Hindawi
Computational and Mathematical Methods in Medicine
Volume 2021, Article ID 5512241, 10 pages
https://doi.org/10.1155/2021/5512241

Research Article

Detection and Classification of Psychopathic Personality Trait from Social Media Text Using Deep Learning Model

Junaid Asghar,[1] Saima Akbar,[2] Muhammad Zubair Asghar,[2] Bashir Ahmad,[3] Mabrook S. Al-Rakhami,[4] and Abdu Gumaei[4,5]

[1]*Faculty of Pharmacy, Gomal University, D.I. Khan (KP), Pakistan*
[2]*Institute of Computing and Information Technology, Gomal University, D.I. Khan (KP), Pakistan*
[3]*Dept. of Computer Science, Qurtaba University, D.I. Khan (KP), Pakistan*
[4]*Research Chair of Pervasive and Mobile Computing; Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia*
[5]*Computer Science Department, Faculty of Applied Sciences, Taiz University, Taiz 6803, Yemen*

Correspondence should be addressed to Mabrook S. Al-Rakhami; malrakhami@ksu.edu.sa

Received 20 February 2021; Revised 20 March 2021; Accepted 26 March 2021; Published 10 April 2021

Academic Editor: Waqas Haider Bangyal

Nowadays, there is a digital era, where social media sites like Facebook, Google, Twitter, and YouTube are used by the majority of people, generating a lot of textual content. The user-generated textual content discloses important information about people's personalities, identifying a special type of people known as psychopaths. The aim of this work is to classify the input text into psychopath and nonpsychopath traits. Most of the existing work on psychopath's detection has been performed in the psychology domain using traditional approaches, like SRPIII technique with limited dataset size. Therefore, it motivates us to build an advanced computational model for psychopath's detection in the text analytics domain. In this work, we investigate an advanced deep learning technique, namely, attention-based BILSTM for psychopath's detection with an increased dataset size for efficient classification of the input text into psychopath vs. nonpsychopath classes.

1. Introduction

According to psychology, traits provide a way of describing a person, such as generous, out-going, and short-tempered. The trait-driven approach is the most focused area in psychology literature. Trait depicts a person's characteristic to make a response and reaction to a certain situation in a specific way [1].

Perceive them as being cunning, antisocial, and manipulative, and such individuals merely exist about 1% of the population [2]. Individuals with psychopathic behavior usually feel no discomfort or guilt while making immoral choices and displaying immoral behavior and committing immoral actions. Hervey Cleckley was the first person who officially coined the concept of psychopathy in 1964 by identifying a group of his patients lacking morality. The modern ideology is about psychopath actions [2]. In connection to the dark triad, psychopathy is termed as the inherent feature of dark triad person to show no regrets for displaying immoral and harmful behavior, because they lack empathy and guilt, while committing wrong and sinful and wicked deeds. Medical science declares such persons as being mentally disordered and harmful to both themselves and society with a high tendency of instability [3]. In addition to their antisocial and criminal behavior, such individuals have certain distinctive features, such as lack of realistic goals, parasitic lifestyle, and lack of responsibility. They are incapable of learning from their experiences and cannot establish fruitful relationships. Being emotionally immature, such individuals are not affected by punishments and continue with their antisocial behavior [4].

1.1. Research Study Motivation. The traditional techniques of identifying psychopaths include the following: Psychopathy Checklist-Revised, PCL-R, Welsh Anxiety Scale [5], PCL-R, Wmatrix linguistic analysis, Dictionary of Affect and

language (DAL) [2], Self-Report Psychopathy Test III (SRP-III) [2], Hare Psychopathy Checklist, Wonderlic Personnel Test, Wide Range Achievement Test [6], and Levenson Self-Report psychopathy scale (LSRP) [7].

Few studies have applied computational models for psychopath detection, covering machine learning and deep learning-based techniques. The machine learning techniques include Support Vector Machine (SVM), Random Forest, Logistic Regression, Multilayer Perceptron (MLP), Random Forest (RF) 100, SVM [8], Naïve Bayes (NB), Decision Tree (DT), and SVM [9], Logistics Regression [10], SVM [11], and clustering [5]. While the deep learning techniques include MLP, Long Short Term Memory(LSTM), Gated Recurrent Unit(GRU), Convolutional Neural Network-1 Dimensional (CNN-1D) [12], and Face++, EmoVu [13].

The rapid evolution of social media networks like Facebook, Twitter, and YouTube has allowed users to communicate information by interacting with the community. Social media's users exploit their status updates, images, text, and public profiles to express themselves [14]. Moreover, the important information on such sites pertaining to individuals assist in the detection and authentication of their personality traits [10]. Additionally, it is observed that the content available on such networks can also be acquired and analyzed to identify individuals with psychopathic behavior.

This work is aimed at the development of an automated method that can filter psychopaths from nonpsychopaths by using textual content available on social media sites. The study conducted by [8] applied different machine learning classifiers for predicting psychopathy from Twitter content. This baseline study used a limited dataset and applied a classical feature set by using a traditional machine learning classifier, which results in poor performance in terms of low classification accuracy.

The aforementioned limitations can be overcome by exploiting BISLTM for classifying reviews into psychopath and nonpsychopath classes. BiLSTM layer retains both the past and future context information using long-term dependency in a review in a sentence to predict the sentence class.

1.2. Problem Statement. In this work, we address the problem of psychopath personality detection from text. To distinguish psychopath review from nonpsychopath, the psychopath classification problem is taken as binary classification task. The training dataset Tr is a set of reviews $\{r1, r2, r3, \cdots .rn$, with the class tag is_psychopath = {yes, no}. Each review is given a class label. The objective is to build a model, which can learn from training dataset and classify a new review as psychopath or nonpsychopath.

1.3. Research Questions. RQ.1: how to classify the input text into psychopath and nonpsychopath class by applying deep learning technique, namely, BILSTM?

RQ.2: what is the performance of the proposed deep learning system, namely, BILSTM as compared to different machine learning and deep learning techniques?

