR	68.98	57.87	62.63	70.48	65.28	72.32
0/1 loss	31.02	42.13	37.37	29.52	34.72	27.68
Ham. loss	16.69	24.11	20.64	14.08	18.40	12.83
Accuracy	83.31	75.89	79.36	85.92	81.60	87.17
Precision	80.91	69.67	72.66	80.73	74.19	82.85
Recall	77.73	74.43	79.36	84.97	82.91	86.23
F1-score	78.98	71.16	75.21	82.73	78.13	84.41
P-weighted	82.73	71.44	74.88	82.48	76.44	84.02
R-weighted	78.56	76.78	81.48	86.70	84.31	87.61
F1-weighted	80.36	73.22	77.41	84.36	80.09	85.55

ERNIE's performance can be attributed to its unique strengths, which stem from diverse pre-training tasks, including masked language modeling, next-sentence prediction, and sentiment analysis. Unlike many other PLMs that rely on a single pre-training task, ERNIE's multifaceted training process enables it to attain a more comprehensive understanding of the English language. These results demonstrate that ERNIE excels, not only in particular individual trait prediction but also, in ensuring the accuracy of multiple personality trait detection.

One of ERNIE's notable attributes is its sequential learning capability, allowing it to accumulate knowledge progressively. This sequential learning enables ERNIE to manage complicated tasks, such as natural language inference, by continually building on its previously acquired knowledge. Moreover, ERNIE distinguishes itself thanks to its larger model size in comparison to other PLMs. This increased model size equips ERNIE with an enhanced capacity to learn and represent complex relationships within the data, contributing to its outstanding performance in personality trait prediction from the text [74].

XLNet achieved comparable results to the ERNIE model, outperforming the other PLMs (BERT, ALBERT, ELECTRA, and RoBERTa). One of the main reasons is that it can obtain a bidirectional context representation of a term by training an autoregressive model over all potential permutations of terms in a sentence, and it has no sequence length limit, unlike BERT, overcoming some problems of the BERT model. On the contrary, ALBERT and RoBERTa obtained the lowest results; their accuracy was improved by ERNIE by 7.81% and 11.28%, respectively.

Compared with previous studies using the same gold standard, the results (see [Table 3](#)) show that all the tested PLMs significantly outperform the state-of-the-art studies in terms of accuracy. This is the primary metric used in these studies, for that reason, it is used for comparison.

**Table 3:** Comparison of this approach with other state-of-the-art models

Reference	Model	EXT	NEU	AGR	CON	OPN	Accur.
Tandera et al. [9]	LSTM+CNN 1D	71.05	58.97	50.00	57.69	75.86	62.71
EL-Demerdash [10]	CNN+ PLMs	80.55	79.00	63.69	65.31	81.00	73.91
Wang et al. [52]	HMAtn-ECBiL	73.94	63.14	70.74	68.65	84.57	72.01
Tadesse et al. [11]	SNA+ XGBoost	78.00	68.00	65.30	69.80	73.30	71.00
Zhao et al. [51]	Att.-based LSTM	–	–	–	–	–	57.95
Darliansyah [12]	GloVe+CNN+LSTM	58.51	62.24	54.77	57.91	72.24	61.13
Our proposal	ERNIE	86.53	90.56	86.53	84.11	88.09	87.16

The techniques used by these previous studies (see [Table 3](#)) are primarily based on classic machine learning models and deep learning architectures, word embeddings, and attention mechanisms, trained on the proposed dataset. These techniques might require a huge amount of data to accurately tune all their parameters, especially for a complicated task such as multi-label classification. Nevertheless, since the PLMs have been trained on huge previous collections over different tasks, the tuning process might be simpler, even though the dataset is small because a great number of lexical, syntactic, and structural features over different tasks have been already learned by the models.

These results corroborate that using transfer learning has a significant impact when there is not available enough data in the target domain, but there is a suitable dataset for the source domain. Hence, transfer learning can be used to adapt models across different tasks, in this case, multiple personality traits detection, increasing the accuracy by at least 10% regarding the previous studies.

To discuss the capabilities of each PLM, the individual personality traits have been broken down in [Table 4](#). ERNIE achieved the best statistics for the F1-score for all traits, except for OPN, where XLNet obtained the best performance. As the F1-score is the harmonic mean of precision and recall, it could be thought that the best results could be also computed by ERNIE; nonetheless, it achieves the best balance between them, but not necessarily the best individual results. Thus, BERT, ALBERT, and XLNet achieved the highest precision scores for AGR, CON, and OPN, respectively; and regarding the recall score, ERNIE also achieved the highest score for EXT and AGR, whereas BERT, XLNet, and ALBERT achieved the highest recall scores for NEU, CON, and OPN, respectively. Finally, according to the F1-score, XLNet just achieved the highest performance score for OPN, whereas ERNIE did for the rest of the personality traits, which makes it the most appropriate model for detecting traits.

