Text-based automatic personality prediction: a bibliographic review

Ali-Reza Feizi-Derakhshi1 · Mohammad-Reza Feizi-Derakhshi1 · Majid Ramezani1 · Narjes Nikzad-Khasmakhi1 · Meysam Asgari-Chenaghlu1 · Taymaz Akan1,2,3 · Mehrdad Ranjbar-Khadivi1,4 · Elnaz Zafarni-Moattar1,5 · Zoleikha Jahanbakhsh-Naghadeh1,6

Received: 16 January 2022 / Accepted: 17 August 2022 / Published online: 6 September 2022
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2022

Abstract
Personality detection is an old topic in psychology and automatic personality prediction (or perception) (APP) is the automated (computationally) forecasting of the personality on different types of human generated/exchanged contents (such as text, speech, image, and video). The principal objective of this study is to offer a shallow (overall) review of natural language processing approaches on APP since 2010. With the advent of deep learning and following it transfer-learning and pre-trained model in NLP, APP research area has been a hot topic, so in this review, methods are categorized into three: pre-trained independent, pre-trained model based, and multimodal approaches. In addition, to achieve a comprehensive comparison, reported results are informed by datasets.

Keywords Automatic personality prediction · Natural language processing (NLP) · Text mining · Personality trait

Introduction

“The development of language is part of the development of the personality, for words are the natural means of expressing thoughts and establishing understanding between people.”, Maria Montessori.

The above quotation becomes the basis of what is present in this article, studying natural language processing in individual personality. Personality is defined as the characteristic set of behaviors, cognitions, and emotional patterns [1] as well as thinking patterns [2], and its external appearance can be seen in writing, speech,
decision and other aspects of the social and personal lives of people. Language is the
most prominent and the most available aspects of individuals’ personality. Mean-
while, written text is one of the most utilized appearance of language. Developing
the Internet based infrastructure, such as social media, e-mails, and different tex-
ting contexts, have made the language appearance of people more available. Con-
sequently, considering the increase in internet-based communications, it would be
so exciting to become aware of individuals’ personality, in spite of their absence.
Therefore, the involvement of computers in determining the personality of people
seems necessary and turned into a study field in computer science.

*Personality* refers to a person’s long-term set of characteristics and behaviors [3].
It encompasses people’s moods, attitudes, and ideas, and it manifests itself in their
relationships with others [4]. In general, it encompasses all the behavioral features
(both inherent and acquired) that may be noticed in people’s social interactions and
even their interactions with their surroundings. The word personality comes from
the Latin *persona*, which refers to a mask worn by an actor in ancient Greek and
Roman plays to symbolize and convey the character’s personality attributes (please
refer to [5–7] for more details).

Automatic personality prediction (or perception) (APP) is the automatic pre-
diction of the personality of individuals and usually done by computers. With the
increasing variety of data types available for analyzing the personality of people,
aspects of view to APP increases likewise. In this point of view to the assortment
of APP, data types can be named as: speech [8–11], image [12–15], video [16, 17],
text [18–20], social media activities [21–23], touch screen interaction [24, 25], and
so on. Also, each of these has subsets and divisions of text-based APP which can
be mentioned are email [26], SMS [27], and tweets and posts on social media [28].
Thereby, the key standpoint of this study is analyzing APP methods through text
data type and NLP.

Personality should be measured and classified to make it more compara-
tive, and this goes back to psychology. Psychologist put forward many personal-
ity trait models, such as Allport’s trait theory [29], Cattell’s 16 Factor Model [30]
(Table 14 shows characteristics of traits), Eysenck Personality Questionnaire (EPQ)
[31], Myers–Briggs Type Indicator (MBTI) [32], and Big Five [33]. Among these,
two models, MBTI and Big Five, are popular and widely used models, especially
in APPs. MBTI has four main dimensions, Introversion versus Extraversion (I-E),
Sensing versus iNtuiting (S-N), Thinking versus Feeling (T-F), and Judging versus
Perceiving (J-P), that each people categorized in two dimensions. Figure 2 defines
each MBTI dimension characteristics. The second popular personality model is Big
Five. This model consists of five traits, and people may get in one or more trait. Also,
two different approaches are taken thus binary modeling (0 and 1 for each trait) or
continuous modeling (each trait get a value in range 0–1); those two approaches are
being used in APP datasets. Openness, Conscientiousness, Extroversion, Agreeable-
ness, and Neuroticism are traits of Big Five which are called OCEAN in abbrevi-
ative. Table 1 illustrates characteristics in each OCEAN trait. Without any doubt,
from the psychological perspective, there are several drawbacks in current APP
methods that basically refer to the existing datasets like assigning binary labels to
individuals’ personality traits which are considered as a continuum. Obviously,
Table 1  An overview of Big Five personality traits model [18]

| Personality trait | Description of LOW values - $\infty \leftarrow$ | Description of HIGH values $\rightarrow +\infty$ |
|-------------------|-----------------------------------------------|-----------------------------------------------|
| **Openness (O)**  | Dislikes changes                              | Very creative                                 |
|                   | Does not enjoy new things                     | Clever, insightful, daring, and varied interests|
|                   | Conventional                                  | Embraces trying new things or visiting new places|
|                   | Resists new ideas                             | Unconventional                                |
|                   | Prefers familiarity                           | Focused on tackling new challenges            |
|                   | Not very imaginative                          | Intellectually curious                        |
|                   | Has trouble with abstract or theoretical concepts | Inventive                                    |
|                   | Skeptical                                     | Happy to think about abstract concepts        |
|                   | Traditional in thinking                       | Enjoys the art                                |
|                   | Consistent and cautious                       | Eager to meet new people                     |
| **Conscientiousness (C)** | Easy going and careless                      | Competent and efficient                       |
|                   | Messy and less detail-oriented                | Goal- and detail-oriented                    |
|                   | Dislikes structure and schedule               | Well organized, self-discipline and dutiful   |
|                   | Fails to return things or put them back, where they belong | Spends time preparing                        |
|                   | Procrastinates on important tasks and rarely completes them on time | Predictable and deliberate                   |
|                   | Fails to stick to a schedule                  | Finishes important tasks on time             |
|                   | Is always late when meeting others            | Does not give in to impulses                 |
|                   |                                               | Enjoys adhering to a schedule                |
|                   |                                               | Is on time when meeting others               |
|                   |                                               | Works hardly                                 |
|                   |                                               | Reliable and resourceful                     |
|                   |                                               | Persevered                                   |
| Personality trait | Description of LOW values - \( \infty \) ⟷ | Description of HIGH values ⟷ +\( \infty \) |
|-------------------|---------------------------------|----------------------------------|
| Extroversion (E)  | Introspective                    | Outgoing and energetic           |
|                   | Solitary and reserved            | Assertive and talkative          |
|                   | Dislikes being at the center of attentions | Able to be articulate          |
|                   | Feels exhausted when having to socialize a lot | Enjoys being the center of attentions |
|                   | Finds it difficult to start conversations | Likes to start conversations     |
|                   | Dislikes making small talks      | Enjoys being with others and meeting new people |
|                   | Carefully thinks things before speaking | Tendency to be affectionate      |
|                   | Thoughtful                       | Finds it easy to make new friends |
|                   |                                  | Has a wide social circle of friends and acquaintances |
|                   |                                  | Says things before thinking about them |
|                   |                                  | Feels organized when around other people |
|                   |                                  | Social confidence                |
| Agreeableness (A) | Challenging and detached         | Friendly and compassionate toward others |
|                   | Takes little interest in others  | Altruist and unselfish           |
|                   | Can be seen as insulting or dismissive of others | Loyal and patient                |
|                   | Does not care about other people’s feelings or problems | Has a great deal of interest in and wants to help others |
|                   | Can be manipulative              | Feels empathy and concern for other people |
|                   | Prefers to be competitive and stubborn | Prefers to cooperate and be helpful |
|                   | Insults and belittles others     | Polite and trustworthy           |
|                   |                                  | Cheerful and considerate         |
|                   |                                  | Modest                           |
| Neuroticism (N)   |                                 |                                  |

*Table 1 (continued)*
| Description of LOW values - $\infty \leftarrow$ | Personality trait | $\rightarrow +\infty$ Description of HIGH values |
|-----------------------------------------------|------------------|-----------------------------------------------|
| Emotionally stable                           |                  | Anxious of many different things and nervous   |
| Deals well with stress                       |                  | Experiences a lot of stress                    |
| Rarely feels sad or depressed                |                  | Irritable                                      |
| Does not worry much and is very relax        |                  | Impulsive and moody                            |
| Confident and secure                         |                  | Jealous                                        |
| Optimist                                      |                  | Lack of confidence                              |
|                                              |                  | Self-criticism                                 |
|                                              |                  | Oversensitive                                   |
|                                              |                  | Instable and insecure                           |
|                                              |                  | Timid                                           |
|                                              |                  | Pessimist                                       |
having comprehensive datasets would remove such drawbacks and improve the performance of APP.

Automatic personality prediction is a young research subject in computer science and artificial intelligence and, in recent years, has become a very active research subject. Further research is demanded to maximize the impact of automated personality prediction, as this is still in its early stages. This review will seek to provide an overview and comprehensive view of the different types of text-based automatic personality prediction methods. The motivation for this review is based on the following:

- Personality classification is still in the early stages of development and requires further investigation. As a challenging and complex issue, investigating additional directions for future research is essential to enrich extant personality classification techniques further.
- The need for a bibliographic literature review is observed after investigating the gradual research regarding personality classification. Hence, the present work is based on the bibliographic literature review.
- This review has been motivated by the rapid advances in personality classification, so the researchers have identified, summarized and evaluated the relevant studies in this field.