RQ.3: how to evaluate the effectiveness of the proposed system w.r.t the baseline techniques.

The rest of the article is organized as follows: Section 2 presents a review of literature; Section 3 portrays proposed methodology; results and discussion are presented in Section 4; and finally, Section 5 presents conclusions and future work.

2. Review of Literature

This section provides a review of the previous works conducted on psychopath's detection.

In their work on identifying the relationship between personality and online behavior, Whitty et al. [10] proposed a novel approach using questionnaires with self-rating items. Experimental results show confirmation of their proposed hypothesis. However, the inclusion of profile images can improve the performance of the system.

The student personality profiles were collected, and feedback was analyzed with respect to their learning experiences [9]. An empirical study was conducted by collecting data from 115 college students. The results depict that personality-based student profiling can act as an effective rule for developing an intelligent feedback analysis system. The limitations of the study include low frequency for particular emotion states, such as anger and delight.

Bukhtawer et al. [15] [R2.5] identified the relation among personality features and drug abuser's in their different phases of education. The data were acquired using different questionnaire templates. The results obtained reveal that self-regulation compared with personality features of drug abusers provide a new avenue in this dimension of research. However, exploring specific personality traits, such as neuroticism and impulsively, if incorporated, can provide interesting results.

Keshtkar et al. [16] applied different data mining and natural language techniques for automatic detection of student's personality traits in an educational game. Input text is classified into six personality traits, and after that, various machine learning algorithms, such as SVM, Naive Bayes, and decision tree, are applied. An N-grams feature model has performed well with different classifiers.

Liu et al. [13] in their work on personality recognition from users social media accounts focused on esthetic and facial features in images. For this purpose, a Bid Five model is trained over 7000 Facebook users using the N-gram feature model, and the personality traits are predicted using linear regression and Elastic Ne regulation with 10-fold cross-validation. Promising results are obtained with respect to baseline methods. However, extended psychological traits and wider collection of user's profile pictures can improve the performance of the system. The language used on Twitter has a relationship with the personality traits of the user. Data set of 90 k users is collected from Twitter. After applying the preprocessing technique, two machine learning algorithms are used, namely, (i) Logistic Regression model, (ii) Random Forest, and (3) SVM. The proposed system achieved a predictive accuracy of 0.661.

YouTube-based personality recognition is performed by [17] using different multiple regression techniques including tree-based models that were applied. For this purpose, 25

multimedia features and 101 textual features were used. This system produced promising results with respect to comparing methods.

Pednekar and Dubey [18] proposed a theoretical framework for identifying personality traits from social media profiles by considering different instruction's commented by the user such as likes, dislikes, comments, and sharing. The text is classified using decision tree classifier over a preposition of personality traits taken from Big Five model.

Khan et al. (2017) developed a personality item pool in Urdu language using a well-known translation model called Darwish. The proposed questionnaire is an important tool in user's own language, i.e., Urdu. The internal consistent checks show that Urdu is more consistent than English.

The aim of this study [19] is to predict the individual's personality through handwriting. For this purpose, 42 handwriting samples are used. SVM technique is used to achieve desired results. The study achieved 82.738% accurate results.

Using Dominance, Influence, Compliance, and Steadiness assessments, Ahmad and Siddique [5] suggested a method for predicting a user's personality based on knowledge mapping available to the public on their personal Twitter. The proposed methods provide satisfactory results with respect to baseline methods. However, images can be incorporated for personal detection and obtained robust results.

Tandera et al. [12] proposed the system to predicate user's personality for Facebook profile. For this, purpose Big Five personality model is used by implementing used deep learning technique. They achieved an accuracy of 74.17% and outperformed the comparing methods.

In their work on Extraversion personality traits, Shaheen et al. [20] analyzed the association among HIV-affected persons with respect to extraversion personality feature. The results show that there is a strong correlation between HIV-affected patients and extraversion personality traits.

The purpose of the work performed by (Hancock et al. [2] is to investigate the psychopathic homicide offenders crime narrative features. Different statistical text analysis tools were used, namely, (i) Psychopathy Checklist-Revised (PCL-R) to learn psychopathy; (ii) Wmatrix linguistic analysis tool to inspect semantic content and parts of speech; and (iii) Dictionary of Affect and language (DAL) tool to inspect the emotional attributes about narratives. According to the results, psychopath speech covers extra rational cause-and-effect descriptors, large occurrence regarding disfluencies, and usage of present tense in small amount with past tense in large amount. Some limitations of the work are the usage of only homicide event in narrative, and the short movie of the event is unavailable. The future focus is on reporting variant emotional and nonemotional events by analyzing the psychopath speech creation attributes and also to explore further varieties of emotional stimuli that contain the ground truth about the event.

Hancock et al. [2] used a Self-Report Psychopathy Test III (SRP-III) to analyze whether there is a relation of psychopathy characteristics with the linguistic pattern made in day-to-day online discussion, such as text messaging, Facebook, and email. The results show that within online discourse, the psychopathy linguistics traces can be uncovered. Moreover, the individuals with large psychopathy attributes are unsuccessful in language alteration across online communication. The major drawbacks of the study include limited sample size ($N = 110$), and a small amount of language samples gathered. However, the addition of the individuals that acquire high SRP-III score will improve the results.

To identify the psychopathy using Twitter account information, Wald et al. [8] proposed ensemble learning, namely, SelectRUSBoost by applying four classification algorithms, namely, LR, MLP, RF100, SVM, and also implementing four feature selection approaches, namely, Kolmogorov-Smirnov statistic (KS), Mutual Information (MI), Area Under the ROC Curve, and Area Under the PRC Curve. The experiments reveal that the best result (AUC = 0.736) is achieved by applying Select RUS Boost plus SVM kernel, with MI feature over a 20 sample size. In the future, it is needed to add the increase of users in the dataset, including more personality attributes, applying further variant techniques, and lastly, using other Twitter datasets or other social media site content.