**Table 4:** Assessment for each trait

Models/Metrics	Extroversion (EXT)	Neuroticism (NEU)	Agreeableness (AGR)	Conscientiousness (CON)	Openness (OPN)
<b>BERT</b>					
Accuracy	83.15	84.82	82.45	83.20	84.95
Precision	66.60	77.20	87.19	84.49	89.07
Recall	73.19	84.56	68.07	79.81	83.06
F1-score	69.74	80.71	76.45	82.08	85.96
<b>RoBERTa</b>					
Accuracy	76.09	77.91	73.32	76.14	78.99
Precision	53.79	76.88	64.79	78.51	74.42
Recall	70.15	58.92	79.39	69.56	94.14
F1-score	60.89	66.71	71.35	73.76	83.13
<b>ALBERT</b>					
Accuracy	78.92	78.61	79.17	78.81	81.29
Precision	59.15	67.50	70.70	87.01	78.95
Recall	66.34	83.08	85.78	65.89	95.74
F1-score	62.54	74.48	77.51	75.00	86.54
<b>ELECTRA</b>					
Accuracy	79.72	82.34	79.82	80.83	85.72
Precision	58.93	77.24	72.11	77.18	85.53
Recall	77.75	75.16	84.45	84.20	93.01
F1-score	67.04	76.19	77.80	80.54	89.11
<b>XLNet</b>					
Accuracy	85.12	87.89	84.41	82.90	89.30
Precision	69.74	85.51	79.77	79.13	89.52
Recall	77.56	81.61	84.09	87.65	93.98
F1-score	73.44	83.51	81.87	83.17	91.69
<b>ERNIE</b>					
Accuracy	86.53	90.56	86.53	84.11	88.09
Precision	71.61	89.97	82.61	83.35	86.75
Recall	81.55	84.29	85.90	83.78	95.66
F1-score	76.26	87.04	84.22	83.56	90.99

To illustrate practical variances in personality trait label predictions among different models and emphasize the importance of the EMR metric in assessing their performance, let us see the results for the following post from the dataset: “just spent the last hour looking at photos from junior

abroad in London and is dying to go back.” ERNIE and XLNet demonstrated their proficiency by accurately predicting all five personality trait labels correctly. These models excelled in comprehending the nuances of the text and making precise judgments. In contrast, the BERT and ELECTRA models exhibited limitations, struggling to predict the CON label accurately, indicating a potential challenge in understanding this specific personality trait in the given context. On the other hand, both ALBERT and RoBERTa models encountered difficulties in accurately predicting NEU and CON, implying that these traits may be particularly challenging to identify from the post.

Considering just the individual labels, it can be noted that the OPN obtained the best results in terms of F1-score in comparison with the other labels. The availability of a larger number of cases for this label in the dataset may be one of the fundamental reasons for explaining these data. In contrast, EXT and NEU comprise the lowest number of cases, being 16.75% and 14.79% of the dataset, respectively. Therefore, the F1-score for them was the lowest value for most of the PLMs. Thus, this corroborates that the greater the number of cases is, the better the identification of traits is.

Focusing on the ERNIE model, the confusion matrices demonstrate that the number of mislabeled instances is low (see Fig. 4). As it can be seen over the principal diagonal, the number of perfectly classified items (the true positives in blue and the true negatives in green) is greater than the number of incorrectly classified ones, i.e., the off-diagonal items (the false negatives in red and the false positives in yellow). Conscientiousness is the trait showing more difficulty in being classified.

Extroversion		Neuroticism		Conscientiousness		Agreeableness		Openness	
1287	170	1168	70	867	160	1003	150	555	182
97	429	117	628	155	801	117	713	54	1192

**Figure 4:** ERNIE’s confusion matrices for each personality trait

It is also interesting to analyze the percentage of misclassified personality traits. As it can be seen in Table 5, the percentage is very low for the three superior PLMs (ERNIE, XLNet, BERT), 12.83%, 14.08%, and 16.69%, respectively. OPN is the most easily classified label by most of the models, possibly as it was explained above because there are more instances in the dataset. Nonetheless, CON and AGR present more errors, in general, for most of the models, although EXT and NEU have fewer instances in the dataset.

**Table 5:** Mislabeled instances per personality trait

Labels/Model	BERT	RoBERTa	ALBERT	XLNet	ELECTRA	ERNIE
Extroversion (EXT)	334	474	418	295	402	267
Neuroticism (NEU)	301	438	424	240	350	187
Agreeableness (AGR)	348	529	413	309	400	267
Conscientiousness (CON)	333	473	420	339	389	315
Openness (OPN)	338	476	371	212	283	236
#Total errors for each model	1,654 (16.69%)	2,390 (24.11%)	2,046 (20.64%)	1,395 (14.08%)	1,824 (18.39%)	1,272 (12.83%)



Another variable that can influence the performance of the models could be the vocabulary coverage of the libraries used. In this case, most of the models share the same vocabulary (see [Table 6](#)); therefore, this factor is not very remarkable to explain the good performance of some models.