In recent years, the NLP field is faced with a revolution that APP does not get its benefits. In this study, APP articles from 2010, which involve textual inputs, are reviewed in three categories: classical text representation and feature extraction methods, articles assisted novel pre-trained word representations, and methods with multimodal approaches (besides text, other data types included).

The rest of this study is organized in the following manner: in the next section, materials and methods which are used as the baselines of APP studies are introduced briefly. The third section is the overview of methods and consist of three sub-sections. The results of studies are structured based on datasets and are shown in the fourth section; the datasets are also explained in this section. Finally, some concluding remarks are given in the last section.

**NLP materials in personality prediction**

The approach of this study is to overview researches in APP, which conducted on texts. To this end, material and methods of text analysis are given in this section briefly. Linguistic inquiry and word count (LIWC) is one of the most used and developed tools of APP, and in “Linguistic inquiry and word count (LIWC)”, it has been reported. The next material based on NLP is called MRC and is a dictionary psycholinguistic in English. The last one is about embedding techniques that represent words for text analysis, typically in the form of a real-valued vector that encodes the meaning of the word such that the words that are closer in the vector space are expected to be similar in meaning.
Linguistic inquiry and word count (LIWC)

Linguistic inquiry and word count (LIWC) is introduced by [34] and developed in years [35, 36] as NLP tools for the psychological purpose. LIWC is a text analysis tool that provides statistical reports that are very useful in determining texts to aim for emotional and cognitive analysis of people. Since 2001, two updated versions of LIWC are introduced in 2007 and 2015. In each version, some features were added, and Table 2 shows all features reported by LIWC and the differences between the last two versions. LIWC reports consist of 91 features in 15 categories. This test could be done online in https://liwc.wpengine.com/.

[38] developed a method based on LIWC on the Essays dataset. The authors fed the Essays dataset to LIWC, and the outputs contain LIWC features with a personality trait label in the Big Five dimension. Since the LIWC queries are not free, Mairesse dataset deploys as LIWC features in most research. Mairesse developed a framework and is available in http://farm2.user.srcf.net/research/personality/recognizer, but it should be noted that it is a while that the framework did not support.

MRC psycholinguistic database

MRC is a publicly available machine usable dictionary that includes different (up to 26) linguistic and psycholinguistic attributes for 150,837 English words. Different semantic, syntactic, phonological, and orthographic details about the words have made it suitable for miscellaneous purposes of researches in psychology, linguistics and artificial intelligence. Word association data are also included in the database. The first version was introduced in 1981 [39] and the second and last versions [40] are now available in https://websites.psychology.uwa.edu.au/school/mrcdatabase/mrc2.html with the details and statistics.

Embedding techniques

Any input should be modeled to be understood by the computer and writings are no exception, so embeddings duty is that. The smallest meaningful segment of writing is words, and that is why the fundamental is the word embedding in this task. Typically, each word or token is represented by a vector. Therefore, basic embedding is one-hot; thus, a dictionary of words is generated, then each word is represented by a vector so that only one cell’s value is one and others are zero, and vector size equals to the dictionary size. In this representation, vectors are orthogonal, so there is no semantics and relation between the words. Moreover, in large corpuses, vector size gets large and needs storage to save and handle it.

The problems and disabilities of one-hot vector generate a felt need for new embeddings. Word2Vec [41, 42] was the first word embedding able to map words to vectors considering the semantic. Word2Vec became the cornerstone of other embedding techniques with their facilities. FastText [43] and GloVe [44] are the
| LIWC dimension          | Output label | LIWC2015 mean | LIWC2007 mean | LIWC 2015/2007 correlation |
|-------------------------|--------------|---------------|---------------|-----------------------------|
| **Summary variable**    |              |               |               |                             |
| Word count              | WC           | 11,921.82     | 11,852.99     | 1.00                        |
| **Analytical thinking** |              |               |               |                             |
| Analytic                | Analytic     | 56.34         |               |                             |
| Clout                   | Clout        | 57.95         |               |                             |
| Authentic               | Authentic    | 49.17         |               |                             |
| **Emotional tone**      |              |               |               |                             |
| Tone                    | Tone         | 54.22         |               |                             |
| **Language metrics**    |              |               |               |                             |
| Words per sentence      | WPS          | 17.40         | 25.07         | 0.74                        |
| Words > 6 letters       | Sixltr       | 15.60         | 15.89         | 0.98                        |
| Dictionary words        | Dic          | 85.18         | 83.95         | 0.94                        |
| **Function words**      |              |               |               |                             |
| Function Words          | function     | 51.87         | 54.29         | 0.95                        |
| Total pronouns          | pronoun      | 15.22         | 14.99         | 0.99                        |
| Personal pronouns       | pppron       | 9.95          | 9.83          | 0.99                        |
| 1st pers singular       | i            | 4.99          | 4.97          | 1.00                        |
| 1st pers plural         | we           | 0.72          | 0.72          | 1.00                        |
| 2nd person              | you          | 1.70          | 1.61          | 0.98                        |
| 3rd pers singular       | shehe        | 1.88          | 1.87          | 1.00                        |
| 3rd pers plural         | they         | 0.66          | 0.66          | 0.99                        |
| Impersonal pronouns     | ipron        | 5.26          | 5.17          | 0.99                        |
| Articles                | article      | 6.51          | 6.53          | 0.99                        |
| Prepositions            | prep         | 12.93         | 12.59         | 0.96                        |
| Auxiliary verbs         | auxverb      | 8.53          | 8.82          | 0.96                        |
| Common adverbs          | adverb       | 5.27          | 4.83          | 0.97                        |
| Conjunctions            | conj         | 5.90          | 5.87          | 0.99                        |
| Negations               | negate       | 1.66          | 1.72          | 0.96                        |
| **Grammar other**       |              |               |               |                             |
| Regular verbs           | verb         | 16.44         | 15.26         | 0.72                        |
| Adjectives              | adj          | 4.49          |               |                             |
| Comparatives            | compare      | 2.23          |               |                             |
| Interrogatives          | interrog     | 1.61          |               |                             |
| Numbers                 | number       | 2.12          | 1.98          | 0.98                        |
| Quantifiers             | quant        | 2.02          | 2.48          | 0.88                        |
| **Affect words**        |              |               |               |                             |
| Affect Words            | affect       | 5.57          | 5.63          | 0.96                        |
| Positive emotion        | posemo       | 3.67          | 3.75          | 0.96                        |
| Negative emotion        | negemo       | 1.84          | 1.83          | 0.96                        |
| Anxiety                 | anx          | 0.31          | 0.33          | 0.94                        |
| Anger                   | anger        | 0.54          | 0.60          | 0.97                        |
| Sadness                 | sad          | 0.41          | 0.39          | 0.92                        |
Table 2  (continued)

| LIWC dimension | Output label | LIWC2015 mean | LIWC2007 mean | LIWC 2015/2007 correlation |
|----------------|--------------|----------------|----------------|-----------------------------|
| **Social words** |              |                |                |                             |
| Social words   | social       | 9.74           | 9.36           | 0.96                        |
| Family         | family       | 0.44           | 0.38           | 0.94                        |
| Friends        | friend       | 0.36           | 0.23           | 0.78                        |
| Female referents| female     | 0.98           |                |                             |
| Male referents  | male         | 1.65           |                |                             |
| **Cognitive processes** |          |                |                |                             |
| Cognitive processes | cogproc | 10.61          | 14.99          | 0.84                        |
| Insight        | insight      | 2.16           | 2.13           | 0.98                        |
| Cause          | cause        | 1.40           | 1.41           | 0.97                        |
| Discrepancies  | discrep      | 1.44           | 1.45           | 0.99                        |
| Tentativeness  | tent         | 2.52           | 2.42           | 0.98                        |
| Certainty      | certain      | 1.35           | 1.27           | 0.92                        |
| Differentiation| differ       | 2.99           | 2.48           | 0.85                        |
| **Perpetual processes** |          |                |                |                             |
| Perpetual processes | percept | 2.70           | 2.36           | 0.92                        |
| Seeing         | see          | 1.08           | 0.87           | 0.88                        |
| Hearing        | hear         | 0.83           | 0.73           | 0.94                        |
| Feeling        | feel         | 0.64           | 0.62           | 0.92                        |
| **Biological processes** |          |                |                |                             |
| Biological processes | bio    | 2.03           | 1.88           | 0.94                        |
| Body           | body         | 0.69           | 0.68           | 0.96                        |
| Health/illness | health       | 0.59           | 0.53           | 0.87                        |
| Sexuality      | sexual       | 0.13           | 0.28           | 0.76                        |
| Ingesting      | ingest       | 0.57           | 0.46           | 0.94                        |
| **Core drives and needs** |          |                |                |                             |
| Core drives and needs | drives   | 6.93           |                |                             |
| Affiliation    | affiliation  | 2.05           |                |                             |
| Achievement    | achieve      | 1.30           | 1.56           | 0.93                        |
| Power          | power        | 2.35           |                |                             |
| Reward focus   | reward       | 1.46           |                |                             |
| Risk/prevention focus | risk | 0.47           |                |                             |
| **Time orientation** |          |                |                |                             |
| Past focus     | focuspast    | 4.64           | 4.14           | 0.97                        |
| Present focus  | focuspresent | 9.96           | 8.10           | 0.92                        |
| Future focus   | focusfuture  | 1.42           | 1.00           | 0.63                        |
| **Relativity** |              |                |                |                             |
| Relativity     | relativ      | 14.26          | 13.87          | 0.98                        |
| Motion         | motion       | 2.15           | 2.06           | 0.93                        |
| Space          | space        | 6.89           | 6.17           | 0.96                        |
| Time           | time         | 5.46           | 5.79           | 0.94                        |
examples of evolution in word embedding techniques. All of the embeddings should be trained, so there are some pre-trained language models (PLMs) with differences in trained database and attitude of training (CBOW and skip-grams).