Le et al. [21] used a primary assessment tool, namely, PCL-R to investigate the discussion/language of offenders during the interview. Additionally, LIWC text analysis software program and Wmatrix method are used. Furthermore, to find the best predictors of psychopathy scores, the regression analysis is performed on the linguistic classes. The results show that psychopathy has limited emotion expression, create less references to others, and use more disfluencies and personal pronoun. Also, less occurrence of anxiety-related words and more repetition of personal pronoun are important predicators regarding PCL-R scores. Some limitations of the work include (i) discussion written from face-to-face interview are used for organizing statistical text analysis, (ii) unavailability of factors F1 and F1 scores for the present samples, and (iii) awareness of the offenders that they are used for evaluation.

3. Materials and Methods

This section describes the proposed methodology developed for classifying the text into psychopaths and nonpsychopath classes using deep learning model, namely, BiLSTM. The job of BiLSTM is to store the past information using Backward LSTM and future information using Forward LSTM [22]. The proposed model (Figure 1) is comprised of different modules, namely, (i) data collection that aims to acquire and annotate the required dataset, (ii) preprocessing that applies some cleansing operation on the noisy textual content, and (iii) applying deep learning model (BiLSTM) to perform the final classification of the given text.

The detail of each module is presented as follows.

3.1. Data Collection. The first module of the proposed methodology is comprised of two steps.

In this step, we acquired the required dataset from different social media sites like Facebook and Twitter. For example, we used the hashtag "#psychopath" to crawl required tweets using a Python-based library, namely, Tweepy [23]. The collected data is stored in the Excel file for further process. To annotate each review text/tweet in the acquired

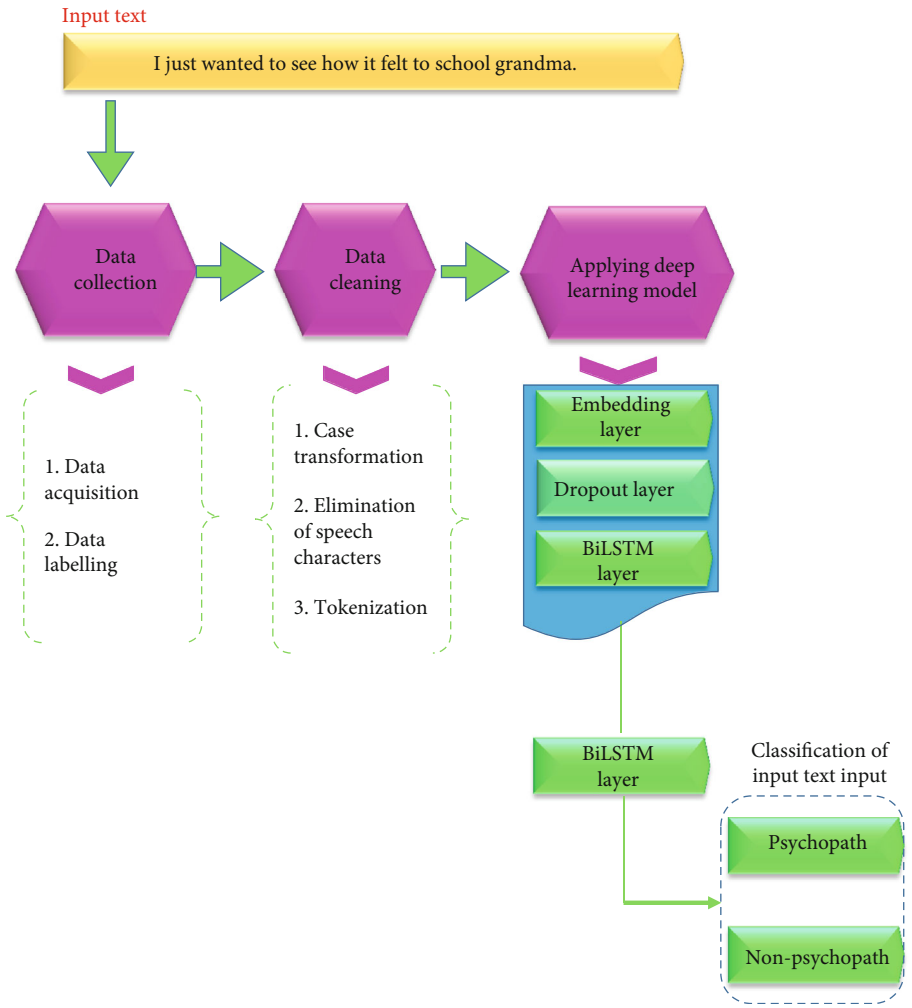

Figure 1: Proposed system.

dataset into psychopath and nonpsychopath classes, we performed manual annotation by assigning the task to three human annotators (psychiatrists), each of them assigned a class label: "psychopath" or "non-psychopath". In this way, we received three votes for each tweet. The class label is selected on the basis of majority voting scheme [24], i.e., a tweet having two votes for "psychopath" and one for "non-psychopath," is tagged as "psychopath." Furthermore, the required dataset contains 601 user input text samples.

To conduct the experiments, first, we split the dataset into two segments, namely, train set and test set by exploiting sklearn train_test_split method [25]. The percentage of train is 90% and test set is 10%.

In Table 1, the description of the given dataset is presented.

3.2. Data Cleaning. In data cleaning task, the cleaning of the data is performed to maintain the original input text. It helps to enhance the accuracy of the text classification process. The reason for applying the data cleaning module is that the user input text in real-world contain a significant amount of noise, so it is needed to clean the data from this noise in order to perform different NLP tasks (text classification) [26]. The data cleaning step is aimed at performing the following tasks on the acquired dataset.

3.2.1. Case Transformation. In this technique, the input text is transformed into lower case by using python based Script.

3.2.2. Elimination of Special Characters. Some special characters like "#," "%," "?," "@," "-," "&," "$," "/," and "∗" are eliminated from the input text during this module that helps to overcome dimensionality reduction [27].

3.2.3. Tokenization. The objective of the tokenization process is to convert the input text into small tokens/pieces. To perform tokenization, we used keras tokenizer [28].

