KMAN PUBLISHING INSTITUTE
Journal Website

Article history:
Received 27 August 2024
Revised 02 November 2024
Accepted 14 November 2024
Published online 28 November 2024

Journal of Assessment and Research in Applied Counseling

In Press

E-ISSN: 3041-8518



# Affective Attention-Based BILSTM Models for Personality Prediction

Zohreh. Anisi Hemaseh[1], Mohamad Ali Afshar Kazemi[2*], Ehsan Mousavi Khanghah[3]

[1] PhD Student, Department of Information Technology Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[2] Associate Professor, Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran
[3] Associate Professor, Department of Computer Engineering, Shahed University, Tehran, Iran

*** Corresponding author email address**: dr.mafshar@gmail.com

A r t i c l e   I n f o

**Article type:**
*Original Research*

**How to cite this article:**
Anisi Hemaseh, Z., Afshar Kazemi, M. A., & Mousavi Khanghah, E.. (IN PRESS). Affective Attention-Based BILSTM Models for Personality Prediction. *Journal of Assessment and Research in Applied Counseling.*
http://dx.doi.org/10.61838/kman.jarac.7.x.x



A B S T R A C T

**Objective:** This study aims to develop a novel, more accurate personality prediction model that integrates affective dimensions and reduces computational complexity.

**Methods and Materials:** The proposed model preprocesses textual data using a novel method that removes noise and enriches semantic features with syntactic information. GloVe word embedding is utilized for semantic representation, and Bidirectional Long Short-Term Memory (BILSTM) networks extract contextual values. Affective dimensions (valence, arousal, and dominance) from the NRC Lexicon are used to create attention layers that focus on emotional aspects of words. The ensemble of these models improves prediction accuracy. The model was evaluated using the MBTI dataset from Kaggle, comprising 8675 users' Twitter data.

**Findings:** The proposed model achieved an average accuracy improvement of 3.35% to 6.23% compared to state-of-the-art models like BERT while reducing execution time significantly (by 1/24 of BERT's time). Incorporating affective dimensions enhanced predictions, particularly in personality traits resistant to change, such as Thinking-Feeling and Judging-Perceiving.

**Conclusion:** The integration of affective attention mechanisms and enhanced preprocessing methods significantly improves personality prediction models. The proposed approach not only boosts accuracy but also reduces computational demands, making it highly applicable in practical scenarios like customer profiling and recommendation systems. Future work could explore additional personality-related dimensions and unsupervised learning techniques.

***Keywords:*** *Personality prediction; MBTI; VAD analysis; Affective Attention mechanism; Ensemble; Bidirectional LSTM*

## 1. Introduction

With the evolution of computing technologies, personality prediction has become a popular field of research. Researchers have found a solid relationship between people's preferences, behaviors, and personality traits (Ozer & Benet-Martínez, 2006). Personality is a psychological structure that reflects individual emotions, thoughts, feelings, motivations, and attitudes characteristics that make people vary from each other (Yuan et al., 2018). Hence it can play an important role in personalization (Abidin & Remli, 2020). Personality recognition helps create a customized and personalized system that improves user experience. So far, different scenarios such as Human resource recruitment systems (Abidin & Remli, 2020), recommendation systems (Wang et al., 2021; Yang & Huang, 2019), chatbots (Galitsky, 2021), and advertainment design (Winter et al., 2021) have benefited personality traits that improve user experience.

Since lingual data is one of the best indicators of personality-related traits such as behaviors, emotions, and motivations, analyzing user content generated is a way of discovering individuals' personalities (Štajner & Yenikent, 2021). The Myers-Briggs Type Indicator (MBTI) (Myers & Myers, 1995) is one of the most popular and widely used personality (Abidin & Remli, 2020) assessments for understanding people in various fields that use behavioral concepts to study personality. Hence, we have chosen the MBTI assessment for our paper, consisting of four dual categories: Introversion(I) vs. Extroversion (E), Sensing (S) vs. Intuition(N), Thinking(T) vs. Feeling(F), and Judging(J) vs. Perceiving(P) (Myers & Myers, 1995).

Users on social networks generate millions of posts (Hassanein et al., 2019) that express their feelings, thoughts, and attitudes. With the growth of computing technologies, the utilization of this rich information resource is increasing. In a few years, one field of attention paid by academic researchers will be personality prediction.

Much research on different machine learning and deep learning on personality prediction. Machine learning methods for extracting the desired features often relied on word counting and statistical calculations of the presence of words and discovering the optimal space of these features. One of the areas related to vocabulary extraction and counting was the LIWC method (Pennebaker et al., 2007). Because access to it is not open and also unsuitable for social network applications (Ren et al., 2021), semantic feature extraction has been mainly used based on the neural network.

In recent years, with the development of computing technologies and emerging deep neural networks, researchers have adopted deep learning methods for personality prediction to improve the study of this field (Yuan et al., 2018). Intrinsically, predicting personality from text is a complicated problem; hence, to achieve a model with optimal performance just with the semantical and contextual dimensions of words, models become more complex and time-consuming, with more layers to be built on understanding the complexity of the problem. Therefore, the objective of this paper is to answer the following question how to provide a model with better performance so that the complexity of the previous models not only doesn't increase but also decreases?

The proposed model includes 4 phases: 1) the cleansing and preprocessing of noisy text data, 2) the extracting of semantic and emotional dimensions from text, 3) the design of the neural network for encoding data and enriching words by emotional dimensions, 4) fusing designed neural networks for improving and stabilizing performance of the final model.

The correlation between emotions and personality has led to using one of these two categories to improve the prediction of the other. Some studies have used personality to analyze emotions (Xue et al., 2018). Contrary, in some others, because personality is a relatively more stable trait, they have improved personality prediction by adding emotional features (Darliansyah et al., 2019). Majumdar et al. (2017) filtered words based on their polarity from the text using NRC Lexicon, which has 14,182 words (Majumder et al., 2017). They focused on words with positive or negative polarity and ignored the semantic effect of other words. Zhao et al. (2020) leverage the sentiment polarity features to word2vec word embed and use attention layer by topic preferences (Zhao et al., 2020). Ren et al. (2021) In another study, the polarity sentiment of each word was added as a feature to BERT's word embedding. Still, due to BERT having 768-dimensional features, one dimension of sentiment could not impact improving personality prediction accuracy (Ren et al., 2021). Hence, they mapped a one-dimensional sentiment vector to 20 dimensions by a linear transformation as data augmentation and then added it to BERT word embedding. As reviewed, emotions in personality prediction have been used based on adding the polarity of words or sentences as features to word embedding or sentence embedding.




JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

In this research, we have used the dimensions of affection. by expanding our study instead of just using polarity. In the NRC lexicon, Mohammad obtained valence, arousal, and dominance of 20,000 words (Majumder et al., 2017)Click or tap here to enter text.. The VAD affection model shows three nearly orthogonal dimensions that comprehensively describe emotional states (Mehrabian, 1996b). There is a correlation between Personality traits and affection analysis. Wen et al. used linear regression formulas (Mehrabian, 1996a) from the Big Five personality traits to calculate VAD values (Wen et al., 2021). Bauerhenne et al. (2020), in their project, used a mapping between the Big five personality traits and VAD space (Bauerhenne et al., 2020). This research shows a strong correlation between personality prediction and emotional dimensions.

Deep learning-based personality prediction has been a new research field in recent years. Previous papers (Ahmad et al., 2020; Artha Agastya et al., 2019; Mehta et al., 2020; Remaida et al., 2020) have reviewed machine learning, deep learning methods, and data sources in the personality prediction field. Deep learning-based personality prediction with the increasing power of computational computers is growing. Deep learning-based personality prediction with the increasing power of computational computers is growing. The following section reviews the previous studies by focusing on the deep learning method for text-based personality prediction.