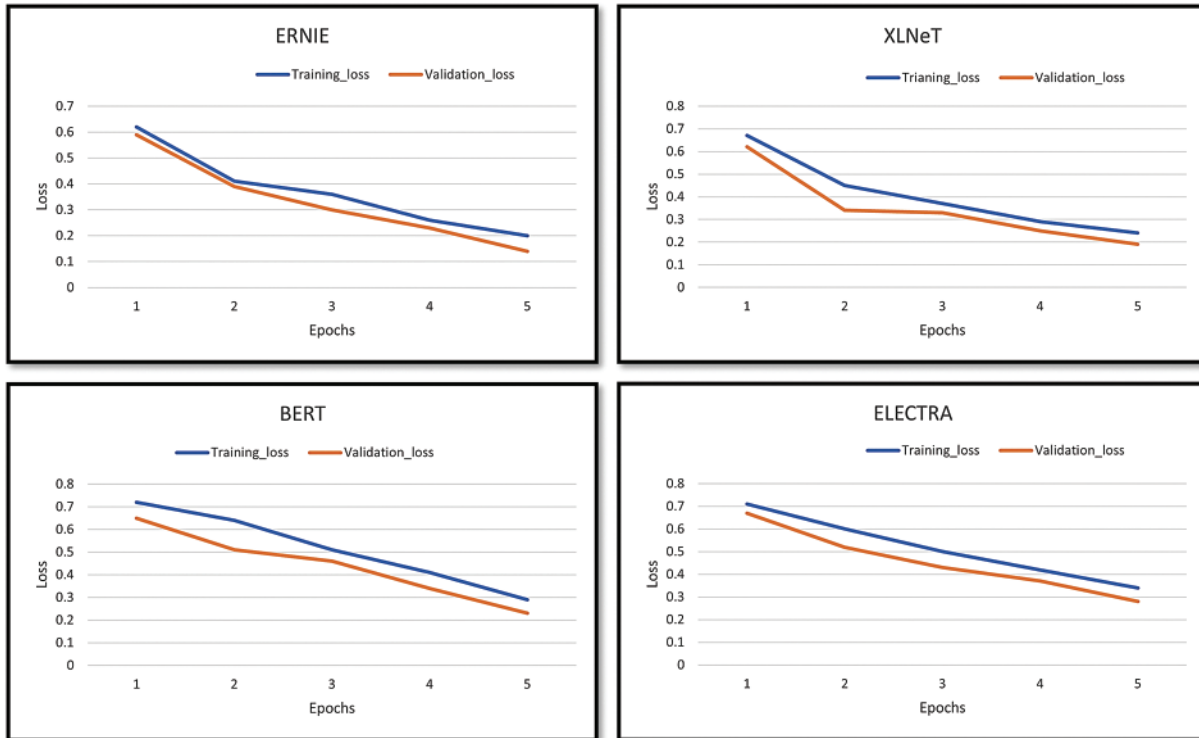
**Table 6:** Terms from the PLMs included in the MyPersonality dataset

Model	Vocabulary	Included	%	Not included	%
ERNIE, ALBERT, ELECTRA, BERT	30,519	8,651	51.27	8,221	48.73
RoBERTa	50,265	3,053	18.09	13,819	81.91

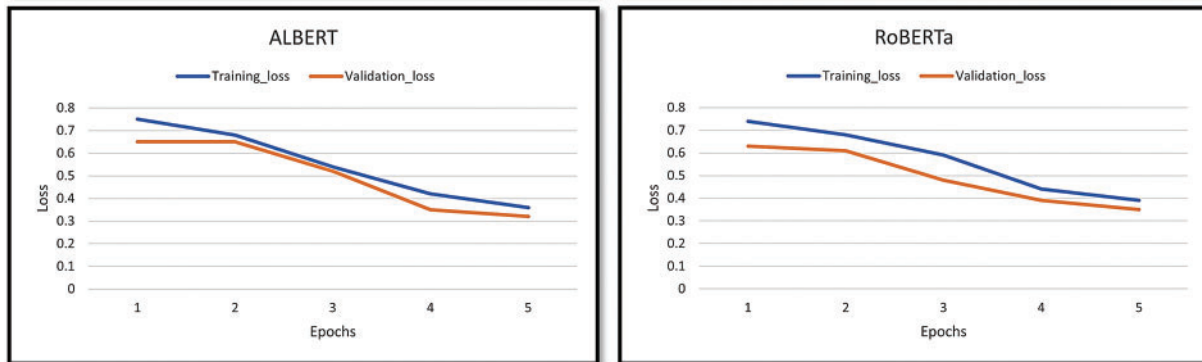
Total number of terms in MyPersonality: 16,872

Nevertheless, this aspect can explain the weak performance of the RoBERTa model. Despite being larger than the other models (50,265 terms), it just shares 3,053 (18.09%) terms with the evaluation dataset. Therefore, the goodness of the ERNIE model is not just based on the vocabulary, but also on the tasks followed to develop the model and the size of the datasets used to train it.

The average training and validation loss was also monitored to study the performance of the PLMs. A consistent pattern was observed, marked by substantial reductions in both training and validation losses up to the fifth epoch (see [Fig. 5](#)).



**Figure 5:** (Continued)



**Figure 5:** Training and validation loss for the PLMs

Beyond this threshold, the losses exhibited fluctuations, leading to the establishment of a pre-defined number of training epochs (5) as a preventive measure against overfitting. The minimal gap between training and validation losses served as additional assurance of the robustness of the model's training process [75]. ERNIE, with the lowest training and validation losses among the evaluated models, clearly demonstrated a remarkable performance, underlining the significant impact of its low loss values on its overall effectiveness.

The complexity of ERNIE's architecture allows carrying out word-aware tasks, structure-aware tasks, and semantic-aware tasks from massive data sources and the multi-task continual learning stage allows learning lexical, syntactic, and semantic information from different tasks. Unlike the other methods which usually only train the model over one specific task forgetting the previously learned knowledge, ERNIE can keep the knowledge through the entire learning process thanks to the multi-task continual learning strategy. This is one of the major ERNIE's key points, not even the huge datasets used for training the model can be considered fundamental because RoBERTa uses similar datasets (see Table 7) but is not able to achieve its results.

**Table 7:** Size of the text corpora used for each implemented PLM

PLMs	Corpora along with the number of words per corpus	Total # Words
BERT	BooksCorpus (800 M) + EnglishWikipedia text(2, 5 B)	3, 3 B
RoBERTa	BooksCorpus (800 M) + EnglishWikipedia text(2, 5 B) + CommonCrawl News (3, 8 B) + OpenWebText (1.9 B) + Stories	1, 6 B
ALBERT	BooksCorpus (800 M) + EnglishWikipedia text(2, 5 B)	3, 3 B
XLNET	Base version: BooksCorpus (1.09 B) + EnglishWikipedia (2, 78 B)	3, 87 B
ELECTRA	BooksCorpus (800 M) + EnglishWikipedia text(2, 5 B)	3, 3 B
ERNIE	For English language: EnglishWikipedia (2,021 M) + BooksCorpus (805 M) + Dialog(4908 M) + Discourse relation data (171 M)	7, 9 B

In Table 7, a summary of the datasets used for each PLM can be seen. ERNIE used the largest corpora (the Chinese corpora are not included in the used version) along with RoBERTa, whereas BERT, ALBERT, and ELECTRA used the same corpora, obtaining different results. Consequently, the vocabulary of the PLMs seems to be an important aspect to analyze the performance of the PLMs;

nonetheless, the datasets and especially, the tasks used to train the different PLMs, clearly determine the goodness of the models.