Transformers [45] introduced in 2018 made a revolution in embedding techniques, thus making more parallelism possible than other architectures (such as CNNs and RNNs). Hence, computers were able to train larger models, since large-scale Transformer-based PLMs appeared. The most well-known Transformer-based PLM could be named BERT [47]. Based on BERT, numerous models arise by a different point of view, namely RoBERTa [48] (robust and larger), ALBERT [49]

Table 2 (continued)

| LIWC dimension | Output label | LIWC2015 mean | LIWC2007 mean | LIWC 2015/2007 correlation |
|----------------|--------------|----------------|---------------|----------------------------|
| **Personal concerns** |              |                |               |                            |
| Work           | work         | 2.56           | 2.27          | 0.97                       |
| Leisure        | leisure      | 1.35           | 1.37          | 0.95                       |
| Home           | home         | 0.55           | 0.56          | 0.99                       |
| Money          | money        | 0.68           | 0.70          | 0.97                       |
| Religion       | relig        | 0.28           | 0.32          | 0.96                       |
| Death          | death        | 0.16           | 0.16          | 0.96                       |
| **Informal speech** |            |                |               |                            |
| Informal speech informal | informal | 2.52           |               |                            |
| Swear words    | swear        | 0.21           | 0.17          | 0.89                       |
| Netspeak       | netspeak     | 0.97           |               |                            |
| Assent         | assent       | 0.95           | 1.11          | 0.68                       |
| Nonfluencies   | nonfl        | 0.54           | 0.30          | 0.84                       |
| Fillers        | filler       | 0.11           | 0.40          | 0.29                       |
| **All punctuation** |           |                |               |                            |
| Periods        | Period       | 7.46           | 7.91          | 0.98                       |
| Commas         | Comma        | 4.73           | 4.81          | 0.98                       |
| Colons         | Colon        | 0.63           | 0.63          | 1.00                       |
| Semicolons     | SemiC        | 0.30           | 0.24          | 0.98                       |
| Question marks | QMark        | 0.58           | 0.95          | 1.00                       |
| Exclamation marks | Exclam    | 1.0            | 0.91          | 1.00                       |
| Dashes         | Dash         | 1.19           | 1.38          | 0.98                       |
| Quotation marks | Quote       | 1.19           | 1.38          | 0.76                       |
| Apostrophes    | Apostro      | 2.13           | 2.83          | 0.76                       |
| Parentheses (pairs) | Parenth  | 0.52           | 0.25          | 0.90                       |
| Other punctuation | OtherP    | 0.72           | 1.38          | 0.98                       |

1 More information in [46].
(high-speed training and lower memory), and DistilBERT [50] (40% smaller and 60% faster).

Methods (overview)

In natural language processing, representation of input is the most important component, and the smallest part of a text is words. In recent years, novel representations called pre-trained word embeddings have been trending and making revolutions in text mining. APP is not unaffected, and novel methods based on word embedding techniques have appeared. In reviewing APPs, at first methods, without pre-trained word embedding techniques describe, second, pre-trained word embedding-based methods detail, and at the end, methods with more than text input introduce.

PLM-free APPs

[51] proposed a combinational algorithm for detecting personality using LIWC and MRC on textual features. 81 LIWC features plus 26 MRC features were extracted from the Essays dataset. The proposed EmoSenticSpace was a novel representation method on a graph of EmoSenticNet [52] and fed to a blending algorithm. The output is a 100-dimensional vector and by “bag of concept”, averaging was done to the vectors to represent a text. The authors trained five Sequential Minimal Optimization (SMO) classifiers, one for each of the Big Five traits.

[53] proposed an ensemble model for recognizing personality from Facebook. The authors trained three SVM classifiers, the first one trained on Facebook. The second is trained on the Essays dataset, and as the meta-classifier, the output of two classifiers is fed to the third SVM classifier.

Part of speech (POS) is a basic feature of texts. In [54], the authors used POS and POS n-grams of texts beside bag of words, word sentiment, negations, and vocabulary size features to predicting personality. As a classifier, SVM is deployed in two formats: 2-class and 3-class. 2588 university students essays were collected on the Big Five personality scale to evaluate the method.

LIWC consists of 19 features, and [55] tried to reduce feature size for the purpose of achieving better results on the Essays. In this way, Information Gain and Principal Component Analysis (PCA) feature reduction techniques have been examined. It is proved that some LIWC features do not have adequate information on Essays for APP.

The early word embedding models, e.g., Word2Vec, have some common practical problems that make using them hard and somehow restricted. The first one is unseen words; if a word does not see in the learning stage, make the model in trouble. The second problem occurs once you want to train a model, and there are too many parameters. Due to these problems, [56] proposed an embedding model to embed personality texts. C2W2S4PT (Character to Word to Sentence for Personality Traits) is a three-stage Bi-RNN-based model; first, characters modeling; second, word modeling based on the first stage; third, sentence embedding using words’ representation.
using feedforward neural network. In Fig. 4, the proposed architecture is illustrated. In [57], the proposed C2W2S4PT is evaluated in English, Spanish and Italian languages to proving language independency of C2W2S4PT.

In [58] research, to take advantage of huge unlabeled data, Pseudo Multi-view Co-training (PMC) [59] is adopted, an effective Semi-supervised learning algorithm, to build a personality prediction model. To extract adequate linguistic features, both LIWC and n-gram, along with the Word2Vec word embedding technique, are trained on myPersonality dataset to predict personality through textual data. Figure 8 illustrates the overall framework of method.

Personality2Vec [60] is the name of a user-personality embedding technique through which a user has generated texts. Semantic and linguistic features of texts construct a graph, and a biased walk strategy has been proposed to divide users into groups with maximum similar personality users in the same group. As shown in Fig. 10, in the first, linguistic and semantic features are extracted in order to pass into the learning part. Linguistic features are 103-dimension of LIWC and 10-dimension of special linguistic features, which have been proposed by the authors.

Paper [61] deployed network representation learning (NRL) as the novelty and for the first time in APP. AdaWalk generates the graph of documents on personality datasets in two approaches: classification and regression. NRL presented node (words or token) by using skip-gram.

Another graph-based approach called personality graph convolutional networks (personality GCN) was introduced by [62]. The authors aimed to create a graph to model users, documents, and words by the core of the co-occurrence of words in a document. The weight of edges is calculated by TF-IDF for a document–word edge and pointwise mutual information (PMI) for a word–word edge. The classification layer is the last one, five classifiers for Big Five traits. Figure 9 illustrates an overview of the three layers of GCN. It is worthwhile to say that words, users, and documents are represented by a one-hot vector.

One of the most recent research in this area proposed five new APP methods: term frequency vector-based, ontology-based, enriched ontology-based, latent semantic analysis (LSA)-based, and deep learning-based (BiLSTM), by a contribution of enhancing accuracy [18]. These five introduced models are used as the base to hierarchical attention network (HAN) ensemble model. The authors evaluated the methods on the Essays dataset and achieved enhancing accuracy on the ensemble method. The architecture of the proposed HAN stacking model is shown in Fig. 1.

Application of knowledgeable representation of the input text in a graph structure (namely knowledge graph) is a powerful technique in artificial intelligence, and Ramezani et al. [63] deployed this technique in APP for the first time. The proposed knowledge graph is a set of interlinked descriptions of concepts that were built by matching the input text’s concepts with DBpedia knowledge base entries. The graph was enriched with the DBpedia ontology, NRC Emotion Intensity Lexicon, and MRC psycholinguistic database to achieve a more robust knowledge representation (Fig. 14 shows the proposed architecture and deployment of knowledge graph). The knowledge graph, which served as a knowledgeable alternative to the input text, was then embedded to produce an embedding matrix. Finally, the resultant embedding matrix was given to four different deep
learning models, including convolutional neural networks (CNN), basic recurrent neural networks (RNN), long short-term memory (LSTM), and bidirectional long short-term memory (bidirectional LSTM) (BiLSTM) (Fig. 2).

Ramezani et al. [64] in the follow of knowledge graph-based approaches deployed KGrAt-Net, which is a knowledge graph attention network text classifier, for the first time to perform automatic personality prediction (APP). As Fig. 15 shows, the proposed approach consists of three stages: pre-processing, knowledge representation, and KGrAt-Net classifier.

Fig. 1 The architecture of a hierarchical attention network (HAN) combined with base methods’ predictions (ensemble) to carry out stacking in document level classification proposed by [18]
PLM-based APPs

The first research that deployed a PLM to APP is done by [65] based on Word2Vec. This research deployed two approaches, document level using Mairesse features and word level using Word2Vec and CNN model to achieve 300-dimension representations of words to model sentences and documents, respectively, by n-gram sight of view. Both approaches’ representations concatenated to be fed to a fully connected layer for classification (Fig. 3 illustrates the architecture of the proposed method). Five models, one for each of five traits, trained on the Essays dataset.