3.3. Applying Deep Learning Model. In this segment, we describe different layers used in the proposed deep neural network model called BILSTM, for the detection of psychopath from input text as shown in Figure 1.

3.3.1. Embedding Layer. This layer performs word embedding using keras embedding layer. The purpose of this layer

Table 1: Dataset description.

Dataset	User reviews	Personality class
Personality data	601	Psychopath (300), nonpsychopath (301)

is to present a word-level representation in which word indices are transformed into embedding.

3.3.2. Dropout Layer. The aim of this layer is to prevent overfitting. The rate parameter of the dropout layer is set to the specified threshold (i.e., 0.7), where its range lies between 0 and 1. This layer is placed after the embedding layer to control the random activation of neurons in the Embedding layer.

3.3.3. Bidirectional LSTM Layer. The BiLSTM layer acts as a second layer of the proposed model that receives input from the embedding layer and then transforms it into new encoding. The BiLSTM performs 2-way encoding by maintaining not only the previous but also the future information.

3.3.4. Output Layer. Finally, the activation function of softmax is used at the output layer for performing the classification task. In this layer, the input text is classified into psychopaths and nonpsychopath's sentences.

3.4. How the Proposed System Works. The proposed architecture of the BILSM model for personality detection into binary classes such as psychopath and nonpsychopath exploits five major phases: (i) embedding layer, (ii) dropout layer, (iii) BILSTM layer, and (iv) output layer.

Initially, a sample input text "I just wanted to see how it felt to shoot grandma," is taken, then, it is moved sequentially through the different layers of the proposed deep neural network model. Each layer is elaborated in the following way.

3.4.1. Integer Encoding via Tokenizer. The sample input text needs to be prepared in numerical form so that the deep learning model can be applied. In this connection, the numerical representation of the given input text starts with the tokenization process. The keras tokenizer method "tokenizer.fit_on_texts" is used to that attaches an integer value to each individual word, such as ["I:1" "just:2" "wanted:3" "to:4" "see:5" "how:6" "it:7" "felt:8" "to:9" "shoot:10" "grandma:11"]. Moving forward, the input text is converted into integers sequence like [1,2,3,4,5,6,7,8,9,10,11] using another keras tokenizer method "tokenizer.text_to_sequences." Finally, the input is prepared and made input to the initial layer of the proposed deep learning model.

3.4.2. Word Embedding Vector via Embedding Layer. The primary layer of the model translates already obtained individual integers into low dimensional feature (embedding) vectors that assist in taking syntactic and semantic information. For instance, a sample word "**I**" having an integer number "**1**" holds the following feature vector [0.1,0.3,0.5,0.4]. As a result, the output of this layer is a word embedding matrix.

3.4.3. Dropout Rate via Dropout Layer. The next layer is the dropout layer that comes after the embedding layer. The values of the dropout rate parameter ranges between [0,1]. Its job is to reduce the problem of overfitting.

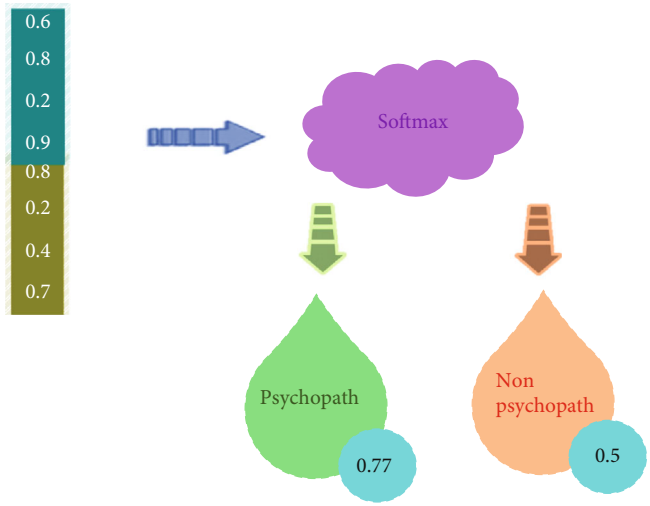

Figure 2: Applying softmax function at **o**utput layer.

3.4.4. Capturing Context Information via BILSTM Layer. We employ deep learning technique, namely, BiLSTM to classify the text into psychopaths and nonpsychopath's sentences. For capturing both syntactic and semantic information of a sentence, Long Short-Term Memory (LSTM) neural network has shown considerable performance improvement. The functionality of Bidirectional Long Short-term memory (Bi-LSTM) architecture includes the processing of text in both directions using dual LSTM layer, which allows us to capture both previous and subsequent context of a given sentence [29]. In this next section, we describe how a Bi-LSTM classifier can be applied for psychopath detection in a given text.

Bi-LSTM model has gained much attention recently due to its superior ability to maintain sequence information by considering both past and future context, as both the contexts have equal importance [30]. It has assisted in overcoming the limitation of Conventional RNN, which can only retain information for a short time interval, and unidirectional LSTM keeps track of only past context [31]. We choose to use the Bi-LSTM, because it uses two separate hidden layers to classify input text into two classes, i.e., "psychopath" and "non- psychopath."

The BiLSTM network is composed of two subnetworks: Forward LSTM and Backward LSTM, respectively [27], as shown in Figure 2. Given an input sequence $x_1, x_2, x_3 \cdots \cdots, x_n$, of "$n$" words, the BiLSTM computes the forward hidden vector "$\overrightarrow{w}$" and the backward hidden vector "$\overleftarrow{w}$". By concatenating the right and left context representations: $\overleftrightarrow{w} = [\overleftarrow{w}, \overrightarrow{w}]$, an output sequence $w_1, w, w_3, \cdots, w_t$, is generated

that is further given as an input to the output layer to get predictions for each sentence/tweet [32].

The following equations are used for the computation of both Forward and Backward LSTM.