Finally, it is necessary to remark that the used dataset has 15,470 unique terms and the average length of each opinion is 8.01 terms, for that reason, it is difficult to measure the effect of several features such as the token CLS over the long opinions in this case. Furthermore, since the used dataset does not provide very specific terms, the issue of out-of-vocabulary terms has not been faced.

Among the practical implications of the proposal, it is necessary to highlight that this framework can help developers focus on designing solutions and applications that need to compute the preferences and characteristics of individual users such as recommender systems or customized decision support systems. These systems can be interesting in different domains such as marketing and advertising for personalizing services according to the different customer personalities, or in healthcare for detecting illnesses such as depression or anxiety which can be related to the different personality features. Furthermore, the capability of detecting personality traits in such a precise manner can reduce the need for interaction with experts such as psychologists or psychiatrists in many studies and developments, being replaced by frameworks like the one here presented.

Nonetheless, this framework also presents some limitations. First, the proposal does not quantify the output, i.e., the intensity of the traits. It just classifies the user as extrovert or introvert but not to what extent. Second, there are other limitations closely related to the nature of PLMs. For instance, these models show computational complexities that limit their ability to handle extended sequences. These complexities can be especially challenging when dealing with sequences that surpass the typical 512-token limit due to GPU memory constraints. Finding more efficient model architectures to capture longer-range contextual information is necessary to deal with long texts. Furthermore, fine-tuning can be parameter-inefficient, as each task often requires its own set of fine-tuned parameters. Improving efficiency involves exploring alternatives where core PLM parameters remain fixed, with small, fine-tunable adaptation modules introduced for specific tasks, thus allowing shared PLMs to serve multiple downstream applications. Moreover, this approach needs substantial computational resources they require for optimal performance, which may be a limitation in resource-constrained environments. Another limitation is the lack of datasets accordingly labeled for personality traits.

## 6 Conclusions and Future Work

Personality detection is a relevant task whose results can be applied to many other areas such as healthcare, e-commerce, recruitment processes, etc.; for that reason, its results must be accurate. In this sense, this research has introduced a comprehensive framework for multi-trait personality classification, which encompasses the adaptation and evaluation of multiple PLMs to fulfill this complex task, revealing promising results.

The meticulous customization and fine-tuning of the specific model components distinguished the methodology here presented. The precise customization and fine-tuning of specific model components were carefully tailored, introducing a dense layer, normalization layer, dropout layer, and tanh activation function. These additional layers enhanced the model's ability to capture complex patterns and relationships over the data, which were crucial for predicting accurately. Furthermore, a specialized output layer was integrated to classify the input text into the Big Five personality traits, ensuring precise and task-specific predictions. The fine-tuning process struck a balance between task-specific feature learning and leveraged the language understanding capabilities of the pre-trained base layers, aided by the Trainer API for streamlined training and hyperparameter optimization, ultimately resulting in an enhanced model performance. Furthermore, the high accuracy of the obtained results

guarantees the correct applicability of the proposed framework in other systems that need information about the user's personality.

Regarding future work, from a practical perspective, it is necessary to assess how the proposed framework can help real applications (recommender systems, decision support systems, expert systems, etc.) to improve their results and to what extent. For that reason, the next step the adaptation of previous recommender systems developed by the authors is proposed, including personality traits as a new variable for customizing the recommendations [76,77]. Moreover, although the performance has been good, it is necessary to study how good the performance of the proposal is on other personality models aside from the Big Five model. For that reason, it is proposed to adapt the framework to detect personality characteristics according to the MBTI model. To do so, the MyPersonality dataset will be first accordingly labeled. And finally, it is necessary to mention that this framework only considers the PLMs working individually; nonetheless, the consideration of all the results can provide a new perspective, therefore, it is planned to group the individual models under the architecture of an ensemble to improve the current results.

**Acknowledgement:** The authors acknowledge the support of the Spanish Ministry of Economy and Competition, General Subdirection for Gambling Regulation of the Spanish Consumption Ministry and Science and Innovation Ministry.

**Funding Statement:** This work has been partially supported by FEDER and the State Research Agency (AEI) of the Spanish Ministry of Economy and Competition under Grant SAFER: PID2019-104735RB-C42 (AEI/FEDER, UE), the General Subdirection for Gambling Regulation of the Spanish Consumption Ministry under the Grant Detec-EMO: SUBV23/00010, and the Project PLEC2021-007681 funded by MCIN/AEI /10.13039/501100011033 and by the European Union NextGenerationEU/ PRTR.

**Author Contributions:** Conceptualization: Jesus Serrano-Guerrero, Bashar Alshouha and Jose A. Olivas; Investigation: Jesus Serrano-Guerrero, Francisco P. Romero and Francisco Chiclana; Formal analysis: Jesus Serrano-Guerrero, Jose A. Olivas and Francisco Chiclana; Writing-original draft: Jesus Serrano-Guerrero, Bashar Alshouha and Francisco P. Romero; Supervision: Jesus Serrano-Guerrero, Jose A. Olivas and Francisco Chiclana; Implementation: Bashar Alshouha and Francisco P. Romero; Writing-review & editing: Jesus Serrano-Guerrero and Francisco Chiclana; Funding acquisition: Jose A. Olivas and Francisco P. Romero.