2CLSTM is the name of the model proposed by [66] that tries to learn structural features based on latent sentence group (LSG). At the first step, each word embedded into a 100-dimension vector through GloVe pre-trained model. 2LSTM encodes the vectors into sentences that pass to the next section. The following sections have to model relationships of sentences, and LSG does this. Latent Sentence Group (LSG) is defined as a synthesis that consists of a number of sentence vectors which are closely connected in some coordinates. To do this, CNN networks deployed to learn 1, 2, 3-grams. Figure 7 illustrates the layers and schema of 2CLSTM. A dense layer and max-pooling layer follow immediately to generate the final vector to pass into a classifier which there was softmax.

In [67], a personality prediction method using sentiment of short texts is introduced by naming SENTIPEDE. The aim was to determine the personality in Big Five using textual features with sentiment features of short texts in which the base was Twitter. The sentiment of text is computed by Valence Aware Dictionary and Sentiment Reasoner (VADER) [68], a rule-based framework for sentiment classification in English, to gain a label from positive, negative, or nonpartisan. Then, GloVe word embedding is deployed to vectorize words to pass to a CNN-LSTM.
model along with the sentiment labels. The case study of the paper was Uber and related tweets.

In introducing a new dataset, some base methods and evaluation of them by a standard dataset are necessary. FriendPersona is the name of a new dataset introduced by [69]. In [69], five models have been developed, whose names are: ABCNN (CNN with attention mechanism), ABLSTM (Bidirectional LSTM attention mechanism), HAN (Hierarchical Attention Network), BERT, and RoBERTa. BERT and RoBERTa as a PLM are fine-tuned on both datasets, Essays and FriendPersona.

In paper [70], two approaches have been studied: personality prediction based on psycholinguistic features and language model features. In both approaches, features extracted from texts are fed to a classifier, SVM and MLP, to classify texts into personality traits. Mairesse, SenticNet [71], NRC Emotion Lexicon [72], VAD Lexicon
[73], and Readability\(^2\) are the extracted psychological features. On the other hand, BERT PLM is deployed as a language model features’ extractor.

[74] segmented documents into 250 tokens’ length sub-documents in order to feed to BERT(base) PLM. Based on the Essays dataset used in this paper, documents length is 650 tokens on average, so 4 layers of [CLS] concatenated with 84 Mairesse features as features of a document. SVM is the classifier of the proposed method, which is trained on sub-documents in parallel mode like a bagged classifier, and at last, the final trait is predicted by majority voting. Figure 11 shows the authors proposed method.

One of the recent APP methods combined semantic and emotional features in order to determine personality trait from multi-text [75]. On the semantic side, BERT is deployed to vectorize texts, and using a self-attention mechanism, sentence-level representation is generated. On the other side, SenticNet5 [71] has extracted the sentiment of the sentences to map to a vector. Both vectors are concatenated and fed to a classification network. CNN, GRU and LSTM are different neural networks that trained to label personality trait.

Semantic-enhanced personality recognition neural network (SEPRNN) [20] is proposed with the goal of avoiding dependency to feature selection in APP and modeling semantic from word-level representations. GloVe PLM is deployed to vectorize words, then a BiGRU model learned to extract left and right contexts of words, but since semantics did not consider. To capture the higher level of semantic representations from contexted textual data, vectors are fed to a fully connected network to text modeling of documents. In the end, a fully connected network with sigmoid activation function is adopted to learn a two-dimensional vector for binary classification of personality. An illustration of the proposed method is displayed in Fig. 12.

Transformer-MD (multi-document Transformer) is the name of the method proposed by [76]. The core of the proposed method is to put together information in multiple posts to represent an overall personality for each user. The authors tried to solve two problems: post-orders bias into posts on personality and individually posts processing of a person to personality detection. In order to this, encoding each post by Transformer-MD allows access to information in the other posts of a user through Transformer-XL’s memory tokens that share the position embedding. For the cases of multi-trait personality detection, a dimension attention mechanism on top of Transformer-MD was set. The overview of the proposed methods is shown in Fig. 13.

Multimodal APPs

One of the first deep learning used methods in multimodal approaches deployed text in couple with the authors information to achieve the writer’s personality [77].

\(^2\) A number of calculated readability measures based on simple surface characteristics of the text. These measures are basically linear regressions based on the number of words, syllables, and sentences.
According to the limitations and obstacles of pre-trained word embedding methods in 2017, Word2Vec, the authors trained their word embedding model based on skip-gram using the myPersonality dataset. In the trained model, modeling words position was not taken into account, so applying N-grams seemed the best approach. In order to apply this approach, two approaches were taken, CNN-based and Bi-RNN-based. Figure 5 shows the architecture of [77]. After the word modeling part, the author’s information is 7-D vector concatenate and goes on. Each of the five personality traits has been trained in a neural network.

[78] explored the use of machine learning techniques for inferring a user’s personality traits from their Facebook status updates. Four kinds of numeric features are:

- LIWC features
- Social Network features: 7 features related to the social network of the user: (1) network size, (2) betweenness, (3) nbetweenness, (4) density, (5) brokerage, (6) nbrokerage, and (7) transitivity.
- Time-related features: 6 features related to the time of the status updates (we assume that all the times are based on one time zone): (1) frequency of status updates per day, (2) number of statuses posted between 6 and 11 am, (3) number of statuses posted between 11 and 16, (4) number of statuses posted between 16 and 21, (5) number of statuses posted between 21 and 00, and (6) number of statuses posted between 00 and 6 am.
- Other features: 6 features not yet included in the categories above: (1) total number of statuses per user, (2) number of capitalized words, (3) number of capital letters, (4) number of words that are used more than once, (5) number of urls, and (6) number of occurrences of the string PROPNAME—a string used in the data to replace proper names of persons for anonymization purposes.

Different concatenation of features is analyzed with 3 classification algorithms named SVM, KNN, and Naive Bayes.

[79] proposed a comprehensive view to APP multimodal approach that accompanied texts, avatars, and emojis. Pearson correlation, Text-CNN, and Bag-of-Word (BOW) clusters are the textual-based features extracted from the Weibo tweets collected in the research. Pearson correlation computed between words and the personality traits to selecting top 2000 words are strongly correlated to personality. Due to the LIWC limited capability in representing users’ linguistic patterns in short and informal texts, 1500 Chinese words and all the punctuations in the bag-of-words format were clustered using the k-means algorithm, and then the count of the number of items within each cluster was used instead of LIWC. As the last feature, a

| Table 3 Distribution of the Essays dataset labels | Label | Opn | Con | Ext | Agr | Neu |
|-----------------------------------------------|-------|-----|-----|-----|-----|-----|
| Yes                                           | 1271  | 1253| 1276| 1310| 1233|
| No                                            | 1196  | 1214| 1191| 1157| 1234|
convolutional architecture called Text-CNN was trained to model words in vector form in reference to [80] model. The structure of the proposed algorithm is shown in Fig. 6. As seen, each type of inputs lasts for a classifier (Logistic Regression) to specify the trait. As the final step, a stacked ensemble algorithm, generalization-based ensemble method, attached to the classifiers as the final result classifier.

It is common to utilize acoustical features simultaneously with transcripts of speech to achieve the personality of people. In [81], four features were demonstrated in order to predict personality, namely acoustic-prosodic low-level descriptor features (LLD), LIWC, Dictionary of Affect in Language (DAL), and word embeddings. DAL features were extracted using Whissell’s Dictionary of Affect in Language (DAL), and 19 features were extracted in this research. In the word embedding part, Google’s pre-trained skip-gram vectors and Stanford’s pre-trained GloVe has been used. Two approaches have been adopted for modeling documents based on embedding vectors, averaging and LSTM neural network. One strange thing in the proposed method is that both of PLMs vectors are fed to the model and concatenated with three other features. Moreover, two strategies are proposed by the authors, first, concatenate features and then five fully connected layers end with five neurons as the classifier; second, each of five features is fed to a three fully connected layers’ block before concatenating and then fed to five neurons for classification similarly.

| Table 4 | The elements count of Essays dataset |
|---------|--------------------------------------|
|         | Sentences   | Characters   | Spaces      | Words       | Punctuations | Stop words |
| Mean    | 46.6250     | 3296.9059   | 702.8232    | 668.6643    | 53.8293     | 318.8873    |
| Std     | 26.7925     | 1287.5402   | 286.0818    | 262.4843    | 31.6477     | 134.2972    |
| Min     | 1           | 159         | 34          | 35          | 0           | 0           |
| 25%     | 28          | 2407.5      | 506         | 482.5       | 32          | 227         |
| 50%     | 43          | 3165        | 676         | 646         | 49          | 302         |
| 75%     | 61          | 4081        | 867.5       | 828         | 69          | 396         |
| Max     | 307         | 12855       | 3859        | 2631        | 348         | 1340        |

| Table 5 | Distribution of labels in myPersonality classificational mode [62] |
|---------|---------------------------------------------------------------|
| Label   | Opn | Con | Ext | Agr | Neu |
| Yes     | 176 | 130 | 96  | 134 | 99  |
| No      | 74  | 120 | 154 | 116 | 151 |

| Table 6 | The distribution of myPersonality dataset labels in regressional mode [67] |
|---------|---------------------------------------------------------------|
|         | Opn | Con | Ext | Agr | Emo |
| Max     | 5   | 5   | 5   | 5   | 4.75|
| Min     | 2.25| 1.45| 1.33| 1.65| 1.25|
| Avg     | 4.0786 | 3.5229 | 3.2921 | 3.6003 | 2.6272|
| Std     | 0.5751 | 0.7402 | 0.8614 | 0.5708 | 0.7768|
Evaluating methods

Every proposed method should be evaluated to prove its performance. In APP, evaluations consist of two parts; dataset is the first and metric of assessment is the second part. Five datasets are available in text-based APP, and metrics vary on the concept of evaluation. This section consists of two parts: part one details the datasets and part two defines results of methods in datasets and metrics.