Forward LSTM equations:

$$\vec{r}_t = (A_r.x_t + B_r.w_{t-1,} + v_r), \tag{1}$$

$$\vec{s}_t = (A_s.x_t + B_s.w_{t-1,} + v_s), \tag{2}$$

$$\vec{u}_t = (A_u.x_t + B_u.w_{t-1,} + v_u), \tag{3}$$

$$\vec{l}\sim_t = (A_{l\sim}.x_t + B_{l\sim}.w_{t-1,} + v_{l\sim}), \tag{4}$$

$$\vec{l}_t = s_t \otimes l\sim_t + r_t \otimes l_{t-1}, \tag{5}$$

$$\vec{w}_t = (l_t) \otimes u_t. \tag{6}$$

Backward LSTM equations:

$$\overleftarrow{r}_t = (A_r.x_t + B_r.w_{t+1,} + v_r). \tag{7}$$

$$\overleftarrow{s}_t = (A_s.x_t + B_s.w_{t+1,} + v_s). \tag{8}$$

$$\overleftarrow{u}_t = (A_u.x_t + B_u.w_{t+1,} + v_u). \tag{9}$$

$$\overleftarrow{l}\sim_t = (A_{l\sim}.x_t + B_{l\sim}.w_{t+1,} + v_{l\sim}). \tag{10}$$

$$\overleftarrow{l}_t = s_t \otimes l\sim_t + r_t \otimes l_{t+1}. \tag{11}$$

$$\overleftarrow{w}_t = (l_t) \otimes u_t, \tag{12}$$

where r, s, u are the forget, input, and output gates. x_t is input state, w_t is hidden state (output) that is sent to the further LSTM layer within the network, l_t is the current cell state, v is called bias, and A and B denote the weight metrics, respectively.

3.5. Predicting Label via Output Layer. The final layer applies softmax activation function to predict the class tag probability like "psychopath" and "non-psychopath" by exploiting Eq. (13).

After passing through the softmax activation function, it is found that the "psychopath" class tag achieved the maximum probability, so the given input text "" is labeled as "psychopath" (see Figure 2)

$$\text{softmax}(l_i) = \frac{e^{l_i}}{\sum_{j=1}^{n} e^{l_j}}. \tag{13}$$

Algorithm 1 represents the pseudocode for personality detection.

4. Results and Discussion

In this section, we present results and their analysis on account of conducting different experiments in response to research questions formulated in Section 1.

4.1. Answer to RQ1: How to Classify the Input Text into Psychopaths and Nonpsychopaths Class by Applying Deep Learning Technique, Namely, BILSTM? In order to classify text into psychopath and nonpsychopath type, we have used different Bi-LSTM models using varying parameters. The parameter setting for the proposed Bi-LSTM model is shown in Table 2. The experimentation is performed with different parameters: a number of Bi-LSTM unit is varied from 100 to 350, batch size with value 32, and vocabulary size of 2000. Similarly, we have also used certain parameters with fixed sizes, such as input vector size, activation function, embedding dimension, and the number of epochs.

We conducted different experiments to answer RQ1; detail is given as follows.

4.1.1. Experiment 1. We conducted an experiment with varying parameter settings of different Bi-LSTM models. Table 3 depicts the estimation metrics (recall, f-score, and precision) results regarding different Bi-LSTM models. It is observed that Bi-LSTM(4) model achieves maximum accuracy of 85% and performs better as compared to other Bi-LSTM models with parameters (batch size = 32, Bi − LSTM unit = 250, and vocabulary size = 2000). It is found through experiments that first, the model accuracy tends to decrease until Bi-LSTM(3), i.e., 0.74%. As we reach Bi-LSTM(4), the accuracy increases with a value of 0.85%, after that, the model accuracy stabilizes with the value of 0.80, as we increase the unit size that is 300, 350.

During experimentation, we recorded the test accuracy, loss score, and training time for all the Bi-LSTM models with different parameter settings, as listed in Table 4.

(1) Justification behind Our Better Results. It is found that the test accuracy decrements up to Bi-LSTM (3) model, whereas the test loss increments (1.39, 1.41, 1.56). As we reach Bi-LSTM (4), the accuracy is incremented because of the decrement (1.39) in the test loss. Finally, the model accuracy gets stabilized at unit size 300 and 350.

4.2. Answer to RQ2: What Is the Performance of the Proposed Deep Learning System, Namely, Attention-Based BILSTM as Compared to Different Machine Learning and Deep Learning Techniques? To answer RQ2, we conducted different experiments on various machine learning classifiers and the proposed deep learning model. Detail is presented in the following subsections.

4.2.1. Comparison with Machine Learning Techniques. To perform comparison of the proposed BI-LSTM model with different machine learning classifiers, we implemented each ML classifiers on the acquired dataset. The performance evaluation results are presented in Table 5. A number of machine learning classifiers are applied, namely, Decision Tree, SVM, KNN, MNB, LR, RF, and XGBoost. It is found that the performance of the XGBoost classifier is higher with an accuracy of 77.05%, while the lowest accuracy (65.57%) is obtained by the LR classifier.

Pseudocode regarding personality detection in input text by exploiting BILSTM.
Input: Personality dataset "E", Train Set "TRS", Test Set "TES"
Output: Personality label regarding input text: Psychopath vs Nonpsychopath
Start
Section 1: Numeric representation of input text
1. **while** each input text T∈ E
2. **while** word w∈ E
3. Allocate index to related word
4. **End while**
5. **End while**
Hyperparameter Initialization
6. train set size=90%, test set size=10%, max-features=2000, embed_dim=128, batch_size=32, epochs=7
Section 2: Developing Deep Neural Network Model
7. **while** each input text T∈ E_{TRS}
8. Create embedding vector of entire words in T = [t_1, t_2, t_3, t_4, … , t_m] //Convert text to machine readable feature(word) vector
9. Apply dropout layer for overfitting reduction $d \in (g, h) = \begin{cases} h \text{ if } g = 1 \\ 1 - h \text{ if } g = 0. \end{cases}$
10. Apply operation of BILSTM using Eq. ((1)-(12), (13)
11. **End while**
Section 3: Evaluating the Model
12. **while** each input text T∈ E_{TES}
13. Developed a Train Model
Apply softmax operation (using Eq. (13) to classify the input text into Psychopath vs Nonpsychopath
14. End while
Terminate

Algorithm 1:

Table 2: Parameters settings of the proposed Bi-LSTM model.