**Availability of Data and Materials:** The used data are accessible on [https://web.archive.org/web/20170313202822/http://mypersonality.org/wiki/lib/exe/fetch.php?media=wiki:mypersonality\\_final.zip](https://web.archive.org/web/20170313202822/http://mypersonality.org/wiki/lib/exe/fetch.php?media=wiki:mypersonality_final.zip).

**Conflicts of Interest:** The author declares that they have no conflicts of interest to report regarding the present study.

## References

- [1] L. R. Goldberg, "An alternative "Description of personality": The big-five factor structure," *J. Pers. Soc. Psychol.*, vol. 59, no. 6, pp. 1216–1229, 1990. doi: [10.1037/0022-3514.59.6.1216](https://doi.org/10.1037/0022-3514.59.6.1216).
- [2] A. Furnham, "The big five versus the big four: The relationship between the myers-briggs type indicator (MBTI) and NEO-PI five factor model of personality," *Pers. Individ. Dif.*, vol. 21, no. 2, pp. 303–307, 1996. doi: [10.1016/0191-8869\(96\)00033-5](https://doi.org/10.1016/0191-8869(96)00033-5).

- [3] T. Ait Baha, M. El Hajji, Y. Es-Saady, and H. Fadili, "The power of personalization: A systematic review of personality-adaptive chatbots," *SN Comput. Sci.*, vol. 4, no. 5, pp. 661, 2023. doi: [10.1007/s42979-023-02092-6](https://doi.org/10.1007/s42979-023-02092-6).
- [4] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," *Artif. Intell. Rev.*, vol. 53, no. 4, pp. 2313–2339, 2020. doi: [10.1007/s10462-019-09770-z](https://doi.org/10.1007/s10462-019-09770-z).
- [5] S. Dhelim, N. Aung, M. A. Bouras, H. Ning, and E. Cambria, "A survey on personality-aware recommendation systems," *Artif. Intell. Rev.*, pp. 1–46, 2022. doi: [10.48550/arXiv.2101.12153](https://doi.org/10.48550/arXiv.2101.12153).
- [6] M. Abdelrahman, "Personality traits, risk perception, and protective behaviors of Arab residents of Qatar during the COVID-19 pandemic," *Int. J. Ment. Health Addict.*, vol. 20, no. 1, pp. 237–248, 2022. doi: [10.1007/s11469-020-00352-7](https://doi.org/10.1007/s11469-020-00352-7).
- [7] D. Xue *et al.*, "Personality recognition on social media with label distribution learning," *IEEE Access*, vol. 5, pp. 13478–13488, 2017. doi: [10.1109/ACCESS.2017.2719018](https://doi.org/10.1109/ACCESS.2017.2719018).
- [8] P. Wang *et al.*, "Predicting self-reported proactive personality classification with weibo text and short answer text," *IEEE Access*, vol. 9, pp. 77203–77211, 2021. doi: [10.1109/ACCESS.2021.3078052](https://doi.org/10.1109/ACCESS.2021.3078052).
- [9] T. Tandra, Hendro, D. Suhartono, R. Wongso, and Y. L. Prasetyo, "Personality prediction system from facebook users," in *Proc. 2nd Int. Conf. Comput. Sci. Comput. Intell. (ICCSCI 2017)*, 2017, vol. 116, pp. 604–611. doi: [10.1016/j.procs.2017.10.016](https://doi.org/10.1016/j.procs.2017.10.016).
- [10] K. El-Demerdash, R. A. El-Khoribi, M. A. Ismail Shoman, and S. Abdou, "Deep learning based fusion strategies for personality prediction," *Egypt. Inform. J.*, vol. 23, no. 1, pp. 47–53, 2022. doi: [10.1016/j.eij.2021.05.004](https://doi.org/10.1016/j.eij.2021.05.004).
- [11] M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality predictions based on user behavior on the facebook social media platform," *IEEE Access*, vol. 6, pp. 61959–61969, 2018. doi: [10.1109/ACCESS.2018.2876502](https://doi.org/10.1109/ACCESS.2018.2876502).
- [12] A. Darliansyah, M. A. Naeem, F. Mirza, and R. Pears, "Sentipede: A smart system for sentiment-based personality detection from short texts," *J. Univers. Comput. Sci.*, vol. 25, no. 10, pp. 1323–1352, 2019. doi: [10.3217/jucs-025-10-1323](https://doi.org/10.3217/jucs-025-10-1323).
- [13] Q. Yang, Y. Zhang, W. Dai, and S. J. Pan, *Transfer Learning*. Cambridge, UK: Cambridge University Press, 2020.
- [14] S. Bashath, N. Perera, S. Tripathi, K. Manjang, M. Dehmer, and F. E. Streib, "A data-centric review of deep transfer learning with applications to text data," *Inf. Sci. (Ny)*, vol. 585, pp. 498–528, 2022. doi: [10.1007/10.1016/j.ins.2021.11.061](https://doi.org/10.1007/10.1016/j.ins.2021.11.061).
- [15] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016. doi: [10.1186/s40537-016-0043-6](https://doi.org/10.1186/s40537-016-0043-6).
- [16] D. Cook, K. D. Feuz, and N. C. Krishnan, "Transfer learning for activity recognition: A survey," *Knowl. Inf. Syst.*, vol. 36, pp. 537–556, 2013. doi: [10.1007/s10115-013-0665-3](https://doi.org/10.1007/s10115-013-0665-3).
- [17] M. T. Bahadori, Y. Liu, and D. Zhang, "A general framework for scalable transductive transfer learning," *Knowl. Inf. Syst.*, vol. 38, pp. 61–83, 2014. doi: [10.1007/s10115-013-0647-5](https://doi.org/10.1007/s10115-013-0647-5).
- [18] O. Day and T. M. Khoshgoftaar, "A survey on heterogeneous transfer learning," *J. Big Data*, vol. 4, pp. 1–42, 2017. doi: [10.1186/s40537-017-0089-0](https://doi.org/10.1186/s40537-017-0089-0).
- [19] P. Prettenhofer and B. Stein, "Cross-lingual adaptation using structural correspondence learning," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 1, pp. 1–22, 2011. doi: [10.1145/2036264.2036277](https://doi.org/10.1145/2036264.2036277).
- [20] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, 2021. doi: [10.1109/TKDE.2021.3070203](https://doi.org/10.1109/TKDE.2021.3070203).
- [21] X. Han *et al.*, "Pre-trained models: Past, present and future," *AI Open*, vol. 2, pp. 225–250, 2021. doi: [10.1016/j.aiopen.2021.08.002](https://doi.org/10.1016/j.aiopen.2021.08.002).
- [22] S. Alaparthy and M. Mishra, "BERT: A sentiment analysis odyssey," *J. Mark. Anal.*, vol. 9, no. 2, pp. 118–126, 2021. doi: [10.1057/s41270-021-00109-8](https://doi.org/10.1057/s41270-021-00109-8).
- [23] P. Do and T. H. V. Phan, "Developing a BERT based triple classification model using knowledge graph embedding for question answering system," *Appl. Intell.*, vol. 52, no. 1, pp. 636–651, 2022. doi: [10.1007/s10489-021-02460-w](https://doi.org/10.1007/s10489-021-02460-w).