Datasets

Each method should evaluate and compare it with other methods, and this requires a fair condition. To achieve fair comparison, ground truths should be same including metrics and datasets. In this part, five datasets, Essays, myPersonality, YouTube, FriendPersona, and Kaggle MBTI, are benchmarked in which text-based APPs are introduced.

Essays

Essays (also called stream-of-consciousness essays) is the first and most cited text dataset in automatic personality prediction. The dataset introduced by [35] that consist of 2468 anonymous essays in English annotated in Big Five scale. The dataset annotated in two modes: classification and regression. Thus, each essay has two Big
Five values: first, each trait has a binary value; second, each trait is real value in the aim of regression. The essay mainly is deployed in classification purpose and Table 3 shows the number of essays in each trait. It should be notated that one row of data was an error and dismissed in the table’s values. Moreover, the distribution of essay’s components is reported in Table 4.

**myPersonality**

myPersonality\(^3\) is a collection of 250 anonymous Facebook users’ profile updates scored in Big Five by questioning users to answer them. Since 2018 creators, Stillwell and Kosinski stopped sharing and developing the dataset. There are some versions available on the internet that do not match the records but approximately they contained 9900 records. The myPersonality annotated in two forms, classification and regression, and Tables 5 and 6 illustrate the distribution of status updates of myPersonality in each form.

**YouTube**

YouTube is the most popular video-sharing platform and has attracted many vloggers. YouTube dataset is a collection of 404 YouTube vloggers personality scores in Big Five. The dataset is recorded talking in front of webcam about a variety of topics and annotated using Amazon Mechanical Turk and the Ten-Item Personality Inventory (TIPI). As mentioned, YouTube dataset is a multimodal, video, speech, and transcript to text [82, 83]. Based on the aim of the article, speech transcripts to text [83] is analyzed in this section. As the routine, distribution of traits has been shown in Table 7 and textual elements details are shown in Table 8.

**FriendPersona**

The newest dataset introduced in the context of text-based APP is FriendPersona. This is developed on Friends TV Show Dataset\(^4\) [84] and 711 conversations were extracted. FriendPersona was annotated by three experts and to make it binary, split from the median. The dataset could be found in github\(^5\).

**Kaggle MBTI**

Kaggle MBTI\(^6\) has gathered 50 last posts through the PersonalityCafe forum in MBTI personality model. There are 8675 rows of data in which each row represents a person. The dataset is started on cognitive functions by Carl Jung, and finally, personality tags are done by Jungian Typology in MBTI.

---

\(^3\) [https://sites.google.com/michalkosinski.com/mypersonality](https://sites.google.com/michalkosinski.com/mypersonality)

\(^4\) [https://github.com/emorynlp/character-mining](https://github.com/emorynlp/character-mining)

\(^5\) [https://github.com/emorynlp/personality-detection](https://github.com/emorynlp/personality-detection)

\(^6\) Available on [https://www.kaggle.com/datasnaek/mbti-type/](https://www.kaggle.com/datasnaek/mbti-type/)
Results

In this part, first evaluation metrics used in APP are introduced and then reported results of methods appeared in the division of datasets. Each following tables survey a dataset presented in the previous part.

Evaluation metrics

Precision, recall, accuracy, and F-measure are well-known classification evaluation metrics that are used in scientific reports. For calculating these measures, four concepts should be defined. TP, TN, FP, and FN are the notions and denote to true positive, true negative, false positive, and false negative, respectively. These concepts make sense based on ground truth in confronting with the output of the system. There are other measurements for evaluating classification (Regression) methods such as RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and Coefficient of Determination ($R^2$). Still, most articles preferred the binary classification measures and reported in four first measures. The following equations are the calculation of the metrics:

\[
\text{Precision} = \frac{\# \text{system predicted true label}}{\# \text{items system predicted as true label}} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)
\]

\[
\text{Recall} = \frac{\# \text{system predicted true label}}{\# \text{ground truth true items}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)
\]

| Method | F-measure | Accuracy |
|--------|-----------|----------|
|        | Opn | Con | Ext | Agr | Neu | Avg | Opn | Con | Ext | Agr | Neu | Avg |
| [20]   | 67.84 | 63.46 | 71.50 | 71.92 | 62.36 | 67.416 | 63.16 | 57.49 | 58.91 | 57.49 | 59.51 | 59.312 |
| [75]   | 80.35 | 80.23 | 79.94 | 80.30 | 80.14 | 80.192 |
| [18]   | 57.37 | 59.74 | 65.80 | 61.62 | 60.69 | 61.04 | 56.30 | 59.18 | 64.25 | 60.31 | 61.14 | 60.24 |
| [62]   | 67 | 68 | 67 | 69 | 69 | 68 | 64.80 | 59.10 | 60 | 57.70 | 63 | 60.92 |
| [69]   | 65.86 | 58.55 | 60.62 | 59.72 | 61.04 | 61.158 |
| [70]   | 64.6 | 59.2 | 60 | 58.8 | 60.5 | 60.62 |
| [74]   | 62.09 | 57.84 | 59.30 | 56.52 | 59.39 | 59.028 |
| [65]   | 62.68 | 57.30 | 58.09 | 56.71 | 59.38 | 58.832 |
| [55]   | 56 | 54 | 53 | 50 | 54 | 53.4 |
| [53]   | 66.1 | 63.3 | 63.4 | 61.5 | 63.7 | 63.6 |
| [51]   | 60.57 | 56.46 | 56.28 | 53.9 | 58.15 | 57.072 |

The boldface values indicate the best values in each trait.
Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)

F – measure = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)

Table 10  Experimental results on myPersonality dataset

| Method | F-measure | Accuracy |
|--------|-----------|----------|
|        | Opn Con Ext Agr Neu Avg | Opn Con Ext Agr Neu Avg |
| [62]   | 76.8 75 85 70 79 77.16 | 80 76 80 68 79 76.6 |
| [58]   | 65 62 71 68 70 67.2 |
| [88]   | 65 |
| [78]   | 61 54 56 45 49 53 |

The boldface values indicate the best values in each trait

Table 11  Experimental results on YouTube dataset

| Method | F-measure | RMSE |
|--------|-----------|------|
|        | Opn Con Ext Agr Emo | Opn Con Ext Agr Emo |
| [20]   | 78.57 77.20 82.35 83.08 79.55 | 0.6813 0.6898 0.8897 0.7743 0.6872 |
| [61]   | |

Table 12  Experimental results on FriendPersona dataset

| Method | F-measure | Accuracy |
|--------|-----------|----------|
|        | Opn Con Ext Agr Neu | Opn Con Ext Agr Neu |
| [69]   | 68.47 56.78 64.21 65.58 60.06 |

Table 13  Experimental results on Kaggle MBTI dataset

| Method | F-measure | Accuracy |
|--------|-----------|----------|
|        | I-E S-N F-T J-P | I-E S-N F-T J-P |
| [89]   | 98.56 99.75 94.72 96.19 | 99.37 99.92 94.55 95.53 |
| [76]   | 66.08 69.10 79.19 67.50 |
| [90]   | 78.17 86.06 71.78 65.70 |
| [70]   | 78.8 86.3 76.1 67.2 |

The boldface values indicate the best values in each trait
F1 score (Eq. 4) translates to harmonic mean of precision (Eq. 1) and recall (Eq. 2) and that is why most studies report F1 score to make readers understand easier instead of reporting precision and recall:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=0}^{N-1} (y_t - \hat{y}_t)^2}{N}}$$  

(5)

$$\text{MAE} = \sum_{t=0}^{N-1} |(y_t - \hat{y}_t)|$$  

(6)

As respect of lack of text-based personality prediction studies, researchers reported the results on their aspect without much comparison, and this makes variety in results’ measurement units. RMSE (Eq. 5) and MAE (Eq. 6) have been used in [85] and [86], however, cannot be compared cause of non-comparable studies in same dataset.

**Discussion**

For evaluating and making results comparative and comprehensive, the results of each dataset are given in separate tables. The Essays is the most popular dataset in APP, and the number of results is more than the others. According to the classification essence of Essays and myPersonality datasets, the evaluations are done by F-measure and accuracy metrics. Augmentation of APP methods have been accelerated more rapidly with the advent of deep learning, and the increase in the quality of results is evident in Essays datasets more obviously. As it is shown in Table 9, the results of each method that powered the newer PLMs and novel deep learning method achieved higher values compared with the previous method. The point that has to be considered is that some methods reported results insufficiently; thus, solely one evaluation metric is listed by personal opinion. Among the methods that do not use word embeddings and evaluated on the Essays dataset, [18] got better results. Since the myPersonality dataset is not available by the creators anymore, it is not used in recent researches. However, [62] achieved the best results, as it is shown in Table 10, which does not apply PLMs for embedding (Table 11).

As mentioned in “Datasets”, FriendPersona is the most recent APP dataset that was only introduced for the proposed method in [69] and is shown in Table 12. But in comparison with reported results on Essays (Table 9), the evaluations are acceptable because the proposed model achieved in range of 60% (63.01% on FriendPersona and 61.158% on Essays).