Majumdar et al. (2017) used deep Covolutional Neural Network (CNN) for extracting n-gram features by convolutional filters on non-neutral sentences (Majumder et al., 2017). To improve the model's performance, they concatenate n-gram features with Mairesse Linguistical features (Mairesse et al., 2007). They showed that CNN alone performs not well, and applying sentiment filters and adding the document level, Mirasse Linguistical features have improved model accuracy Xue et al. (2018) used a Bidirectional Gated recurrent unit (GRU) with an attention mechanism layer at the sentence level, concatenated outputs and fed them into the CNN layer as input (Xue et al., 2018). While Lynn et al. (2018) use of attention mechanisms in a hierarchical way at the message level and word level (Lynn et al., 2020). Wang et al. (2020) All personality traits were correlated together, and a model was created using graph convolutional networks (Wang et al., 2020). Jeremy and Suhartono (2021) used Recurrent Neural Network (RNN) to create word embedding for neural networks to predict personalities from Indonesian tweets (Jeremy & Suhartono, 2021). Sun et al. (2018) exploited BILSTM and CNN models to detect personality traits (Sun et al., 2018).

However, with the popularity of CNN and LSTM-based methods in personality prediction, Yang et al. (2022) addressed that generic attention-based CNNs, Graph Convolutional Networks (GCNs), and Long Short-Term Memory networks (LSTMs) have a big limitation in personality prediction space in that they are unable to acquire linguistic cues at different granularities including person-level characteristics, psychological concepts, and syntactic at the word level, so they were designed personality model with transfer approach and hierarchical by adding the psychological feature (Yang et al., 2022). Zhao et al. (2020) enriched insufficient semantic context information by adding sentiment features to the model (Zhao et al., 2020). Also, they applied the attention mechanism layer (Vaswani et al., 2017) to LSTM, given the proven importance of the attention mechanism (Kim et al., 2017). They used the distribution of theme preferences extracted by LDA (Blei et al., 2003) for the weight of attention in the attention mechanism. And also mentioned that we need to add extra knowledge to the text to enrich the personality prediction model. Ren et al. (2021) noted two limitations of personality prediction methods: the lack of sentiment analysis and ignoring information about the context of words. They leveraged BERT to generate sentence-level embedding and then added sentiment information (Ren et al., 2021).

Some researchers have used the ensemble approach to improve personality prediction model performance. The use of ensemble models in the deep learning method for text in various fields has improved the results over single models (Alsayat, 2022; Heikal et al., 2018; Mohammadi & Shaverizade, 2021; Rokach, 2019). The ensemble methods combine base classifiers with improving model accuracy and stability. Ensemble methods are categorized into two groups: dependent and independent methods. In dependent methods, The output of one classifier impacts another. But in independent methods, each classifier executes independently and is combined with obtained outcomes. The Averaging strategy of independent methods ensemble is widely used (Mohammed & Kora, 2021). Voting is a method of averaging. In this method of determining class, prediction is achieved based on a label with the most frequent vote on base classifiers. This method of voting is called hard voting, shown in Eq 1.

$$y_i=\text{mode}\{cl_1,cl_2,\ldots,cl_n\} \quad (1)$$





JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

$cl_i$ is the label of each classifier. In contrast, soft voting in Eq 2. considers the probability value of each classifier rather than the prediction labels of each classifier.

$$y_i = argmax_i \frac{1}{n} \sum_{j=1}^{n} p_{ij} \qquad (2)$$

That $p_{ij}$ is the probability of *ith* class label of the *jth* classifier of *n* classifier. One of the popular methods in weighted majority voting ensemble is using accuracy as a weight for each classifier (Rokach, 2019).

In personality prediction, EI-Demerdash et al. (2020) used various pre-trained language models to get more semantics from the context of the text (El-Demerdash et al., 2022). They fused classifier methods for improving accuracy with fixed rule methods. Teragawa and Li (2020) used Feature fusion and voting mechanisms to ensemble LSTM and CNN and CNN with LSTM output models (Teragawa & Li, 2020).

The primary purpose of this paper is to achieve a novel more accurate personality prediction model based on informal texts such as social networks texts. To achieve this purpose, again, we return to the definition of personality. Personality is a psychological construct reflecting an individual's behavior, emotions, thought patterns, and motivational characteristics (Han et al., 2020; Mehta et al., 2020; Yuan et al., 2018); in other words, personality combines the mentioned components. Hence, we leverage personality prediction by adding affection analysis.

Some studies have shown a relationship between personality and emotions (Remaida et al., 2020). This concept indicates that people with similar personalities use words with similar emotional weight. Therefore, the model should consider the emotional aspect of the words in the embedding and encoding of words; By although some previous research, it has mentioned adding emotional dimensions to the words, still two limitations can be seen in them: The first limitation is that these studies mainly have used the sentiment polarity of words or sentences (Ren et al., 2021; Zhao et al., 2020) While a word emotionally, in addition to its polarity, has other dimensions such as dominance and arousal that can be affected by the personality trait of individuals. The second one is that sentiment information has been applied to the model just by concatenating sentiment dimension(s) to semantical dimensions. Due to the sentiment dimensions (almost one dimension) being much smaller than semantic dimensions (200, 300, or 768 dimensions), they will not significantly impact the final models (Majumder et al., 2017; Ren et al., 2021). To tackle these limitations, we add arousal and dominance of dimensions to the polarity of words. Hence, we used the lexicon provided by the National Research Council Canada (NRC) because it has these three dimensions of emotion, and the model has paid attention to sentences at the word level from the aspect of emotions.

On the other hand, social network text has many noises, and the usual preprocessing method is insufficient. If Input data is noisy, we can't expect a good model performance even if complicated models with more layers are used. We prefer cleansing data by a new method. This method removes potential noisy data and stopwords by positional words and Latent Dirichlet Allocation (LDA) modeling.

Pre-trained Global Vector (GloVe [*]) word embedding on Twitter data applies to clean data for extracting semantical feature dimensions. We encode word embedding dimensions by Bidirectional long short-term memory (BILSTM) to improve semantical and contextual dimensional values. We extract the emotional value of each word based on the VAD (Valence, Arousal, Dominance) NRC Lexicon[†]. This lexicon includes three important dimensions of the meaning of the word: valence, Arousal, and dominance for 20,000 words. To enrich encoded words' values extracted by BILSTM with emotional awareness, we create an attention layer and focus on each word by its emotional level. We make three models based on each dimension of emotion and ensemble them to improve and stabilize the model's performance.

Indeed, by modeling the attention mechanism on preprocessed and encoded data, we present an effective method in a shorter time than Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) algorithms for personality recognition. We have defined the attention weights based on three emotional dimensions. Our proposed model is the fusion of three BILSTM affective attention models. To evaluate the performance model, we apply it to the MBTI public dataset, which includes users' tweets with personality type labels.

The main contributions of this study are as follows:

1) We propose a novel multi-label personality detection model that combines BILSTM encoded value with attention mechanism and ensemble models to improve and stabilize final models.
2) We propose a novel preprocessing method by using syntactic role of words that reduces text noise and

[*] https://nlp.stanford.edu/

[†] https://nrc.canada.ca/





JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

improves models' performance. This method can apply to all text pattern recognition problems.

3) We propose a method that pays attention to semantical features by the level of emotion of each word. Indeed, it strengthens or weakens the impact of words by using these dimensions. The combination of Valence, Arousal, and dominance dimensions adds some ability to explain personality and contributes to the analysis of personality traits.
4) Experiment results on the MBTI datasets demonstrate that our model outperforms the state-of-the-art techniques in personality detection with execution time much less than BERT.