- [24] Y. Liu and M. Lapata, "Text summarization with pretrained encoders," arXiv Preprint. arXiv:1908.08345, 2019.
- [25] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," in *Acc. 5th Workshop Eng. Effic. Mach. Learn. Cogn. Comput.-NeurIPS 2019*, 2020. abs/1910.0.
- [26] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for self-supervised learning of language representations," in *8th Int. Conf. Learn. Represent.*, 2019, pp. 1–17.
- [27] Y. Liu *et al.*, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019.
- [28] F. Mi, W. Zhou, F. Cai, L. Kong, M. Huang, and B. Faltings, "Self-training improves pre-training for few-shot learning in task-oriented dialog systems," in *Conf. Empir. Methods Nat. Lang. Process.*, 2021, pp. 1–12.
- [29] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 16000–16009.
- [30] Q. Wu, T. Yang, Z. Liu, B. Wu, Y. Shan, and A. B. Chan, "DropMAE: Masked autoencoders with spatial-attention dropout for tracking tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 14561–14571.
- [31] Y. Gong *et al.*, "Contrastive audio-visual masked autoencoder," in *Proc. Elev. Int. Conf. Learn. Represent. (ICLR)*, Kigali, Rwanda, May 1–5, 2022.
- [32] G. Li, H. Zheng, D. Liu, C. Wang, B. Su, and C. Zheng, "Semmae: Semantic-guided masking for learning masked autoencoders," *36th Conference on Neural Information Processing Systems (NeurIPS 2022)*, vol. 35, pp. 14290–14302.
- [33] S. C. Fanni, M. Febi, G. Aghakhanyan, and E. Neri, Natural language processing. in *Introduction to Artificial Intelligence*, Springer, 2023, pp. 87–99. doi: [10.1007/978-81-322-3972-7\\_19](https://doi.org/10.1007/978-81-322-3972-7_19).
- [34] P. Kaur and R. K. Singh, "A review on optimization techniques for medical image analysis," *Concurr. Comput. Pract. Exp.*, vol. 35, no. 1, pp. e7443, 2023. doi: [10.1002/cpe.7443](https://doi.org/10.1002/cpe.7443).
- [35] M. A. Haq, A. K. Jilani, and P. Prabu, "Deep learning based modeling of groundwater storage change," *Comput. Mater. Contin.*, vol. 70, no. 3, pp. 4599–4617, 2021. doi: [10.32604/cmc.2022.020495](https://doi.org/10.32604/cmc.2022.020495).
- [36] M. A. Haq, "CDLSTM: A novel model for climate change forecasting," *Comput. Mater. Contin.*, vol. 71, no. 2, pp. 2363–2381, 2022. doi: [10.32604/cmc.2022.023059](https://doi.org/10.32604/cmc.2022.023059).
- [37] M. A. Haq *et al.*, "Analysis of environmental factors using AI and ML methods," *Sci. Rep.*, vol. 12, no. 1, pp. 13267, 2022. doi: [10.1038/s41598-022-16665-7](https://doi.org/10.1038/s41598-022-16665-7).
- [38] S. Bharadwaj, S. Sridhar, R. Choudhary, and R. Srinath, "Persona traits identification based on myers-briggs type indicator(MBTI)—A text classification approach," in *2018 Int. Conf. Adv. Comput., Commun. Inform. (ICACCI)*, 2018, pp. 1076–1082. doi: [10.1109/ICACCI.2018.8554828](https://doi.org/10.1109/ICACCI.2018.8554828).
- [39] M. Gjurković and J. Šnajder, "Reddit: A gold mine for personality prediction," in *Proc. Second Workshop Comput. Model. People's Opin., Pers., Emot. Soc. Media*, 2018, pp. 87–97. doi: [10.18653/v1/w18-1112](https://doi.org/10.18653/v1/w18-1112).
- [40] N. R. Ngatirin, Z. Zainol, and T. L. Chee Yoong, "A comparative study of different classifiers for myers-brigg personality prediction model," in *2016 6th IEEE Int. Conf. Control Syst., Comput. Eng (ICCSCE)*, 2016, pp. 435–440. doi: [10.1109/ICCSCE.2016.7893613](https://doi.org/10.1109/ICCSCE.2016.7893613).
- [41] V. Ong *et al.*, "Personality prediction based on Twitter information in Bahasa Indonesia," *Proc. 2017 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2017*, vol. 11, pp. 367–372, 2017. doi: [10.15439/2017F359](https://doi.org/10.15439/2017F359).
- [42] M. H. Amirhosseini and H. Kazemian, "Machine learning approach to personality type prediction based on the Myers–Briggs type indicator," *Multimodal Technol. Interact.*, vol. 4, no. 1, pp. 9, 2020. doi: [10.3390/mti4010009](https://doi.org/10.3390/mti4010009).
- [43] D. Azucar, D. Marengo, and M. Settanni, "Predicting the big 5 personality traits from digital footprints on social media: A meta-analysis," *Pers. Individ. Dif.*, vol. 124, pp. 150–159, 2018. doi: [10.1016/j.paid.2017.12.018](https://doi.org/10.1016/j.paid.2017.12.018).
- [44] S. Han, H. Huang and Y. Tang, "Knowledge of words: An interpretable approach for personality recognition from social media," *Knowl.-Based Syst.*, vol. 194, pp. 105550, 2020. doi: [10.1016/j.knosys.2020.105550](https://doi.org/10.1016/j.knosys.2020.105550).