The evaluations on the Kaggle dataset are shown in Table 13 and done using F-measure and accuracy metrics. The best-performed method is [89] that deployed XGboost ensemble algorithm to gain high values on both metrics. To the rest methods, three reported accuracies and one reported F-measure only that does not have any subscription, so methods with the same reported metrics should be compared.
In the end, as it can be concluded from the tables, deployment of PLMs for words and documents embeddings in couple with deep neural network models makes a significant improvement in textual-based APPs. However, somehow, it seems that the hybrid and ensemble models (especially PLM-free) are achieving better results in their own class and could be a progressive area of research. Also, the APP research area suffers from a lack of golden standard datasets and evaluations. As the final recommendation for future work, introducing novel datasets labeled on more than one personality trait model, with more samples and different lengths of documents, can make the APP research area more attractive and competitive.

Conclusion

Automatic personality prediction (perception) (APP) system provides an opportunity to predict personality traits based on human behavior manifestations, especially in texts in this review. This paper is reviewed text-based APP methods since 2010 and reported results from five well-known benchmark datasets. To do so, articles have been categorized into three PLM-free approaches, PLM-based methods, and multimodal techniques. Also, the framework of overviewed methods has been collected. The aim of this review is to give a general overview of the steps of getting meliorated of APPs to researchers in this field. This review solely looks at existing personality trait identification methods from a computational standpoint and ignores psychological studies on personality trait recognition. Finally, we introduced several open issues that outline promising directions for future research on text-based automatic personality prediction:

- As psychologists recognize that individuals personality traits change through changes in their speaking/writing, similarly, when the appeared words and expressions in the given input text change, the machine assigns different labels for traits. Therefore, the dataset, in fact, is the basis of the APP system, because a computer is learning from the dataset. All things considered generating new datasets labeled by psychologists through essays and other features would be an excellent benchmark for evaluating personality prediction methods. Meaningfully, generate a dominant dataset and benchmark, so that the results are based on convincing a psychologist and learning a computer.
- Making available datasets in several personality trait models simultaneously. This would make results more comprehensive and represent APP systems output psychologistic.
- Generating datasets in low-resource languages. Since languages characterize different features of people, different languages would make various personality-related features through texts.

A Appendix

See Table 14 and Figs. 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15.
Table 14  An overview of Cattell’s 16 [91]

| Description of LOW values - $\infty \leftarrow$ | Personality trait $\rightarrow +\infty$ | Description of HIGH values |
|-----------------------------------------------|--------------------------------------|---------------------------|
| **Warmth (A)**                                |                                     |                           |
| Impersonal                                    | Distant                             | Warm                      |
| Cool                                          | Reserved                             | Attentive to others       |
| Detached                                      | Formal                               | Easygoing                 |
| Aloof                                         |                                      | Likes people              |
| **Reasoning (B)**                             |                                     |                           |
| Concrete-thinking                             | Less intelligent                    | Abstract-thinking         |
| Lower general mental capacity                 | Unable to handle abstract problems  | Higher general mental capacity |
| **Emotional stability (C)**                   |                                     |                           |
| Reactive emotionally                          | Changeable                           | Emotionally stable        |
| Affected by feelings                          | Emotionally less stable              | Faces reality calmly      |
| Easily upset                                  |                                      | Mature                    |
| **Dominance (E)**                             |                                     |                           |
| Deferential                                   | Cooperative                         | Dominant                  |
| Avoids conflict                               | Submissive                          | Assertive                 |
| Humble                                        | Obedient                            | Competitive               |
| Easily led                                    | Docile                              | Bossy                     |
| Accommodating                                 |                                      |                           |
| **Liveliness (F)**                            |                                     |                           |
| Serious                                       | Restrained                          | Lively                    |
| Prudent                                       | Taciturn                            | Spontaneous               |
| Introspective                                 | Silent                               | Happy-go-lucky            |
|                                              |                                      |                           |
Table 14 (continued)

| Description of LOW values - $\infty \leftarrow$ | Personality trait $\rightarrow +\infty$ | Description of HIGH values |
|-----------------------------------------------|----------------------------------------|-----------------------------|
| **Rule-consciousness (G)**                    | Expressive                             | Impulsive                   |
| Expedient                                     | Rule-conscious                         | Dutiful                     |
| Disregards rules                              | Conscientious                          | Conforming                  |
|                                               | Moralistic                             | Staid                       |
|                                               | Rule-bound                             |                             |
| **Social boldness (H)**                       |                                       |                             |
| Shy                                           | Socially bold                          | Venturesome                 |
| Timid                                         | Thick-skinned                          | Uninhibited                 |
| Intimidated                                   |                                        |                             |
| **Sensitivity (I)**                           |                                       |                             |
| Utilitarian                                   | Sensitive                              | Esthetic                    |
| Unsentimental                                 | Sentimental                            | Tender-minded               |
| Self-reliant                                  | Intuitive                              | Refined                     |
| Rough                                         |                                        |                             |
| **Vigilance (L)**                             |                                       |                             |
| Trusting                                      | Vigilant                               | Suspicious                  |
| Accepting                                     | Skeptical                              | Distrustful                 |
| Easy                                          | Oppositional                           |                             |
| **Abstractedness (M)**                        |                                       |                             |
| Grounded                                      | Abstract                               | Imaginative                 |
| Prosaic                                       | Absentminded                            | Impractical                 |
| Steady                                        | Absorbed in ideas                      |                             |
| **Privateness (N)**                           |                                       |                             |
| Forthright                                    | Private                                | discreet                    |
Table 14 (continued)

| Low Values (LOW) | Personality Trait | High Values (HIGH) |
|------------------|-------------------|-------------------|
| Artless          | Open              | Nondisclosing      |
| Guileless        | Naive             | Polished           |
| Unpretentious    | Involved          | Astute             |

**Apprehension (O)**
- Self-assured: Unworried
- Complacent: Secure
- Free of guilt: Confident
- Self-satisfied: Self-blaming

**Openness to change (Q1)**
- Traditional: Attached to familiar
- Conservative: Respecting traditional ideas

**Self-reliance (Q2)**
- Group-oriented: Affiliative
- A joiner and follower dependent: Resourceful

**Perfectionism (Q3)**
- Tolerates disorder: Unexacting
- Flexible: Undisciplined
- Lax: Self-conflict
- Impulsive: Careless of social rules
- Uncontrolled: Self-sentimental

**Description of LOW values - ∞ ←**
- Artless
- Guileless
- Unpretentious

**Description of HIGH values → +∞**
- Nondisclosing
- Polished
- Astute
- Shrewd
- Worldly
- Diplomatic
- Apprehensive
- Worried
- Insecure
- Self-doubting
- Guilt-prone
- Worrying
- Open to change
- Liberal
- Critical
- Freethinking
- Self-reliant
- Resourceful
- Individualistic
- Perfectionistic
- Compulsive
- Socially precise
- Control
- Exacting will power
- Self-disciplined
- Self-sentimental
Table 14 (continued)

| Description of LOW values - \( \infty \) ← | Personality trait \( \rightarrow +\infty \) | Description of HIGH values |
|-------------------------------------------|---------------------------------|---------------------------|
| Tension (Q4)                              |                                 |                           |
| Relaxed                                   | Placid                          | Tense                     | High-energy               |
| Tranquil                                  | Torpid                          | Impatient                 | Driven                    |
| Patient                                   | Composed low drive              | Frustrated                | Over-wrought              |
|                                           |                                 | Time-driven               |                           |
Fig. 4 The architecture of proposed model for representation of texts called C2W2S4PT (Character to Word to Sentence for Personality Traits). Dotted boxes indicate concatenation [56, 57]
Fig. 5  The proposed architectures in [77]

(a) Framework of Heterogeneous Information Ensemble (HIE)

(b) The structure of Text-CNN.

Fig. 6  The structure of proposed algorithm called HIE and text representation model [79]

Fig. 7  The architecture of proposed method in [66] called 2CLSTM
Fig. 8  Pseudo-multi-view co-training (PMC)-based framework for personality prediction [58]

Fig. 9  Three-layer architecture overview of proposed personality GCN [62]
Fig. 10  Overall framework of Personality2Vec [60]

Fig. 11  The architecture of Bagged SVM over BERT word embedding technique [74]

Fig. 12  SEPRNN (semantic-enhanced personality recognition neural network) [20]

Fig. 13  The schema of proposed method named Transformer-MD [76]
**Conflict of interest**

The authors declare that they have no conflict of interest.

**Data availability**

Not applicable.

**Acknowledgements**

This project is supported by a research grant from the University of Tabriz (number S/806).

**Funding**

This project is supported by a research grant of the University of Tabriz (number S/806).

**Declarations**

**Conflict of interest**

The authors declare that they have no conflict of interest.
Ethics approval and consent to participate  Not applicable.

Consent for publication  Not applicable.