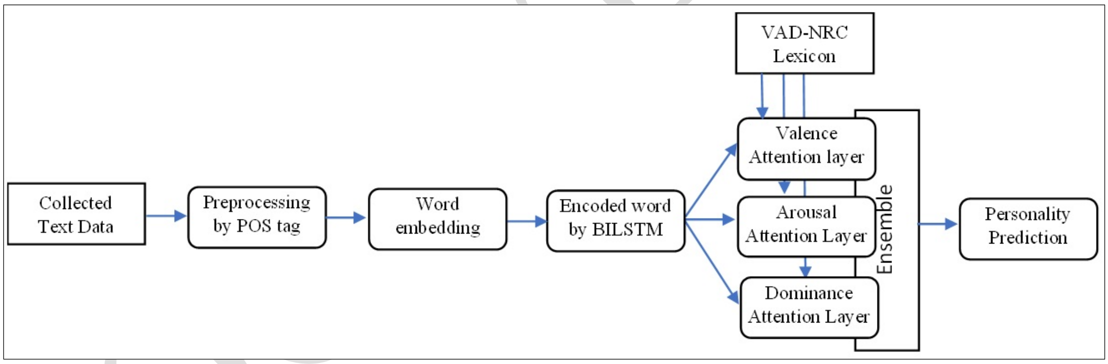
## 2. Methods and Materials

### *2.1. Study Design*

This paper uses a Twitter data set with labels of personality types for each user. Hence, the model is learned by supervised deep learning. For the MBTI personality model, we have 16 different character types. Each type includes four dimensions that convert to four labels for each personality type. Since these labels are independent, a neural network model for each label has been transformed from a multiclass classification problem to a multilabel classification problem. Figure 1 shows the methodology steps of this paper. Data are preprocessed by a new method for removing stopwords to reduce the noise of texts. For representing, words of sequences used, the Twitter GloVe pre-trained model. The next step encodes the GloVe word representation by BILSTM to extract left and right semantics. Encoded semantic words are fed to attention layers built by the affection weights of each word instead of focusing on all words equally. VAD NRC Lexicon has created the affection weights for each personality trait.

**Figure 1**

*The steps of the proposed model*




### *2.2. Preprocessing*

Preprocessing is one of the most significant steps to improve performance models in text classification problems. It includes several steps that are usually the same in all research for formal and informal text. While informal texts, especially social networks, have a lot of noise, the usual pre-processing methods are not enough to clean the data. Hence to reduce noise from these, we add some steps in preprocessing phase. We have two activities for preprocessing of data. The first, change and update of default stopwords list by using the LDA method. We create the LDA model on texts of the dataset and extract 20 topics with 20 words. Meaningless repetitive words existing in 400 words be chosen and added to the default stopwords list. The second enrich each word by syntactic feature. Some words, based on their role in the sentences, are ignorable. On the other hand, in personality prediction, using some parts-of-speech tags is more important than others. We add syntactic information to each word by leveraging NLTK tools. Part-of-Speech tagging annotate a role for each word in the sentence that is the lowest level of syntactic analysis. So, it is used for achieving syntactic information and word sense disambiguation and could distinguish each morphism in polymorphism words.




JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

**Table 1**

*List of Filtered tags from NLTK POS Tag*

| Tags | Description | Examples |
|---|---|---|
| CC | coordinating conjunction | And, both, but, et, for, etc. |
| CD | cardinal digit | One, mid-1980,000, 271, etc. |
| DT | determiner | All, an, another, the that, etc. |
| IN | preposition/subordinating conjunction | Out, inside, pro, on, by, etc. |
| dEX | existential there | there |
| FW | foreign word | |
| LS | list market | |
| NNP(s) | proper noun | Motown, Liverpool, ODI, etc. |
| POS | possessive ending | ‘, 's |
| TO | infinite marker (to) | To |
| UH | Interjection | Gosh, whammy, Oops, golly, etc. |

We extract of syntactic of words via NLTK Pos of tag and create a vector of role of words in sentences related to tokenized sentences. In Figure 2 shown the steps of our proposed method for preprocessing data. We remove the URLs, digits, emojis, and emoticons from the text without removing punctuation. For removing emojis, we use code of emojis and emoticons. Then we correct the words that the letters have been repeated more than twice. In the second step, we mark words in the text as a particular part of speech using NLTK pos tagging. In the third step, we add a list of extracted customized stopwords by LDA (Blei et al., 2003) to NLTK stopwords so that each word is in the top ten topics and does not meaningful; it removes from the text. Then we add punctuation to the stopwords list. Social media texts are informal texts and contain noisy words or misspellings. We chose the words based on POS tagging in the sentence. As described in the second step, we have given a pos tag to each word. We selected just words that their tags exist in this list ['VBN','WP','VBG', 'JJ', 'VBZ', 'VBP', 'NN', 'PRP', 'WP$', 'PRP$', 'WDT', 'RB', 'RBR', 'RBS', 'VBD', 'IN', 'JJR', 'JJS', 'PDT', 'MD', 'VB', 'WRB', 'EX', 'NNS']. This filtering helps us to remove some noise from the text. Table 1 shows tags that their related words are removed. In the Figure 2 shown the flowchart of preprocessing data.

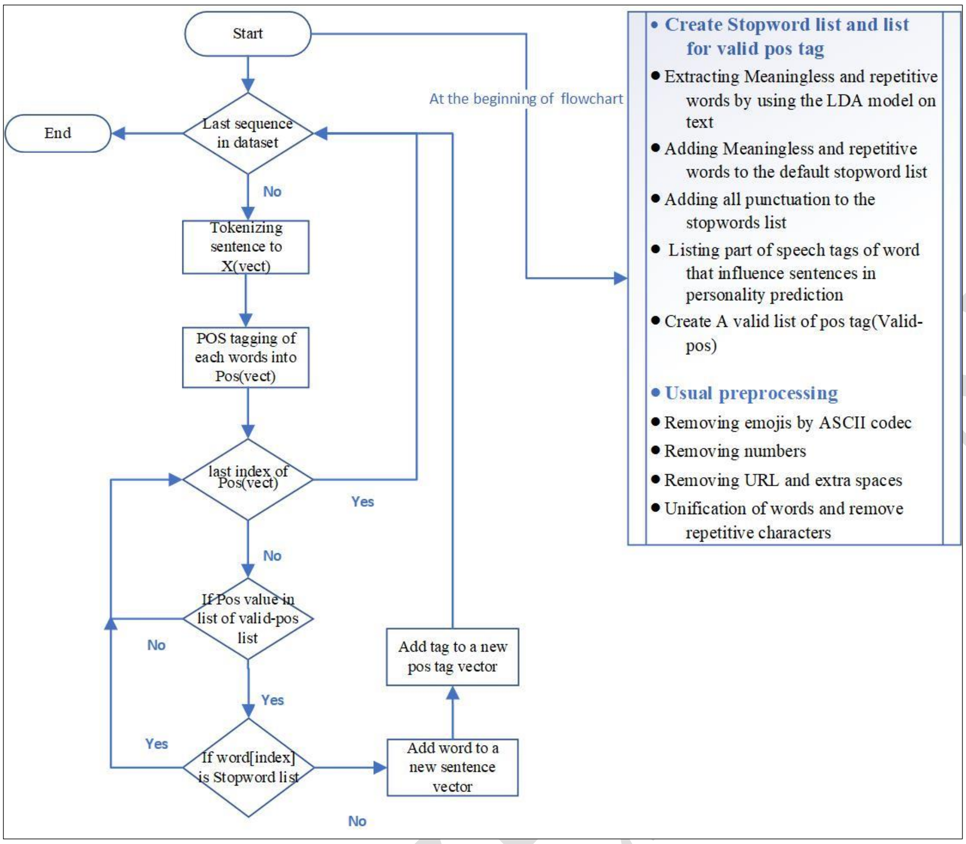
**Figure 2**

*Flowchart of the proposed method of preprocessing*





JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518




### *2.3. Encoder block*

We have used pre-trained tweeter GloVe word embedding to benefit local and global contexts. Max length of each document is 600, and the dimension of the word in this pre-trained word representation is 200; the space of word embedding is (601*200) for N users. To stabilize the learning process and reduce the number of training epochs, we add a Batch Normalization Layer (Ioffe & Szegedy, 2015). After normalization, data is fed as input into LSTM Networks. LSTM is a popular type of RNN that is powerful for encoding input sequences. These networks perform as well for long sentences and overcome vanishing gradients (Hochreiter, 1998; Hochreiter & Schmidhuber, 1997).