- [45] K. N. P. Kumar and M. L. Gavrilova, "Personality traits classification on twitter," in *2019 16th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, 2019, pp. 1–8. doi: [10.1109/AVSS.2019.8909839](https://doi.org/10.1109/AVSS.2019.8909839).
- [46] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intell. Syst.*, vol. 32, no. 2, pp. 74–79, 2017. doi: [10.1109/MIS.2017.23](https://doi.org/10.1109/MIS.2017.23).
- [47] X. Sun, B. Liu, J. Cao, J. Luo, and X. Shen, "Who am I? Personality detection based on deep learning for texts," in *IEEE Int. Conf. Commun.*, Kansas City, MO, USA, 2018, pp. 1–6. doi: [10.1109/ICC.2018.8422105](https://doi.org/10.1109/ICC.2018.8422105).
- [48] H. Ahmad, M. U. Asghar, M. Z. Asghar, A. Khan, and A. H. Mosavi, "A hybrid deep learning technique for personality trait classification from text," *IEEE Access*, vol. 9, pp. 146214–146232, 2021. doi: [10.1109/ACCESS.2021.3121791](https://doi.org/10.1109/ACCESS.2021.3121791).
- [49] M. A. Rahman, A. Al Faisal, T. Khanam, M. Amjad, and M. S. Siddik, "Personality detection from text using convolutional neural network," in *1st Int. Conf. Adv. Sci., Eng. Robot. Technol. (ICASERT)*, Dhaka, Bangladesh, 2019, pp. 1–6. doi: [10.1109/ICASERT.2019.8934548](https://doi.org/10.1109/ICASERT.2019.8934548). 2019.
- [50] D. Xue *et al.*, "Deep learning-based personality recognition from text posts of online social networks," *Appl. Intell.*, vol. 48, no. 11, pp. 4232–4246, 2018. doi: [10.1007/s10489-018-1212-4](https://doi.org/10.1007/s10489-018-1212-4).
- [51] J. Zhao, D. Zeng, Y. Xiao, L. Che, and M. Wang, "User personality prediction based on topic preference and sentiment analysis using LSTM model," *Pattern Recogn. Lett.*, vol. 138, pp. 397–402, 2020. doi: [10.1016/j.patrec.2020.07.035](https://doi.org/10.1016/j.patrec.2020.07.035).
- [52] X. Wang, Y. Sui, K. Zheng, Y. Shi, and S. Cao, "Personality classification of social users based on feature fusion," *Sens.*, vol. 21, no. 20, pp. 6758, 2021. doi: [10.3390/s21206758](https://doi.org/10.3390/s21206758).
- [53] Z. Ren, Q. Shen, X. Diao, and H. Xu, "A sentiment-aware deep learning approach for personality detection from text," *Inf. Process. Manag.*, vol. 58, no. 3, pp. 102532, 2021. doi: [10.1016/j.ipm.2021.102532](https://doi.org/10.1016/j.ipm.2021.102532).
- [54] M. S. Anari, K. Rezaee, and A. Ahmadi, "TraitLWNet: A novel predictor of personality trait by analyzing Persian handwriting based on lightweight deep convolutional neural network," *Multimed. Tools Appl.*, vol. 81, no. 8, pp. 10673–10693, 2022. doi: [10.1007/s11042-022-12295-3](https://doi.org/10.1007/s11042-022-12295-3).
- [55] P. William, A. Badholia, B. Patel, and M. Nigam, "Hybrid machine learning technique for personality classification from online text using HEXACO model," in *2022 Int. Conf. Sustain. Comput. Data Commun. Syst. (ICSCDS)*, 2022, pp. 253–259.
- [56] M. D. Kamalesh and B. Bharathi, "Personality prediction model for social media using machine learning technique," *Comput. Electr. Eng.*, vol. 100, pp. 107852, 2022. doi: [10.1016/j.compeleceng.2022.107852](https://doi.org/10.1016/j.compeleceng.2022.107852).
- [57] M. Ramezani *et al.*, "Automatic personality prediction: An enhanced method using ensemble modeling," *Neural Comput. Appl.*, vol. 34, no. 21, pp. 18369–18389, 2022. doi: [10.1007/s00521-022-07444-6](https://doi.org/10.1007/s00521-022-07444-6).
- [58] C. Suman, S. Saha, A. Gupta, S. K. Pandey, and P. Bhattacharyya, "A multi-modal personality prediction system," *Knowl.-Based Syst.*, vol. 236, pp. 107715, 2022. doi: [10.1016/j.knosys.2021.107715](https://doi.org/10.1016/j.knosys.2021.107715).
- [59] K. El-Demerdash, R. A. El-Khoribi, M. A. I. Shoman, and S. Abdou, "Psychological human traits detection based on universal language modeling," *Egypt. Inform. J.*, vol. 22, no. 3, pp. 239–244, 2021. doi: [10.1016/j.eij.2020.09.001](https://doi.org/10.1016/j.eij.2020.09.001).
- [60] S. Aslan and U. Gdkbay, "Multimodal video-based apparent personality recognition using long short-term memory and convolutional neural networks," arXiv preprint arXiv:1911.00381, 2019.
- [61] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, and S. Eetemadi, "Bottom-up and top-down: Predicting personality with psycholinguistic and language model features," in *Proc. IEEE Int. Conf. Data Min., ICDM*, 2020, pp. 1184–1189. doi: [10.1109/ICDM50108.2020.00146](https://doi.org/10.1109/ICDM50108.2020.00146).
- [62] J. Devlin, M. W. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. 2019 Conf. N. Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol.*, 2019, pp. 4171–4186. doi: [10.18653/v1/n19-1423](https://doi.org/10.18653/v1/n19-1423).
- [63] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. the 33rd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, pp. 5753–5763.
- [64] K. Clark, M. T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training text encoders as discriminators rather than generators," in *8th Int. Conf. Learn. Represent.*, 2020, pp. 1–18.