Parameter	Value
Input vector size	100
Vocabulary size	2000
Embedding dimension	128
Bi-LSTM unit size	100, 150, 200, 250, 300, 350
Number of hidden layers	2
Activation function	Softmax
Number of epochs	7
Batch size	32

Table 3: Estimation metrics results of Bi-LSTM models.

Model name	Recall	*F*1-score	Precision
Bi-LSTM(1)	0.84	0.84	0.84
Bi-LSTM(2)	0.82	0.82	0.82
Bi-LSTM(3)	0.74	0.74	0.74
Bi-LSTM(4)	0.85	0.85	0.85
Bi-LSTM(5)	0.80	0.80	0.80
Bi-LSTM(6)	0.80	0.80	0.80

Table 4: Test accuracy, loss, and training time of Bi-LSTM models.

Model name	Test accuracy	Test loss	Training time(s)
Bi-LSTM(1)	0.84%	1.39	5 s
Bi-LSTM(2)	0.82%	1.41	8 s
Bi-LSTM(3)	0.74%	1.56	15 s
Bi-LSTM(4)proposed	0.85%	1.39	24 s
Bi-LSTM(5)	0.80%	1.42	49 s
Bi-LSTM(6)	0.80%	1.44	39 s

Table 5: Comparison with machine learning techniques.

Machine learning classifiers and proposed model	Accuracy	Precision	Recall	*F*1 -score
Decision tree	75.41%	0.76%	0.77%	0.75%
SVM	72.13%	0.76%	0.76%	0.72%
KNN	70.49%	0.73%	0.70%	0.71%
MNB	68.85%	0.76%	0.74%	0.69%
LR	65.57%	0.65%	0.66%	0.65%
RF	73.77%	0.76%	0.74%	0.74%
XGBoost	77.05%	0.77%	0.77%	0.77%
Proposed (BiLSTM)	85%	85%	85%	85%

It is obvious that the proposed DL model outperformed different ML classifiers in terms of better precision (0.85%), recall (0.85%), *F*-score (0.85%), and accuracy (0.85%).

(1) Justification behind Our Better Results. The feature representation scheme exploited in the proposed DL model is word embedding that stores the syntactic as well as sematic information related to the given word. However, the ML classifiers exploited the traditional feature representation scheme, namely, BOW (TF-IDF, CountVectorizer) that only stores the syntactic information without tackling the sematic

Table 6: Comparison with deep learning techniques.

Deep learning classifiers	Accuracy	Precision	Recall	$F1$-score
CNN	0.69%	0.74%	0.69%	0.69%
LSTM	0.79%	0.79%	0.79%	0.78%
GRU	0.82%	0.82%	0.82%	0.82%
RNN	0.72%	0.75%	0.72%	0.72%
Proposed (BiLSTM)	0.85%	0.85%	0.85%	0.85%

information due to the one hot vector representation of words. Therefore, it is proved experimentally that the proposed DL model (BI-LSTM) performed better than the other ML techniques.

4.2.2. Comparison with Deep Learning Techniques. To perform comparison of the proposed BILSTM model with different deep learning classifiers, we implemented each DL classifiers and the proposed BILSTM classifier on the acquired dataset. The performance evaluation results are presented in Table 6. We have used several DL classifiers such as CNN, LSTM, GRU, and RNN. Through experimental results, it is found that the highest accuracy (0.82%) is obtained by the GRU model, where the lowest value of accuracy (0.69%) is achieved by the CNN model.

It is obvious that the proposed BILSTM model outperformed different DL classifiers in terms of better precision (0.85%), recall (0.85%), F-score (0.85%), and accuracy (0.85%).

(1) Justification behind Our Better Results. The proposed Bi-LSTM model used forward and backward LSTM that are able to capture the contextual information in two directions, i.e., forward (future) and backward (previous). The Bi-LSTM model generates an enhanced representation regarding an input text through capturing the information from left towards right and from right towards left that saves the information from loss. However, the single CNN model only executed the task of feature extraction without retaining the information of sequences within a text that results in a poor performance in terms of precision (0.74%), recall (0.69%), F-score (0.69%), and accuracy (0.69%) of CNN model. For the unidirectional LSTM, the reason behind low performance in terms of precision (0.79%), recall (0.79%), F-score (0.78%), and accuracy (0.79%) is that it only memorizes the prior information excluding the future information. In the case of the GRU model, the performance decay in terms of precision (0.82%), recall (0.82%), F-score (0.82%), and accuracy (0.82%) is because of saving the context information in a single direction (prior information) lacking the future information. Finally, the RNN model is not able to memorize the long-term dependencies which is significant to keep the sequence information for a long time duration. So, the limitation of the RNN model of keeping the information for a short time interval results in a performance decay (precision = 0.75%, recall = 0.72%, $F-$Score = 0.72%, and accuracy = 0.72%). Therefore, it is observed experimentally that the proposed DL model (BI-LSTM) performed better than the other DL techniques.

Table 7: Performance evaluation with baseline studies.

Study	Technique (s)	Results
Wald et al. [8]	(i) LR (ii) MLP (iii) RF (iv) SVM	73% (accuracy) SVM
Sumner et al. [33]	(i) SVM (ii) RF (iv) J48 (v) NB	0.639% (accuracy)
Preotiuc-Pietro et al. [34]	(i) Unigram (ii) Liwc (iii) Word cluster	Pearson correlation R .25
Proposed (our work)	Bi-LSTM	0.85 (precision) 0.85 (recall) 0.85 (F-score) 0.85 (accuracy)

4.3. Answer to RQ3: How to Evaluate the Effectiveness of the Proposed System w.r.t the Baseline Techniques? While answering RQ3 we evaluated the performance of the proposed system with respect to the baseline study. The performance evaluation results of comparing studies and the proposed model are presented in Table 7. Detail of comparing studies is presented as follows.