Because LSTM does not consider the whole context provided by the next words, we use Bidirectional LSTM. BILSTM trains two LSTM instead of one. One processes input sequences and the other processes reverse input sequences. For encoding sequences in BILSTM, we can get the hidden forward state from forwarding LSTM $\vec{h}_i$ and the backward hidden state $\overleftarrow{h}_i$ from backward LSTM for each sentence. And then, as shown in Eq 3, we merge two hidden states.

$$h_i = \vec{h}_i \oplus \overleftarrow{h}_i \qquad (3)$$

### *2.4. Affective attention mechanism block*

One of interest advancement of machine learning has been attention mechanism (Yang et al., 2022). Attention is becoming a most important concept in deep learning. In their study, Niu et al. (2021) reviewed methods and applications that used attention mechanisms in computer vision and natural language processing (Niu et al., 2021). The human biological system inspires the attention mechanism, which takes more attention to specific elements to understand phenomena. Humans don't tend to perceive all information at once. Instead, they to selectivity focus on the part of information when needed and about other information less pay attention to them or ignore them altogether. We used the attention concept in natural language processing and created custom attention weights by valence-arousal-dominance orthogonal dimensions of affection.

This section has two parts. In the first part, we addressed how emotional attention weight calculates and explained creating an attention mechanism layer.





JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518

### *2.5. Affection & emotion analysis*

Words play a significant role in understanding and describing the world around us. In several studies three most important, largely independent, dimensions of word meaning are valence (positiveness–negativeness/pleasure–displeasure that is sentiment polarity), arousal (active-passive degree), and dominance (dominant-submissive that is active power) (Mohammad, 2018; Russell, 1980, 2003). hence for evaluating the similarity of two words in addition to the semantical feature, we also could compare their degree of Valence, arousal, and dominance.

So far, the number of applications include automatic sentiment and emotional analysis of the text (Xiang et al., 2020; Xiang et al., 2021), cognitive science (Reuderink et al., 2013), and psychology for understanding how people view the world around them (Grgic & Podobnik, 2021; Wen et al., 2021), and social sciences for understanding relationships between people (Mehu & Scherer, 2015; Parkinson & Manstead, 2015) benefited from valence, arousal, and dominant dimensions. In this regard, some researchers created lexicons to take advantage of the emotional dimension in the analysis of texts. Bradley and Lang (2008), by the theory of emotions for the first time, three types of ratings were carried out for 1,034 words called the ANEW lexicon (Bradley & Lang, 2008). Warriner et al. (2013) extended the ANEW lexicon to 1,340 words (Warriner et al., 2013). Mohammad (2018), in VAD NRC Lexicon, obtained a human rating of valence (V), arousal (A), and dominance (D) for 20,000 words by using the Best–Worst Scaling to obtain fine-grained scores and address issues of annotation consistency (Mohammad, 2018). He showed statistically significant differences in understanding valence, arousal, and dominance across age, gender, and personality. Because VAD NRC Lexicons cover more words and are vastly more reliable than those in existing lexicons (Mohammad, 2018), this paper has used this lexicon to

**Figure 3**

*The proposed personality model*

analyze emotional dimensions as the most important aspects of each word besides the semantical aspect. These dimensions are independent of each other (Russell, 2003). Hence, we focus on each dimension separately.

### *2.6. Affective attention layer for personality prediction*

Affection values in each dimension of valence, Arousal, and dominance in important dimensions, meaning words for creating attention weights. Eq 4 shows the calculation of three weights for each dimension of VAD space. Each model of prediction personality dimension coefficient of vector values could be different.

$$\begin{cases} W_{Vi} = 1 + \alpha_i v_V \\ W_{Ai} = 1 + \beta_i v_A \\ W_{Di} = 1 + \gamma_i v_D \end{cases} \quad (4)$$

$\alpha_i, \beta_i, \gamma_i$ where $i$ valued by E-I, N-S, T-F, J-P, indicates a non-negative value for increased impact of each VAD dimension relative to each personality type. For not ignoring semantical dimensions in subsequent steps, the number one has been added to the equation. So, we have three weights for each personality trait dimension, and overall, we calculate 12 attention weights.

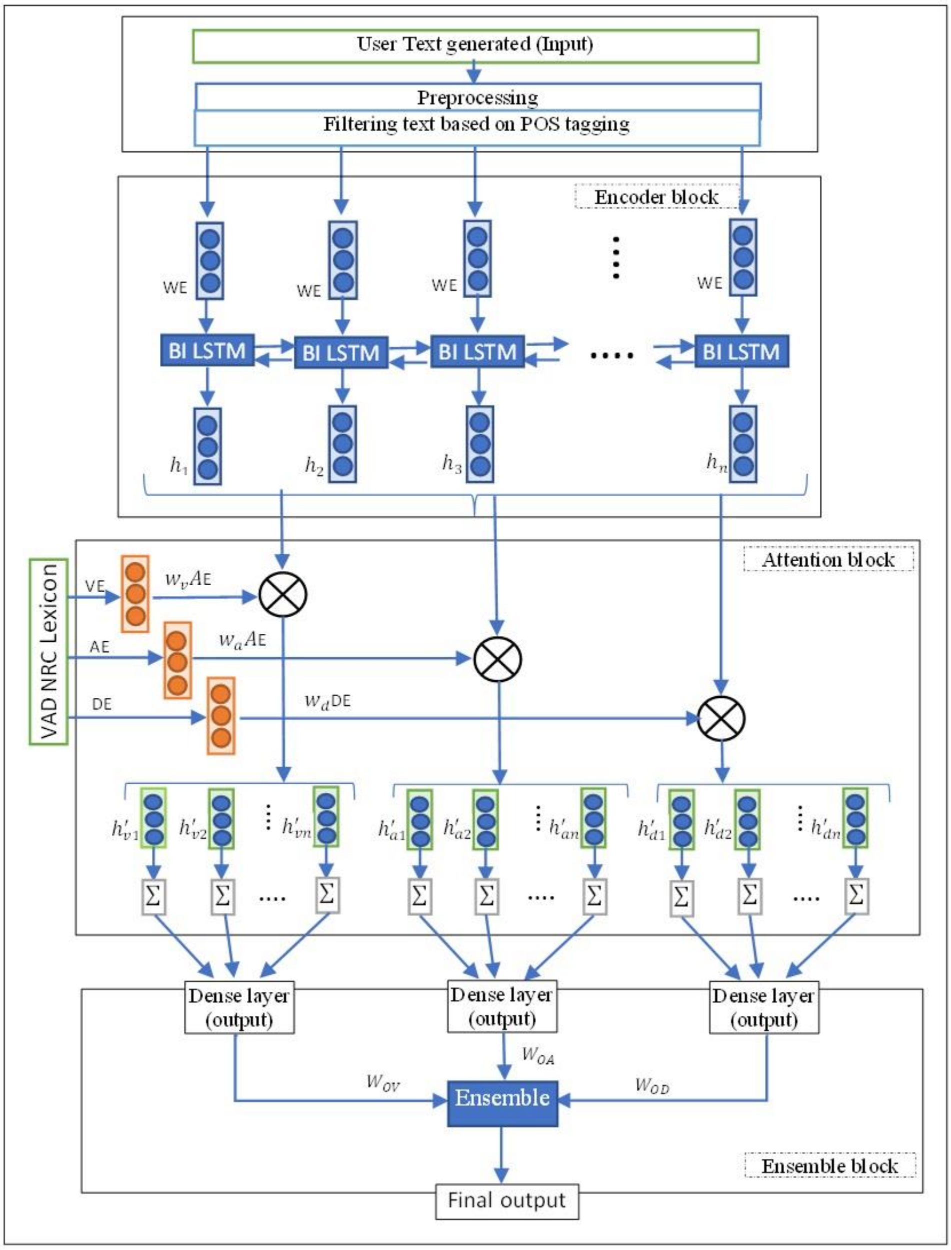
For the application of the attention mechanism, have been taken help from calculated affective weights to highlight words based on the degree of their emotions. To build of attention layer in this paper, weights of affection multiplied by the output of the BILSTM encoders.

$$h_i' = h_i * W_i \quad (5)$$

$h_i$ is the output encoding word embedding dimensions from BILSTM and $W_i$ includes $W_{Vi}, W_{Ai}, W_{Di}$ that are the weight of each dimension of VAD spaces calculated in Eq 5.




JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518




In Encoder Block, WE are shown GloVe word embedding, and $h_i$ is word sequences encoded which id the output of BILSTM. In the attention block, we use the VAD NRC lexicon for 20000 words with three dimensions valence, arousal, and dominance. We create three vectors for each dimension: VE, AE, and DE. Then each vector multiplies in a non-negative value $w_v$ for the Valence vector, $w_a$ for the arousal vector, and $w_d$ for the dominance vector for weighting them. After we multiply the output of BILSTM in these weighted affection vectors separately, we have three model outputs $h'_{vi}, h'_{ai}. h'_{di}$. Then we sum up ($\sum$) dimensions of each word. And each model used a dense layer for predicting the class of personality. In the ensemble block, used accuracy of each attention-based BILSTM model as weight - $W_{OV}$ for valence attention based BILSTM model, $W_{OA}$ for arousal attention based BILSTM model, and $W_{OD}$ for dominance attention based BILSTM model for majority voting.

Indeed, we multiply each value from the encoder word values obtained by BILSTM in its related affective attention weight vectors and sum all values for each word. These two steps are formulated in Eq 6.

$$\begin{cases} \sum h'_{vi} = \sum(h_i * W_{Vi}) \\ \sum h'_{ai} = \sum(h_i * W_{Ai}) \\ \sum h'_{di} = \sum(h_i * W_{Di}) \end{cases} \quad (6)$$

After the attention mechanism, we create a dense layer with a sigmoid activation function for training the network's output. Figure 1 shows three outputs. Each output has resulted from performing V, A, and D weight attention vector to encoder output in this step. We fused three models for each personality trait to take advantage of the performance of all three models.

For evaluating of model performance be used accuracy metric that in Eq 7.

$$Accuracy = \frac{True_{Positive}+True_{Negative}}{True_{Positive}+True_{Negative}+False_{Positive}+False_{Negative}} \quad (7)$$

### 2.7. Ensemble Block

Because of the variety of accuracy of models, we fuse three models to reach the best accuracy between models. We build three models for each valence, Arousal, and Dominance Dimension as attention points to each word and fuse these models with the majority voting technique. These models have different aggregation values; hence, we use weighted soft majority voting in Eq 8. These weights are the accuracy of the efficiency of each classifier.

$$y_i = argmax_i \frac{1}{n} \sum_{j=1}^{n} m_j s_{ij} \quad (8)$$

That $s_{ij}$ is the probability of belonging to one or zero class of the $jth$ classifier of three classifiers, and $m_j$ is the accuracy of each classifier. $y_i$ is the final value for predicting class.

## 3. Findings and Results

### 3.1. Data Set and Preprocessing

The MBTI personality type dataset is a standard dataset that contains 8675 volunteer users' data on Twitter, which is one of the largest datasets available with personality type labels. This set includes two fields, post, and type. Each post has 50 of each tweet sent by the user, and each tweet has the character "|||" separated, which contains 422,845 tweets. The type field addresses the personality of the tweet generator user. Each type includes four letters; each letter signifies a key aspect of an individual's personality. The personality café forum has collected this dataset, and permission has been given for its use in academic research.

Preprocessing is one of the most significant steps that effectively reduces noise in textual data. Data preprocessing often involves several specific parts. We create a list of stopwords, remove them from text, and filter words based on their role in sentences.

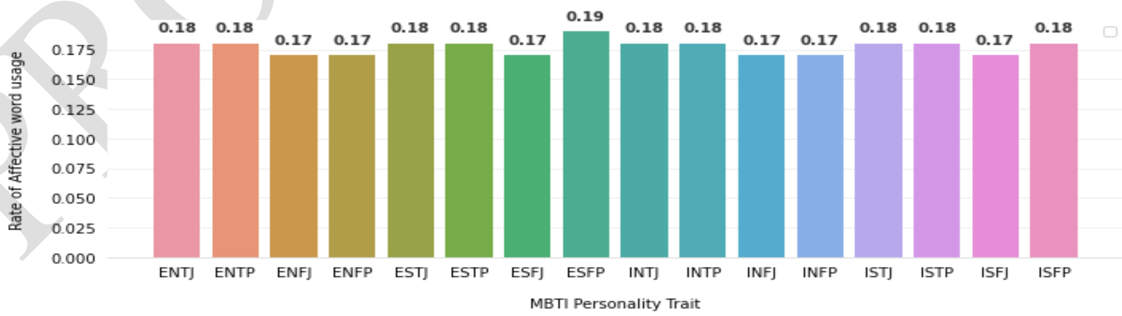
### 3.2. Affective Relevance Analysis

We have used VAD NRC Lexicon to add emotional features to this paper. This Lexicon contains 20,000 words; each word has a real value between 0 and 1 in three dimensions valence, arousal, and dominance. Before implementing the model, we analyzed this lexicon and the distribution of using it on the data set.

We first discuss the proportion of affective words in texts generated by each personality type. Figure 4 shows that each personality type has an approximate same portion in the usage of affective words (between 17.5 and 19.5).

**Figure 4**

*The portion of affective words used by users in each personality type*



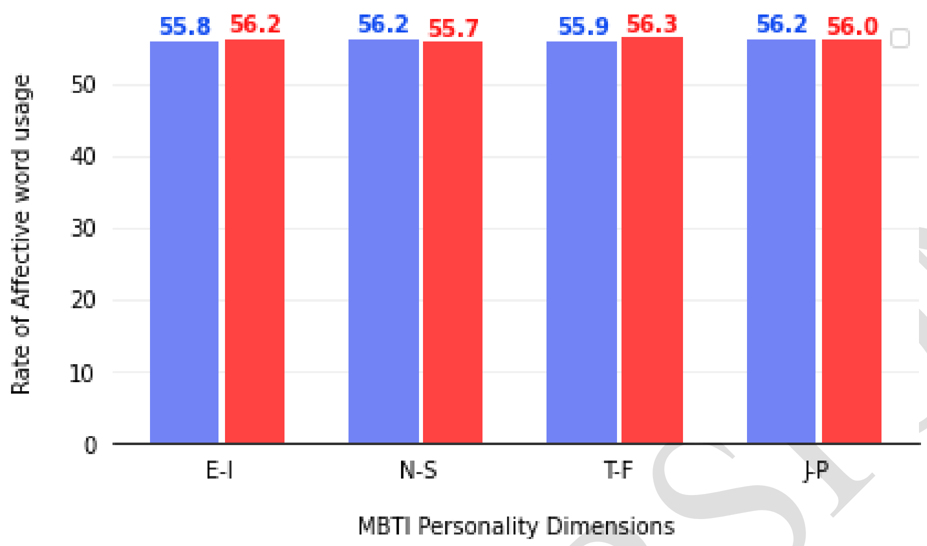

For a more detailed analysis, considering that we have used the multi-label models for predicting personality, we analyzed the share of usage of this word for each label. Figure 5 shows that each personality dimensions have a share of more than 50% (55.7-56.3) in the use of affective words. Therefore, we could result that these words include information in addition to semantics content that needs more attention.




JARAC Assessment and Research in Applied Counseling
E-ISSN: 3041-8518

**Figure 5**

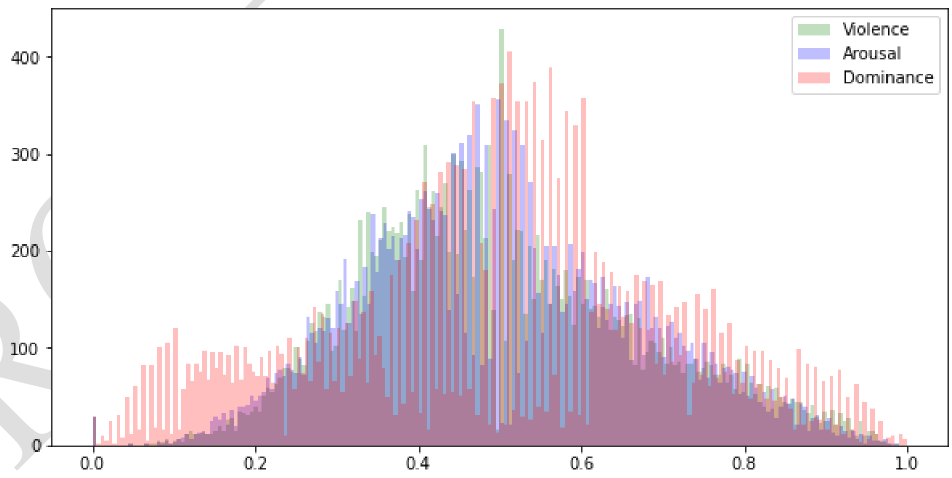
*The portion of the affective word in each dual personality dimension*




At last, we analyze the distribution of the values of each dimension in the VAD NRC Lexicon. Figure 6 shows that the value of words in the valence, Arousal and dominance dimension is different, and it expresses that each dimension of affection has different distribution between words. Paying attention to each of these dimensions can affect identifying personality dimensions differently.