References

1. Corr, P. J., & Matthews, G. (2009). *The Cambridge handbook of personality psychology*. Cambridge: Cambridge University Press.
2. Kazdin, A. E. (2000). *Encyclopedia of psychology*, volume 2. American Psychological Association Washington, DC. http://www.scielo.br/scielo.php?script=sci_nlinks&pid=S1516-4446201300040041600015&lng=en
3. Bergner, R. M. (2020). What is personality? Two myths and a definition. *New Ideas in Psychology*, 57, 100759. https://doi.org/10.1016/J.NEWIDEAPSYCH.2019.100759.
4. Peters, E., & Killcoyne, H. L. (2015). Psychology. Encyclopaedia Britannica.
5. Schultz, D. P., & Schultz, S. Ellen. (2016). *Theories of personality*. Cengage Learning, 11 Edn. https://books.google.com/books/about/Theories_of_Personality.html?id=dXccCgAAQB AJ
6. Ewen, R. B. (2010). An introduction to theories of personality. Psychology Press. https://www. routledge.com/An-Introduction-to-Theories-of-Personality-7th-Edition/Ewen/p/book/9781841697468
7. Eysenck, H. (1998). *Dimensions of personality*. Routledge. https://doi.org/10.4324/97800203793268.
8. Jothilakshmi, S., Sangeetha, J., & Brindha, R. (2017). Speech based automatic personality perception using spectral features. *International Journal of Speech Technology*, 20(1), 43–50. https://doi.org/10.1007/s10772-016-9390-0.
9. Su, M. H., Wu, C. H., Huang, K. Y., Hong, Q. B., & Wang, H. M. (2018). Personality trait perception from speech signals using multiresolution analysis and convolutional neural networks. In *Proceedings—9th Asia-Pacific signal and information processing association annual summit and conference, APSIPA ASC 2017* (vol. 2018-February, pp. 1532–1536). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/APSIPA.2017.8282287.
10. Gilpin, L. H., Olson, D. M. & Alrashed, T. (2018). Perception of speaker personality traits using speech signals. In *Conference on human factors in computing systems—Proceeedings* (vol. 2018-April, pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3170427.3188557.
11. Mohammadi, G., & Vinciarelli, A. (2012). Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing*, 3(3), 273–284. https://doi.org/10.1109/TAFFC.2012.5.
12. Sang, J., Zhang, H., & Xu, C. (2016). Visual BFI: An exploratory study for image-based personality test. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (vol. 9916 LNCS, pp. 95–106). Springer. https://doi.org/10.1007/978-3-319-48890-5_10.
13. Allen, M. S., & Walter, E. E. (2016). Personality and body image: A systematic review. *Body Image*, 19, 79–88. https://doi.org/10.1016/j.bodyim.2016.08.012.
14. Chaudhari, K., & Thakkar, A. (2019). Survey on handwriting-based personality trait identification. *Expert Systems with Applications*, 124, 282–308. https://doi.org/10.1016/j.eswa.2019.01.028.
15. Lokhande, V. R., & Gawali, B. W. (2017). Analysis of signature for the prediction of personality traits. In *Proceedings—1st international conference on intelligent systems and information management, ICISIM 2017* (vol. 2017-Janua, pp. 44–49). Institute of Electrical and Electronics Engineers Inc., nov. https://doi.org/10.1109/ICISIM.2017.8122145.
16. Kindiroglu, A. A., Akarun, L., & Aran, O. (2017). Multi-domain and multi-task prediction of extraversion and leadership from meeting videos. *Eurasip Journal on Image and Video Processing*, 1, 1–14. https://doi.org/10.1186/s13640-017-0224-z.
17. Aslan, S., & Güdükbay, U. (2019). Multimodal video-based apparent personality recognition using long short-term memory and convolutional neural networks. arXiv:1911.00381.
18. Ramezani, M., Feizi-Derakhshi, M.-R., Balafar, M.-A., Asgari-Chenaghlu, M., Feizi-Derakhshi, A.-R., Nikzad-Khasmakhi, N., et al. (2022). Automatic personality prediction: An enhanced method using ensemble modeling. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-022-07444-6.
19. Han, S., Huang, H., & Tang, Y. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. Knowledge-Based Systems, 194, 105550. https://doi.org/10.1016/j.knosys.2020.105550.

20. Xue, X., Feng, J., & Sun, X. (2021). Semantic-enhanced sequential modeling for personality trait recognition from texts. Applied Intelligence. https://doi.org/10.1007/s10489-021-02277-7.

21. Zhu, H., Li, L., Jiang, H., & Tan, A. (2020). Inferring personality traits from attentive regions of user liked images via weakly supervised dual convolutional network. Neural Processing Letters. https://doi.org/10.1007/s11063-019-09987-7.

22. Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2018). Personality predictions based on user behavior on the Facebook social media platform. IEEE Access, 6, 61959–61969. https://doi.org/10.1109/access.2018.2876502.

23. Lima, A. C. E. S., & de Castro, L. N. (2014). A multi-label, semi-supervised classification approach applied to personality prediction in social media. Neurocomputing, 58, 122–130. https://doi.org/10.1016/j.neucom.2014.05.020.

24. Küster, L., Trahms, C., & Voigt-Antons, JN. (2018). Predicting personality traits from touchscreen based interactions. In 2018 Tenth international conference on quality of multimedia experience (QoMEX) (pp. 1–6). IEEE.

25. Roy, S., Roy, U., & Sinha, D. D. (2018). The probability of predicting personality traits by the way user types on touch screen. Innovations in Systems and Software Engineering, 15(1), 27–34. https://doi.org/10.1007/10.1016/j.isso.2013-017-000.

26. Shen, J., Brdiczka, O., & Liu, J. (2013). Understanding email writers: Personality prediction from email messages. In Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) (vol. 7899 LNCS, pp. 318–330). Springer. https://doi.org/10.1007/978-3-642-38844-6_29.

27. Yakoub, F., Zein, M., Yasser, K., Adl, A., & Hassanien, A.E. (2015). Predicting personality traits and social context based on mining the smartphones SMS data. In Advances in intelligent systems and computing (vol. 370, pp. 511–521). Springer. https://doi.org/10.1007/978-3-319-21206-7_44.

28. Arnoux, P.H., Xu, A., Boyette, N., Mahmoud, J., Akkiraju, R., & Sinha, V. (2017). 25 Tweets to know you: A new model to predict personality with social media. In Eleventh international AAAI conference on web and social media. http://www.aaai.org.

29. Allport, G. W. (1937). Personality: A psychological interpretation. Holt. https://psycnet.apa.org/record/1938-01964-000

30. Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1970). Handbook for the sixteen personality factor questionnaire (16 PF) in clinical, educational, industrial, and research psychology, for use with all forms of the test. III Institute for Personality and Ability Testing.

31. Eysenck, H., Eysenck, S. B., & Giuliett. (1975). Manual of the Eysenck personality questionnaire. Hodder and Stoughton.

32. Briggs, K. C., & Myers, I. B. (1976). Myers–Briggs type indicator. Consulting Psychologists Press.

33. John, O. P., & Srivastava, S. (1990). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives—PsycNET. In Handbook of personality: Theory and research (pp. 102–138). Guilford Press. https://psycnet.apa.org/record/1999-04371-004

34. Pennebaker, J. W., & Francis, M. E. (1996). Cognitive, emotional, and language processes in disclosure. Cognition and Emotion, 10(6), 601–626. https://doi.org/10.1080/026999396380079.

35. Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. Journal of Personality and Social Psychology, 77(6), 1296–1312. https://doi.org/10.1037/0022-3514.77.6.1296.

36. Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29(1), 24–54. https://doi.org/10.1177/0261927X09351676.

37. LIWC. (2015). Comparing LIWC 2015 and LIWC 2007|LIWC. http://liwc.wpengine.com/comparing-dictionaries/

38. Mairesse, F., Uk, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. Journal of Artificial Intelligence Research, 30, 457–500. https://doi.org/10.5555/1622637.1622649.

39. Coltheart, M. (1981). The mrc psycholinguistic database. The Quarterly Journal of Experimental Psychology Section A, 33(4), 497–505. https://doi.org/10.1080/14640748108400805.
40. Wilson, M. (1988). Mrc psycholinguistic database: Machine-readable dictionary, version 2.00. Behavior Research Methods, Instruments and Computers, 20(1), 6–10. https://doi.org/10.3758/bf03202594.

41. Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013a). Distributed representations of words and phrases and their compositionality. arXiv:1310.4546.

42. Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013b). Efficient estimation of word representations in vector space. In 1st International conference on learning representations, ICLR 2013—Workshop track proceedings. International Conference on Learning Representations, ICLR. http://roman.collobert.com/senna/

43. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. Transactions of the Association for Computational Linguistics, 5:135–146. https://doi.org/10.1162/tacl_a_00051. http://www.isthe.com/chongo/tech/comp/fnv

44. Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532–1543).

45. Vaswani, A, Shazeer, R., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need.

46. Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2020). Deep learning based text classification: A comprehensive review.

47. Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In NAACL HLT 2019—2019 Conference of the North American chapter of the association for computational linguistics: Human language technologies (vol. 1, pp. 4171–4186). Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/N19-1423.

48. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT training approach. arXiv:1907.11692.

49. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). Albert: A lite bert for self-supervised learning of language representations.

50. Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2020). Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.

51. Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., & Howard, N. (2013). Common sense knowledge based personality recognition from text. In Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) (vol. 8266 LNAI, pp. 484–496). Springer. https://doi.org/10.1007/978-3-642-45111-9_42.

52. Havasi, C., Speer, R., & Alonso, J. (2007). Conceptnet 3: A flexible, multilingual semantic network for common sense knowledge. Recent advances in natural language processing (pp. 27–29). John Benjamins.

53. Verhoeven, B., Daelemans, W., & Smidt, T. D. (2013). Ensemble methods for personality recognition. In Seventh international AAAI conference on weblogs and social media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/view/6177

54. Wright, W. R., & Chin, D. N. (2014). Personality profiling from text: Introducing part-of-speech n-grams. In Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics) (vol. 8538, pp. 243–253). Springer. https://doi.org/10.1007/978-3-319-08786-3_21.

55. Tighe, E. P, Ureta, J. C, Andrei Pollo, B. L., Cheng, C. K., & de Dios Bulos, R. (2016). Personality trait classification of essays with the application of feature reduction. In 4th Workshop on sentiment analysis where AI meets psychology (SAAP 2016) (pp. 22–28).