4.3.1. Wald et al. [8]. Wald et al. [8] suggested ensemble training, in particular SelectRUSBoost, using four classification algorithms. Experiments show that the SelectRUSBoost plus SVM kernel achieved the best result (with a 20 MI attribute). However, stronger results can be achieved by studying the specifics of word-based input using in-depth learning techniques.

4.3.2. Sumner et al. [33]. Sumner et al. [33] applied different machine learning classifiers such as SVM, FR, NB, and J48 for predicting dark trait personality trait from the users. An accuracy of 0.67% was achieved on the Twitter dataset. However, there is a need to investigate word embedding-based feature by applying deep learning technique. This may significant performance improvement as compared to the classical machine learning technique.

4.3.3. Preotiue-Pietro et al. [34]. Preotiuc-Pietro et al. [34] worked on the detection of dark trait of personality from twitter by taking into account unigram features supported by LWIC dictionary and word clustering techniques. Experimental are statically significant in term of Pearson correlation ($R = 25$). However, there is a need to exploited psychopathic trait explicitly from its superset, called dark trait.

4.3.4. Proposed Model (Our Work). The proposed BILSTM model when applied on the labeled dataset yielded the best performance results in terms of precision (85%), recall (85%), $F1$-score (86%), and accuracy (85%).

(1) Justification behind Our Better Results. The proposed Bi-LSTM model used forward and backward LSTM that are able to capture the contextual information in two directions that is forward (future) and backward (previous). The B–LSTM

model generates an enhanced representation regarding an input text through capturing the information from left towards right and from right towards left, resulting in saving the information from loss.

5. Conclusion and Future Work

The current research work is aimed at performing the text classification into psychopath and nonpsychopath by exploiting a deep neural network model called Bi-LSTM. The proposed study consists of different modules: (i) data collection, (ii) preprocessing, and (iii) applying deep learning model (BiLSTM).

First, the numerical representation (continuous values) of input text is performed at embedding layer, then proposed Bi-LSTM stores the sequence information in the two directions that are from left towards right and right towards left because it holds two layers known as forward layer and backward layer. These two layers of the Bi-LSTM model assist in preserving the context information in forward and backward directions generating a rich depiction of input text. Finally, the classification of input text is executed at the output layer using the softmax activation function into two classes known as psychopath and nonpsychopath. During experiments, the implementation of different machine learning models and deep learning models is performed on the given dataset. The conducted experiments reveal that the proposed Bi-LSTM model outperformed all the other comparing models and obtaining the improved results in terms of precision (85%), recall (85%), $f1$-score (85%), and accuracy (85%).

5.1. Limitations. The following are the limitations of the current research work.

(1) The dataset collected for experimentation is insufficient that may decay the performance of the proposed model

(2) The present study is limited to the implementation of the BI-LSTM model without applying the fusion of different deep neural networks such as CNN + LSTM, CNN + Bi-LSTM, CNN + RNN, and CNN + Bi-RNN

(3) We exploited only the random word embedding for the input layer, while other different word representation techniques like Glove and FastText may improve the system performance

(4) The focus of the present study is on English textual content

(5) The current study lacks the exploitation of different kinds of features like audio, video, and images that may assist in system performance

5.2. Future Directions

(1) The future aim is to enhance the size of the dataset that can assist in attaining improved results regarding the proposed Bi-LSTM model

(2) Exploitation of other different deep neural networks like CNN + LSTM, CNN + Bi-LSTM, CNN + RNN, and CNN + Bi-RNN

(3) In the future, we aim to apply different word representation schemes like Glove and FastText

(4) The possible extension of the present work will be to exploit different languages for personality classification other than English text

(5) In the future, we inspect various features like audio, video, and images other than textual features

(6) We aimed to extend the preprocessing modules by exploiting other text cleaning steps like grammar correction, spell checker, and stemming

Data Availability

Underlying data supporting the results can be provided by sending a request to the 3rd author or corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors are grateful to the Deanship of Scientific Research, King Saud University for funding through Vice Deanship of Scientific Research Chairs.

References

[1] P. Shrestha, "Trait theory of personality," 2017, https://www.psychestudy.com/general/personality/trait-theory.

[2] J. T. Hancock, M. T. Woodworth, and S. Porter, "Hungry like the wolf: a word-pattern analysis of the language of psychopaths," *Legal and Criminological Psychology*, vol. 18, no. 1, pp. 102–114, 2013.

[3] Illimitable Man, "Understanding the dark triad–a general overview," 2014, https://illimitablemen.com/2013/11/17/understanding-the-dark-triad/.

[4] All things Psychopath, "Inside the mind of a psychopath essay," https://www.bartleby.com/essay/Inside-the-Mind-of-a-Psychopath-FKKMZEZVC.

[5] N. Ahmad and J. Siddique, "Personality assessment using Twitter tweets," *Procedia Computer Science*, vol. 112, pp. 1964–1973, 2017.

[6] M. E. Hastings, J. P. Tangney, and J. Stuewig, "Psychopathy and identification of facial expressions of emotion," *Personality and Individual Differences*, vol. 44, no. 7, pp. 1474–1483, 2008.

[7] D. E. Reidy, A. Zeichner, K. Hunnicutt-Ferguson, and S. O. Lilienfeld, "Psychopathy traits and the processing of emotion words: results of a lexical decision task," *Cognition and Emotion*, vol. 22, no. 6, pp. 1174–1186, 2008.

[8] R. Wald, T. M. Khoshgoftaar, A. Napolitano, and C. Sumner, "Using Twitter content to predict psychopathy," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2, pp. 394–401, Boca Raton, FL, USA, 2012.

[9] J. Robison, S. McQuiggan, and J. Lester, "Developing empirically based student personality profiles for affective feedback models," in *Intelligent Tutoring Systems. ITS 2010. Lecture Notes in Computer Science*, V. Aleven, J. Kay, and J. Mostow, Eds., vol. 6094, pp. 285–295, Springer, Berlin, Heidelberg, 2010.