**Figure 6**

*Distribution weight of Valence, Arousal, and Dominance words in NRC Lexicon*




### 3.3. *Experimental setting*

#### 3.3.1. *Baseline Methods*

We have listed neural network methods most recently used in feature extraction and predicting personality traits.

GloVe: uses a GloVe word embedding for representing the semantics of words that benefit from combining advantages of global matrix factorization and local context window methods (Pennington et al., 2014).

CNN: uses a convolution layer to extract importance and similar feature in the text (Jacovi et al., 2018).

LSTM: That is a famous RNN architecture with a gated mechanism to encode longer sentences and memorize





JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

important words. LSTMs were developed to handle exploding and vanishing gradient problems when training traditional RNNs (Hochreiter & Schmidhuber, 1997).

BILSTM: That is Called Bidirectional LSTM. It Concatenates LSTM encoding in two directions to model semantic words from left and right neighbor words (Graves & Schmidhuber, 2005).

Transformer Encoder: uses word embedding and positional embedding with a multi-head attention mechanism for encoding text representation (Vaswani et al., 2017).

BERT: A Transformer Bidirectional Encoder Representation that models the context of words via a pre-trained language model (Devlin et al., 2019).

We have compared its accuracy with baseline methods to evaluate the impact of improving our proposed model.

*Training*

**Parameter setting:** In deep neural networks, hyperparameter choice is a critical task. We investigated the different values to get the best values. Table 2 shows the final values of these parameters. Due to the complexity of deep neural networks, one of the most likely problems is the overfitting of the models. We use four techniques to prevent this problem: batch normalization (Sun et al., 2018), dropout (Russell, 2003), regularization, and early stopping (Xiang et al., 2020). Batch normalization and dropout techniques have been used before the BILSTM layer and the regularization method in the dense layer after the BILSTM layer. And as the final technique, we used early stopping in fitting models. Sequences have been vectorized using pre-trained GloVe word embedding trained on tweeter data with 200 dimensions for each word.

For the configuration of baseline methods, we have used 64 channels for CNN and pool with size=1. And we have designed a network with dual head attention and 32 dimensions, and 32 neurons for the Encoder transformer implementation.

The coefficients of the mentioned weights in equation 4 in this study are considered equal to 1.

**Evaluation:** We use accuracy to evaluate the performance of the prediction personality trait model. Our model consists of four binary class classifications. So, Accuracy is not only a popular metric for measuring performance personality trait predictions but is also suitable for binary classification problems. Besides, the average accuracy is obtained for comparing the results of models in conclusion.

**Table 2**

*Configuration of BILSTM model in the proposed model*

| Hyperparameters | Values |
|---|---|
| LSTM neurons | 32 |
| Dropout | 0.5 |
| L1 Regularize | 0.01 |
| Optimizer | RMSprop |
| Learning Rate | 0.005 |
| No. epochs | 10 |
| Batch size | 64 |

### *3.4. Result & Discussion*

In this paper, because we have a multi-label classification, and each label in MBTI is independent of another, we model a network for each label separately. We investigate the deep learning methods for feature extraction and classifying. This section shows that our proposed model outperforms in comparison with state-of-the-art methods. We trained Encoder transformer, CNN, LSTM, and BILSTM on the pretrained GloVe word embedding and pre-trained BERT with a dense layer. Table 3 shows the result of all models. We have highlighted the best values in this table.

Our proposed model improves some dimensions of personality traits, including Extroversion-Introversion and Intuition-Sensing, which are resistant to change. As evident in Table 3, among the state-of-the-art models, the BERT model and a dense layer with a sigmoid activation function have outperformed the other models. However, our proposed model (both with the normal pre-processing and the new pre-processing method) has performed better than the BERT model. We found that GloVe word embedding with LSTM performs better than BILSTM, which means the best algorithm's performance in each problem could be different. Adding affection information to word embedding on the baseline methods shows T-F personality trait could be improved.

**Table 3**

*Comparison of Accuracy Scores proposed model with baseline models*

| Model | E-I | N-S | T-F | J-P | Avg |
|---|---|---|---|---|---|
| GloVe +LSTM | 76.91 | 86.21 | 61.47 | 62.54 | 71.78 |
| GloVe +BILSTM | 76.91 | 86.21 | 60.51 | 60.32 | 70.98 |
| BERT | 76.95 | 86.20 | 70.53 | 60.93 | 73.65 |
| ENCODER+MLP | 76.91 | 86.21 | 59.59 | 60.43 | 70.78 |
| ENCODER+LSTM | 76.91 | 86.21 | 64.81 | 60.85 | 72.19 |
| GloVe +CNN+LSTM | 76.91 | 86.21 | 57.16 | 60.43 | 70.17 |
| GloVe +AFF+LSTM | 76.91 | 86.21 | 63.00 | 60.05 | 71.54 |
| GloVe +AFF+BILSTM | 76.91 | 86.21 | 68.15 | 60.15 | 72.85 |
| Proposed model | 77.43 | 86.40 | 76.35 | 63.40 | 75.92 |
| Proposed model + novel preprocessing | 78.30 | 86.66 | 76.85 | 65.76 | 76.90 |

Three affective attention weights have been created based on three important dimensions of the word meaning. These affective attention weights include valence attention weight (VATT), Arousal attention weight (AATT), and Dominance attention weight (DATT) vectors. Also, we have used the average of three dimensions as another affective attention weight (AFFATT). And then, we applied them to the BILSTM encoder outputs. Table 4 shows the results of four experiments compared to our proposed model with the typical preprocessing method.

**Table 4**

*Comparison of BILSTM with different affective attention models with the proposed model (with the typical preprocessing method)*

| Model | E-I | N-S | T-F | J-P | Avg |
|---|---|---|---|---|---|
| GloVe +BILSTM+VATT | 78.06 | 86.21 | 74.41 | 60.20 | 74.72 |
| GloVe + BILSTM+AATT | 77.10 | 86.21 | 74.62 | 62.47 | 75.10 |
| GloVe +BILSTM+DATT | 75.26 | 85.75 | 75.64 | 60.24 | 74.22 |
| GloVe +BILSTM+AFFATT | 77.30 | 86.32 | 74.41 | 60.70 | 74.68 |
| Proposed | 77.43 | 86.40 | 76.35 | 63.40 | 75.92 |

Table 5 shows the of four experiments in comparison with our proposed model with the new preprocessing method.

**Table 5**

*Comparison of BILSTM with different affective attention models with the proposed model (with the novel preprocessing method)*

| Model | E-I | N-S | T-F | J-P | Avg |
|---|---|---|---|---|---|
| GloVe +BILSTM+VATT | 78.90 | 86.35 | 75.68 | 65.08 | 76.50 |




E-ISSN: 3041-8518

| | | | | | |
|---|---|---|---|---|---|
| GloVe + BILSTM+AATT | 77.76 | 86.25 | 74.68 | 65.58 | 76.06 |
| GloVe +BILSTM+DATT | 77.10 | 86.29 | 76.00 | 62.93 | 75.58 |
| GloVe +BILSTM+AFFATT | 77.50 | 86.51 | 75.26 | 62.47 | 75.43 |
| Proposed | 78.30 | 86.66 | 76.85 | 65.76 | 76.90 |

As expected, because of the different distribution of each affection dimension in texts, also the outcome and accuracy of each model are different. Some personality traits have been affected more than others by each affection dimension. As shown in Table 4 and Table 5, when using the Dominance attention weight vector (DATT), E-I, N-S, and J-P prediction accuracy is improved than in other models. While applying the valence attention weight (VATT), performance in the T-F personality dimension has improved.