56. Liu, F., Perez, J., & Nowson, S. (2016). A recurrent and compositional model for personality trait recognition from short texts. In The workshop on computational modeling of people’s opinions, personality, and emotions in social media (pp. 20–29). https://www.aclweb.org/anthology/W16-4303.pdf

57. Liu, F., Perez, J., & Nowson, S. (2017). A language-independent and compositional model for personality trait recognition from short texts. In 15th Conference of the European chapter of the association for computational linguistics (vol. 1, pp. 754–764). https://www.aclweb.org/anthology/E17-1071

58. Zheng, H., & Wu, C. (2019). Predicting personality using facebook status based on semi-supervised learning. In ACM international conference proceeding series (vol. Part F148150, pp. 59–64). Association for Computing Machinery. https://doi.org/10.1145/3318299.3318363.
59. Chen, M., Weinberger, K. Q., & Chen, Y. (2011). Automatic feature decomposition for single view co-training. In ICML.

60. Guan, Z., Wu, B., Wang, B., & Liu, H. (2020). Personality2vec: Network representation learning for personality. In 2020 IEEE Fifth international conference on data science in cyberspace (DSC) (pp. 30–37). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/dsc50466.2020.00013.

61. Sun, X., Liu, B., Meng, Q., Cao, J., Luo, J., & Yin, H. (2020). Group-level personality detection based on text generated networks. World Wide Web, 23(3), 1887–1906. https://doi.org/10.1007/s11280-019-00729-2.

62. Wang, Z., Wu, C. H., Biao Li, Q., Yan, B., & Zheng, K. F. (2020). Encoding text information with graph convolutional networks for personality recognition. Applied Sciences (Switzerland), 10(12), 4081. https://doi.org/10.3390/app10124081. www.mdpi.com/journal/appsci

63. Ramezani, M., Feizi-Derakhshi, M.-R., & Balafar, M.-A. (2022b). Knowledge graph-enabled text-based automatic personality prediction. Computational Intelligence and Neuroscience, 2022, 1–18. https://doi.org/10.1155/2022/3732351. https://www.hindawi.com/journals/cin/2022/3732351/.

64. Ramezani, M., Feizi-Derakhshi, M.-R., & Balafar, M.-A. (2022c). Text-based automatic personality prediction using KGAt-Net: A knowledge graph attention network classifier. https://doi.org/10.48550/arxiv.2205.13780. arXiv:2205.13780.

65. Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. IEEE Intelligent Systems, 32(2), 74–79. https://doi.org/10.1109/MIS.2017.23.

66. Sun, X., Liu, B., Cao, J., Luo, J., & Shen, X. (2018). Who am I? Personality detection based on deep learning for texts. In IEEE International conference on communications (vol. 2018-May). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICC.2018.8422105.

67. Darliansyah, A., Asif Naeem, M., Mirza, F., & Pears, R. (2019). SENTIPEDE: A smart system for sentiment-based personality detection from short texts. Journal of Universal Computer Science, 25(10), 1323–1352.

68. Hutto, C., & Gilbert, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (vol. 8).

69. Jiang, H., Zhang, X., & Choi, J. D. (2020). Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings. In The thirty-fourth aaai conference on artificial intelligence (pp. 13821–13822). Association for Computational Linguistics (ACL).

70. Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E., & Eetemadi, S. (2020). Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. IEEE International Conference of Data Mining. https://doi.org/10.1109/ICDM50108.2020.00146.

71. Cambria, E., Poria, S., Hazarika, D., & Kwok, K. (2018). Senticnet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16839

72. Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon.

73. Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In Proceedings of the 56th annual meeting of the association for computational linguistics (Vol. 1: Long Papers, pp. 174–184). Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-1017.

74. Kazameini, A., Fatehi, S., Mehta, Y., Eetemadi, S., & Cambria, E. (2020). Personality trait detection using bagged SVM over BERT word embedding ensembles. arXiv:2010.01309

75. Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. Information Processing and Management, 58(3), 102532. https://doi.org/10.1016/j.ipm.2021.102532.

76. Yang, F., Quan, X., Yang, Y., & Yu, Jg. (2021). Multi-document transformer for personality detection. AAAI. https://www.aaai.org/AAAI21Papers/AAAI-2371.YangF.pdf

77. Yu, J., & Markov, K. (2017). Deep learning based personality recognition from Facebook status updates. In Proceedings—2017 IEEE 8th international conference on awareness science and technology, iCAST 2017 (vol. 2018-January, pp. 383–387). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICAwST.2017.8256484.
78. Farnadi, G., Zoghbi, S., Moens, M.-F., & De Cock, M. (2013). Recognizing personality traits using facebook status updates. In *Computational personality recognition (shared task)* (pp. 14–18). AAAI. http://www.aaai.org.

79. Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., Rui, Y., & Ma, W. Y. (2017). Beyond the words: Predicting user personality from heterogeneous information. In *WSDM 2017—Proceedings of the 10th ACM international conference on web search and data mining* (pp. 305–314). Association for Computing Machinery, Inc. https://doi.org/10.1145/3018661.3018717.

80. Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1746–1751). Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1181.

81. An, G., & Levitan, R. (2018). Lexical and acoustic deep learning model for personality recognition. In *Interspeech 2018* (vol. 2018-September, pp. 1761–1765). ISCA. https://doi.org/10.21437/Interspeech.2018-2263.

82. Biel, J. I., & Gatica-Perez, D. (2013). The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Transactions on Multimedia, 15*(1), 41–55. https://doi.org/10.1109/TMM.2012.2225032.

83. Biel, J. I., Tsiminaki, V., Dines, J., & Gatica-Perez, D. (2013). Hi YouTube! Personality impressions and verbal content in social video. In *2013 ACM international conference on multimodal interaction* (pp. 119–126). ACM Press. https://doi.org/10.1145/2522848.2522877.

84. Chen, Y.-H., & Choi, J. D. (2016). Character identification on multiparty conversation: Identifying mentions of characters in TV shows. In *Proceedings of the 17th annual meeting of the special interest group on discourse and dialogue* (pp. 90–100). Association for Computational Linguistics (ACL). https://doi.org/10.18653/v1/w16-3612. http://www.freebase.com

85. Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). TwitPersonality: Computing personality traits from tweets using word embeddings and supervised learning. *Information, 9*(5), 127. https://doi.org/10.3390/info9050127.

86. Xue, D., Wu, L., Hong, Z., Guo, S., Gao, L., Wu, Z., et al. (2018). Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence, 48*(11), 4232–4246. https://doi.org/10.1007/s10489-018-1212-4.

87. Mohammad, S. M., & Kiritchenko, S. (2013). Using nuances of emotion to identify personality. In *ICWSM workshop on computational personality recognition*. AAAI Publications. https://nrc-publications.canada.ca/eng/view/accepted?id=0d2da4fd-27fd-42f1-a867-d3de866ad8ea

88. Hassanein, M., Hussein, W., Rady, S., & Gharib, T. F. (2019). Predicting personality traits from social media using text semantics. In *Proceedings—2018 13th international conference on computer engineering and systems, ICCES 2018* (pp. 184–189). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCES.2018.8639408.

89. Khan, A. S., Ahmad, H., Asghar, M. Z., Saddozai, F. K., Arif, A., & Khalid, H. A. (2020). Personality classification from online text using machine learning approach. *International Journal of Advanced Computer Science and Applications, 11*(3), 460–476. https://doi.org/10.14569/ijacsa.2020.0110358. http://www.ijacsa.thesai.org

90. Amirhosseini, M. H., & Kazemian, H. (2020). Machine learning approach to personality type prediction based on the Myers-Briggs type indicator. *Multimodal Technologies and Interaction, 4*(1), 9. https://doi.org/10.3390/mti4010009.

91. 16PF Questionnaire—Wikipedia. https://en.wikipedia.org/wiki/16PF_Questionnaire

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
Authors and Affiliations

Ali-Reza Feizi-Derakhshi1 · Mohammad-Reza Feizi-Derakhshi1 · Majid Ramezani1 · Narjes Nikzad-Khasmakhi1 · Meysam Asgari-Chenaghlu1 · Taymaz Akan1,2,3 · Mehrdad Ranjbar-Khadivi1,4 · Elnaz Zafarni-Moattar1,5 · Zoleikha Jahanbakhsh-Naghadeh1,6

Ali-Reza Feizi-Derakhshi
derakhshi96@ms.tabrizu.ac.ir; a.f.derakhshi@gmail.com

Majid Ramezani
m_ramezani@tabrizu.ac.ir

Narjes Nikzad-Khasmakhi
n.nikzad@tabrizu.ac.ir

Meysam Asgari-Chenaghlu
m.asgari@tabrizu.ac.ir

Taymaz Akan
taymaz.farshi@gmail.com

Mehrdad Ranjbar-Khadivi
mehrdad.khadivi@iaushab.ac.ir

Elnaz Zafarni-Moattar
e.zafarani@iaut.ac.ir

Zoleikha Jahanbakhsh-Naghadeh
zoleikha.jahanbakhsh@iau.ac.ir

1 Computerized Intelligence Systems Laboratory, Department of Computer Engineering, University of Tabriz, Tabriz, Iran

2 Department of Software Engineering, Aysansaray University, Istanbul, Turkey

3 Clinical Informatics, Louisiana State University Health Sciences Center Shreveport, Shreveport, USA

4 Department of Computer Engineering, Shabestar Branch, Islamic Azad University, Shabestar, Iran

5 Department of Computer Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran

6 Department of Computer Engineering, Naghadeh Branch, Islamic Azad University, Naghadeh, Iran