[10] M. T. Whitty, J. Doodson, S. Creese, and D. Hodges, "A picture tells a thousand words: what Facebook and Twitter images convey about our personality," *Personality and Individual Differences*, vol. 133, pp. 109–114, 2018.

[11] M. Z. Asghar, F. Subhan, M. Imran et al., "Performance evaluation of supervised machine learning techniques for efficient detection of emotions from online content," *Computers, Materials & Continua*, vol. 63, no. 3, pp. 1093–1118, 2020.

[12] T. Tandera, D. Suhartono, R. Wongso, and Y. L. Prasetio, "Personality prediction system from Facebook users," *Procedia Computer Science*, vol. 116, pp. 604–611, 2017.

[13] L. Liu, D. Preotiuc-Pietro, Z. R. Samani, M. E. Moghaddam, and L. H. Ungar, "Analyzing personality through social media profile picture choice," in *Proceedings of the International AAAI Conference on Web and Social Media*, pp. 211–220, Cologne, Germany, 2016, May.

[14] H. A. A. Akram and A. Mahmood, "Predicting personality traits, gender and psychopath behavior of Twitter users," *International Journal of Technology Diffusion*, vol. 5, no. 2, pp. 1–14, 2014.

[15] N. Bukhtawer, S. Muhammad, and A. Iqbal, "Personality traits and self regulation: a comparative study among current, relapse and remitted drug abuse patients," *Health*, vol. 6, no. 12, pp. 1368–1375, 2014.

[16] F. Keshtkar, C. Burkett, H. Li, and A. C. Graesser, "Using data mining techniques to detect the personality of players in an educational game," in *Educational Data Mining. Studies in Computational Intelligence*, A. Peña-Ayala, Ed., vol. 524, pp. 125–150, Springer, Cham, 2014.

[17] G. Farnadi, S. Sushmita, G. Sitaraman, N. Ton, M. De Cock, and S. Davalos, "A multivariate regression approach to personality impression recognition of vloggers," *Proceedings of the 2014 ACM Multi Media on Workshop on Computational Personality Recognition - WCPR '14*, 2014, pp. 1–6, Orlando, FL, USA, 2014.

[18] J. Pednekar and S. Dubey, "Identifying personality trait using social media: a data mining approach," *International Journal of Current Trends in Engineering & Research*, vol. 2, pp. 489–496, 2016.

[19] W. Wijaya, H. Tolle, and F. Utaminingrum, "Personality analysis through handwriting detection using android based mobile device," *Journal of Information Technology and Computer Science*, vol. 2, no. 2, 2017.

[20] A. Shaheen, U. Ali, and H. Kumar, "Extraversion personality traits and social support as determinants of coping responses among individuals with HIV/AIDs," *Journal of Psychology and Clinical Psychiatry*, vol. 4, article 00188, 2015.

[21] M. T. Le, M. Woodworth, L. Gillman, E. Hutton, and R. D. Hare, "The linguistic output of psychopathic offenders during a PCL-R interview," *Criminal Justice and Behavior*, vol. 44, no. 4, pp. 551–565, 2017.

[22] T. S. N. Ayutthaya and K. Pasupa, "Thai sentiment analysis via bidirectional LSTM-CNN model with embedding vectors and sentic features," *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, 2018, pp. 1–6, Pattaya, Thailand, 2018, November.

[23] J. Roesslein, "Tweepy an Easy-to-Use Python Library for Accessing the Twitter API," *Github repository*, 2009.

[24] R. Rodrigues do Carmo, A. M. Lacerda, and D. H. Dalip, "A majority voting approach for sentiment analysis in short texts using topic models," in *Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web*, pp. 449–455, New York, USA, 2017.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort et al., "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[26] H. Ahmad, M. Z. Asghar, A. S. Khan, and A. Habib, "A systematic literature review of personality trait classification from textual content," *Open Computer Science*, vol. 10, no. 1, pp. 175–193, 2020.

[27] G. Lefebvre, S. Berlemont, F. Mamalet, and C. Garcia, "Inertial Gesture Recognition with Blstm-Rnn," in *Artificial Neural Networks. Springer Series in Bio-/Neuroinformatics*, P. Koprinkova-Hristova, V. Mladenov, and N. Kasabov, Eds., vol. 4, pp. 393–410, Springer, Cham, 2015.

[28] A. Khattak, W. T. Paracha, M. Z. Asghar et al., "Fine-grained sentiment analysis for measuring customer satisfaction using an extended set of fuzzy linguistic hedges," *International Journal of Computational Intelligence Systems*, vol. 13, no. 1, pp. 744–756, 2020.

[29] L. Nio and K. Murakami, "Japanese sentiment classification using bidirectional long short-term memory recurrent neural network," in *Annual Meeting of the Association for Natural Language Processing*, pp. 1119–1122, Okayama, Japan, 2018.

[30] P. Zhou, Z. Qi, S. Zheng, J. Xu, H. Bao, and B. Xu, "Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling," 2016, https://arxiv.org/abs/1611.06639.

[31] X. Ma and E. Hovy, "End-to-end sequence labeling via bidirectional lstm-cnns-crf," 2016, https://arxiv.org/abs/1603.01354.

[32] Z. Yu, V. Ramanarayanan, D. Suendermann-Oeft et al., "Using bidirectional LSTM recurrent neural networks to learn high-level abstractions of sequential features for automated scoring of non-native spontaneous speech," in *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 338–345, Scottsdale, Arizona, USA, 2015.

[33] C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets," in *2012 11th International Conference on Machine Learning and Applications*, vol. 2, pp. 386–393, Boca Raton, FL, USA, 2012.

[34] D. Preotiuc-Pietro, J. Carpenter, S. Giorgi, and L. Ungar, "Studying the dark triad of personality through Twitter behavior," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 761–770, Indianapolis Indiana, USA, 2016.