This paper has three significant findings: the first is that paying attention to text by the aspects of the emotion of words as three important dimensions of word meaning affected the performance of personality trait prediction in all dimensions and improved them. Although these effect for all type of personality trait prediction is not equal; It has a more significant impact on the prediction of some personality dimensions, such as **T**hinking-**F**eeling and **J**udging-**P**erceiving, which means these two personality traits have more related to emotional words. and the second is data preprocessing by 'PoS tagging' and extracting new stopwords by LDA help to reduce noise and improve the performance of models effectively. And finally, our proposed model has surpassed BERT's advanced and efficient model in terms of model prediction accuracy and time complexity reduction, as shown in Table 6.

**Table 6**

*Comparison of execution time and average accuracy scores proposed model with baseline models*

| Selection Models | Avg Acc | Execution Time (~) |
|---|---|---|
| GloVe +LSTM | 71.53 | 15 Minutes |
| Pretrained BERT (base) | 73.65 | 3515 Minutes |
| ENCODER+LSTM | 72.19 | 75 Minutes |
| GloVe +AFF+BILSTM | 72.85 | 20 Minutes |
| Proposed | 76.40 | 90 Minutes |

## 4. Discussion and Conclusion

Our proposed model achieved better performance in comparison with the state-of-the-art models. This paper encodes and revalues the features extracted with pre-trained GloVe by the BILSTM model for word representation. In parallel, an attention weight vector has been created based on each dimension of VAD dimensions separately. These affective attention weights have been applied to the BILSTM encoder for the attention mechanism layer. Three affective attention models have been fused to achieve maximum accuracy of models in all types of personality dimensions with various performances. We applied attention mechanism in the word level. Model evaluated on Kaggle MBTI dataset. Significantly, attending to each word based on its valence, arousal, and dominance level to semantical and contextual values have been achieved better performance than only concatenating three dimensions to semantical and contextual dimensions.

The average accuracy of our proposed model with the new preprocessing method improves by 4.41% compared to the LSTM and multi-encoder model and 3.35% in contrast with BERT. Novel preprocessing with the proposed model positively affects the model's performance and is suggested for all classification problems; this model could improve the performance of some resistant personality dimensions. In addition to, some personality dimensions, such as J-P and T-F, are improved more than others because of these two traits could be more correlated with emotional words.

## 5. Limitations & Suggestions





JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

While the proposed model demonstrates significant improvements in personality prediction accuracy and computational efficiency, it has several limitations. The reliance on the MBTI dataset, which is limited in size and diversity, may affect the model's generalizability to other datasets or real-world applications. Additionally, the focus on textual data from social media might not fully capture the multifaceted nature of personality traits, which are influenced by various non-textual cues such as behavior, tone, and context. The study also manually determined hyperparameters, which could be optimized further using automated methods. Finally, the computational benefits of the proposed model are achieved at the cost of excluding advanced contextual embeddings like those in larger transformer models, which might limit its applicability to tasks requiring deeper semantic understanding.

In this paper, we achieved worthwhile finding in personality prediction from text. We found the following significant implications:

1) removing noise effectively from social networks text that has a informal and noisy text improving performance.
2) semantical and contextual dimensional of word representation based on previous and next words cause the accurate meaning of words effect on accuracy, but sometimes we need to consider cost-benefit between computational resources and selective models. Our model has considered this issue.
3) Significant implication is adding valence, arousal, and dominance of words dimension to semantical dimension. Indeed, attention to words is based on the level of affection values in each dimension.
4) We found and verified the impact of three dimensions of the emotion of words, Valence (positiveness-negativeness/ pleasure-displeasure which is sentiment polarity), arousal (active-passive degree), and dominance (dominant-submissive that is active power) as extra information on the personality model. Indeed, we find the impact of these three dimensions of emotion–one of the building blocks of personality - as attention aspects of words in improving personality prediction.
5) Each dimension of valence, arousal, and dominance have a different distribution, which causes some personality traits to improve with applying each affection dimension. We find that using of ensemble algorithm for fusing models gets us an optimal final model that has stable

Creating the user profile by leveraging personality types helps businesses implement personalization and hyper-personalization strategies. Because personality refers to the differentiation of the individuals and how they perceive the world. By knowing users' personalities, we get a deep insight into customers. Our model use text generated by user, in this case, the social network, which is an enrichment resource for recognizing personality. Personality profiling is used in the recommender system, the personality-based chatbot, the personalized answer to customers via email or comment, and design products and advertisements. The trend of modern CRM and relational marketing is hyper-individualization. Personality is the primary key in this trend for perceiving customers and effective relations with them.

Considering that this paper presents a model for identifying the customer's personality in businesses, several research directions have emerged for us after conducting this research, which are our future research directions. Focusing on another aspect of information related to each personality trait could be a beneficial guideline. We will review the definition of personality to find another aspect of information. Indeed, we should pay attention to other new dimensions besides affective dimensions, such as the preferences of the user in texts. Then we will develop a personality model with three advantages: accurate, economic, and comprehensive. Second, we want to evaluate a short text dataset. Third, this paper manually determined and tested the model hyperparameter; In the future, we will use the optimization method for setting it. And last, because the dataset with labels in this field is rare, we will predict personality from text using unsupervised methods.

**Acknowledgments**

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

**Declaration of Interest**

The authors of this article declared no conflict of interest.

**Ethical Considerations**

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

**Transparency of Data**

The data that support the findings of this study are openly available in https://psy.takelab.fer.hr/datasets/all/ and available in Kaggle.

## Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

## Authors' Contributions

All authors equally contributed in this article.

## References


Abidin, N. H. Z., & Remli, M. A. A. N. M. P. D. N. E. Y. N. A. H. K. B. A. H. (2020). Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier. *Int. J. Adv. Comput. Sci. Appl.*, *11*(11). https://doi.org/10.14569/IJACSA.2020.0111125

Ahmad, H., Asghar, M. Z., Khan, A. S., & Habib, A. (2020). A systematic literature review of personality trait classification from textual content. *Open Comput. Sci.*, *10*(1). https://doi.org/10.1515/comp-2020-0188

Alsayat, A. (2022). Improving Sentiment Analysis for Social Media Applications Using an Ensemble Deep Learning Language Model. *Arab. J. Sci. Eng.*, *47*(2). https://doi.org/10.1007/s13369-021-06227-w

Artha Agastya, I. M. A., Handayani, D. O. D., & Mantoro, T. (2019). A Systematic Literature Review of Deep Learning Algorithms for Personality Trait Recognition. ICCED 2019,

Bauerhenne, C., Gammoudi, A., Moussaoui, M., Reichelt, J., & Wang, J. (2020). Emotional states and personality profiles in Conversational AI. https://www.mdsi.tum.de/en/di-lab/vergangene-projekte/horvath-emotional-states-and-personality-profiles-in-conversational-ai/

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.*, *3*(4-5). https://doi.org/10.1016/b978-0-12-411519-4.00006-9

Bradley, M. M., & Lang, P. J. (2008). Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings. https://pdodds.w3.uvm.edu/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf

Darliansyah, A., Naeem, M. A., Mirza, F., & Pears, R. (2019). Sentipede: A smart system for sentiment-based personality detection from short texts. *J. Univers. Comput. Sci.*, *25*(10). https://www.researchgate.net/publication/354074066_SENTIPEDE_A_Smart_System_for_Sentiment-based_Personality_Detection_from_Short_Texts

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, https://aclanthology.org/N19-1423/

El-Demerdash, K., El-Khoribi, R. A., Ismail Shoman, M. A., & Abdou, S. (2022). Deep learning based fusion strategies for personality prediction. *Egyptian Informatics Journal*, *23*(1). https://doi.org/10.1016/j.eij.2021.05.004

Galitsky, B. (2021). Adjusting Chatbot Conversation to User Personality and Mood. 2021,

Graves, A., & Schmidhuber, J. (2005). Framewise Phoneme Classification with Bidirectional LSTM Networks.