- [65] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun and Q. Liu, "Enhanced language representation with informative entities," in *Proc. 57th Conf. Assoc. Comput. Linguist.*, 2020, pp. 1441–1451. doi: [10.18653/v1/p19-1139](https://doi.org/10.18653/v1/p19-1139).
- [66] X. P. Qiu, T. X. Sun, Y. G. Xu, Y. F. Shao, N. Dai, and X. J. Huang, "Pre-trained models for natural language processing: A survey," *Sci. China Technol. Sci.*, vol. 63, no. 10, pp. 1872–1897, 2020. doi: [10.1007/s11431-020-1647-3](https://doi.org/10.1007/s11431-020-1647-3).
- [67] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, pp. 5998–6008, 2017.
- [68] A. Rogers, O. Kovaleva, and A. Rumshisky, "A primer in BERTology: What we know about how BERT works," *Trans. Assoc. Comput. Linguist.*, vol. 8, pp. 842–866, 2021. doi: [10.1162/tacl\\_a\\_00349](https://doi.org/10.1162/tacl_a_00349).
- [69] Y. Wang, Y. Hou, W. Che, and T. Liu, "From static to dynamic word representations: A survey," *Int. J. Mach. Learn. Cybern.*, vol. 11, pp. 1611–1630, 2020. doi: [10.1007/s13042-020-01069-8](https://doi.org/10.1007/s13042-020-01069-8).
- [70] K. Sun, X. Luo, and M. Y. Luo, "A survey of pretrained language models," in *Int. Conf. Knowl. Sci., Eng. Manag.*, 2022, pp. 442–456.
- [71] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines," *Am. Psychol.*, vol. 70, no. 6, pp. 543–556, 2015. doi: [10.1037/a0039210](https://doi.org/10.1037/a0039210).
- [72] T. Wolf *et al.*, "Transformers: State-of-the-art natural language processing," in *Proc. 2020 Conf. Empir. Methods Nat. Lang. Process.: Syst. Demonstratio.*, 2020, pp. 38–45. doi: [10.18653/v1/2020.emnlp-demos.6](https://doi.org/10.18653/v1/2020.emnlp-demos.6).
- [73] M. S. Sorower, "A literature survey on algorithms for multi-label learning," Oregon State University, Corvallis, USA, vol. 18, pp. 1–25, 2010.
- [74] Y. Sun *et al.*, "ERNIE 2.0: A continual pre-training framework for language understanding," in *Proc. AAAI Conf. Artif. Intell.*, 2020, pp. 8968–8975.
- [75] M. A. Haq and M. A. Rahim Khan, "DNNBoT: Deep neural network-based botnet detection and classification," *Comput. Mater. Contin.*, vol. 71, no. 1, pp. 1729–1750, 2022. doi: [10.32604/cmc.2022.020938](https://doi.org/10.32604/cmc.2022.020938).
- [76] J. Serrano-Guerrero, F. Chiclana, J. A. Olivas, F. P. Romero and E. Homapour, "A T1OWA fuzzy linguistic aggregation methodology for searching feature-based opinions," *Knowl.-Based Syst.*, vol. 189, pp. 105131, 2020. doi: [10.1016/j.knosys.2019.105131](https://doi.org/10.1016/j.knosys.2019.105131).
- [77] J. Serrano-Guerrero, J. A. Olivas and F. P. Romero, "A T1OWA and aspect-based model for customizing recommendations on eCommerce," *Appl. Soft Comput. J.*, vol. 97, pp. 106768, 2020. doi: [10.1016/j.asoc.2020.106768](https://doi.org/10.1016/j.asoc.2020.106768).