Grgic, D., & Podobnik, V. (2021). Application of Data Science for Understanding Emotional Dimensional Behavior and Their Connection to Uncertainty and Risk Behavior. *IEEE Access*, *9*. https://doi.org/10.1109/ACCESS.2021.3079877

Han, S., Huang, H., & Tang, Y. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Syst.*, *194*. https://doi.org/10.1016/j.knosys.2020.105550

Hassanein, M., Hussein, W., Rady, S., & Gharib, T. F. (2019). Predicting Personality Traits from Social Media using Text Semantics.

Heikal, M., Torki, M., & El-Makky, N. (2018). Sentiment Analysis of Arabic Tweets using Deep Learning. Procedia Computer Science,

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int. J. Uncertainty, Fuzziness Knowlege-Based Syst.*, *6*(2). https://doi.org/10.1142/S0218488598000094

Hochreiter, S., & Schmidhuber, J. (1997). Long Short Term Memory. *Neural Comput.*, *9*(8). https://doi.org/10.1162/neco.1997.9.8.1735

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. 32nd International Conference on Machine Learning, ICML 2015, https://dl.acm.org/doi/10.5555/3045118.3045167

Jacovi, A., Shalom, O. S., & Goldberg, Y. (2018). Understanding Convolutional Neural Networks for Text Classification.

Jeremy, N. H., & Suhartono, D. (2021). Automatic personality prediction from Indonesian user on twitter using word embedding and neural networks. Procedia Computer Science,

Kim, Y., Denton, C., Hoang, L., & Rush, A. M. (2017). Structured attention networks. https://www.semanticscholar.org/paper/Structured-Attention-Networks-Kim-Denton/13d9323a8716131911bfda048a40e2cde1a76a46

Lynn, V., Balasubramanian, N., & Schwartz, H. A. (2020). Hierarchical Modeling for User Personality Prediction: The Role of Message-Level Attention. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics,

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.

Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep Learning-Based Document Modeling for Personality Detection from Text. *IEEE Intell. Syst.*, *32*(2). https://doi.org/10.1109/MIS.2017.23

Mehrabian, A. (1996a). Analysis of the big-five personality factors in terms of the PAD temperament model. *Aust. J. Psychol.*, *48*(2). https://doi.org/10.1080/00049539608259510

Mehrabian, A. (1996b). Pleasure-Arousal.Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. https://doi.org/10.1007/BF02686918

Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2020). Recent trends in deep learning based personality detection. *Artif. Intell. Rev.*, *53*(4). https://doi.org/10.1007/s10462-019-09770-z

Mehu, M., & Scherer, K. R. (2015). Emotion categories and dimensions in the facial communication of affect: An






JARAC Assessment and Research in Applied Counseling E-ISSN: 3041-8518

integrated approach. *Emotion*, *15*(6). https://doi.org/10.1037/a0039416
Mohammad, S. M. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers),
Mohammadi, A., & Shaverizade, A. (2021). Ensemble deep learning for aspect-based sentiment analysis. *Int. J. Nonlinear Anal. Appl.*, *12*(Special Issue). https://doi.org/10.22075/IJNAA.2021.4769
Mohammed, A., & Kora, R. (2021). An effective ensemble deep learning framework for text classification. *J. King Saud Univ. - Comput. Inf. Sci.* https://doi.org/10.1016/j.jksuci.2021.11.001
Myers, I. B., & Myers, P. B. (1995). *Gifts Differing: Understanding Personality Type*. https://books.google.de/books/about/Gifts_Differing.html?id=WfR8DAAAQBAJ&redir_esc=y
Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep learning. *Neurocomputing*, *452*. https://doi.org/10.1016/j.neucom.2021.03.091
Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.*, *57*. https://doi.org/10.1146/annurev.psych.57.102904.190127
Parkinson, B., & Manstead, A. S. R. (2015). Current Emotion Research in Social Psychology: Thinking About Emotions and Other People. *Emot. Rev.*, *7*(4). https://doi.org/10.1177/1754073915590624
Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). *Operator's Manual: Linguistic Inquiry and Word Count - LIWC2007*. Depts.Ttu.Edu. http://www.gruberpeplab.com/teaching/psych231_fall2013/documents/231_Pennebaker2007.pdf
Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation.
Remaida, A., Abdellaoui, B., Moumen, A., & El Idrissi, Y. E. B. (2020). Personality traits analysis using Artificial Neural Networks: A Literature Survey.
Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. *Inf. Process. Manag.*, *58*(3). https://doi.org/10.1016/j.ipm.2021.102532
Reuderink, B., Mühl, C., & Poel, M. (2013). Valence, arousal and dominance in the EEG during game play. *Int. J. Auton. Adapt. Commun. Syst.*, *6*(1). https://doi.org/10.1504/IJAACS.2013.050691
Rokach, L. (2019). *Ensemble Learning: Pattern Classification Using Ensemble Methods*. https://doi.org/10.1142/11325
Russell, J. A. (1980). A circumplex model of affect. *J. Pers. Soc. Psychol.*, *39*(6). https://doi.org/10.1037/h0077714
Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychol. Rev.*, *110*(1). https://doi.org/10.1037/0033-295X.110.1.145
Štajner, S., & Yenikent, S. (2021). Why is MBTI personality detection from texts a difficult task?
Sun, X., Liu, B., Cao, J., Luo, J., & Shen, X. (2018). Who am I? Personality detection based on deep learning for texts. IEEE International Conference on Communications,
Teragawa, S., & Li, M. (2020). Text Classification based on Model of Feature Fusion and Voting Mechanism for Personality Recognition.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, M. (2017). Attention is all you need. Advances in Neural Information Processing Systems, https://dl.acm.org/doi/10.5555/3295222.3295349
Wang, H., Zuo, Y., Li, H., & Wu, J. (2021). Cross-domain recommendation with user personality. *Knowledge-Based Syst.*, *213*. https://doi.org/10.1016/j.knosys.2020.106664
Wang, Z., Wu, C. H., Li, Q. B., Yan, B., & Zheng, K. F. (2020). Encoding Text Information with Graph Convolutional Networks for Personality Recognition. *Appl. Sci.*, *10*(12), 4081. https://doi.org/10.3390/app10124081
Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behav. Res. Methods*, *45*(4). https://doi.org/10.3758/s13428-012-0314-x
Wen, Z., Cao, J., Yang, R., Liu, S., & Shen, J. (2021). Automatically Select Emotion for Response via Personality-affected Emotion Transition. Findings of the Association for Computational Linguistics: ACL
Winter, S., Maslowska, E., & Vos, A. L. (2021). The effects of trait-based personalization in social media advertising. *Comput. Human Behav.*, *114*. https://doi.org/10.1016/j.chb.2020.106525
Xiang, R., Long, Y., Wan, M., Gu, J., Lu, Q., & Huang, C. R. (2020). Affection driven neural networks for sentiment analysis. https://aclanthology.org/2020.lrec-1.14/
Xiang, R., Lu, Q., Zhang, Z., & Huang, C. R. (2021). Affective awareness in neural sentiment analysis. *Knowledge-Based Syst.*, *226*. https://doi.org/10.1016/j.knosys.2021.107137
Xue, D., L., W., Z., H., S., G., L., G., Z., W., X., Z., & J., S. (2018). Deep learning-based personality recognition from text posts of online social networks. *Appl. Intell.*, *48*(11). https://doi.org/10.1007/s10489-018-1212-4
Yang, H. C., & Huang, Z. R. (2019). Mining personality traits from social messages for game recommender systems. *Knowledge-Based Syst.*, *165*. https://doi.org/10.1016/j.knosys.2018.11.025
Yang, K., Lau, R. Y. K., & Abbasi, A. (2022). Getting Personal: A Deep Learning Artifact for Text-Based Measurement of Personality. *Inf. Syst. Res.* https://doi.org/10.1287/isre.2022.1111
Yuan, C., Wu, J., Li, H., & Wang, L. (2018). Personality Recognition Based on User Generated Content.
Zhao, J., Zeng, D., Xiao, Y., Che, L., & Wang, M. (2020). User personality prediction based on topic preference and sentiment analysis using LSTM model. *Pattern Recognit. Lett.*, *138*. https://doi.org/10.1016/j.patrec.2020.07.035




E-ISSN: 3041-8518