Social Network Analysis and Mining (2022) 12:67
https://doi.org/10.1007/s13278-022-00884-x

ORIGINAL ARTICLE

A hybrid approach for the detection and monitoring of people having personality disorders on social networks

Mourad Ellouze[1] · **Lamia Hadrich Belguith**[1]

Received: 31 December 2021 / Revised: 11 March 2022 / Accepted: 3 May 2022 / Published online: 27 June 2022

Abstract
Research in the medical field does not stop evolving. This evolution obliges doctors to be up-to-date in order to well manage every situation that may occur with their patients. However, the medical field is very sensitive and requires a great deal of precision, all of that poses a major problem. Consequently, there is a recourse to computer science, to resolve all of these issues. In this context, we propose in this paper an architecture, taking advantage of artificial intelligence (AI) and text mining techniques to: (i) identify individuals with personality disorder from their textual production on social networks by classifying their set of tweets into distinct classes representing respectively the presence, the category and the type of the disease and (ii) guarantee personalized monitoring by filtering inappropriate tweets according to patient's circumstance. The first phase was achieved by taking advantage of a deep neuronal approach that benefits of: (i) CNN layers for features extraction from the textual part, (ii) two LSTM layers to preserve long-term dependencies between different lexical units, (iii) SVM classifier to detect the sick person using the dependency links found from the previous layer. The second phase was accomplished by applying a hybrid approach that combined linguistic and statistical techniques in order to filter inappropriate tweets according to the state of each patient. Following the evaluation of our approach, we acquire an F-measure rate equivalent to 84% for the detection of personality disorder, 64% for the detection of the type of disease and 70% for the task of filtering inappropriate content. The obtained results are motivating and may encourage researchers to improve them in view of the interest and the importance of this research axis.

Keywords Personality disorder · Social media · Deep learning · Natural language processing · Text mining · Semantic analysis

1 Introduction

Psychiatric diseases or mental illnesses are disorders that affect the behavior, thinking and emotions of individuals and result in difficulties integrating into society. This problem arises from the sick person's instability, which puts him in a critical state, including socially irresponsible, mentally unbalanced and uncontrollable reactions. Therefore, we expect him to react badly toward any circumstance (curse, contempt, threat, ...). All these reactions are due to the symptoms of paranoid illness such as aggressiveness, grudge, feelings of superiority, expectation of attacks from others, etc. These effects have contributed to the emergence of several dangerous consequences existing frequently nowadays such as suicide, terrorism, etc. Despite the danger of these diseases, we notice an increase in the number of people with psychological problems, particularly in less developed countries Kõlves et al. (2006), where various issues such as economic, social, and political issues are ignored. In this context, the World Health Organization (WHO) has stated that one among four people in the world suffers from mental disorders and in half of the countries of the world, there is only one psychiatrist for 100,000 inhabitants. Moreover, 40% of the countries have less than one hospital bed for mental disorders compared to 10,000 inhabitants Organization (2001). All these consequences, limit the under control situation of people with personality disorders, even with existing systems like Baumgartl et al. (2020) that worked on Electroencephalographic data and Wang et al. (2021) that

✉ Mourad Ellouze
ellouzemourad@yahoo.fr

Lamia Hadrich Belguith
lamia.belguith@fsegs.usf.tn

[1] ANLP Research Group, MIRACL Laboratory, FSEGS, University of Sfax, Sfax, Tunisia

based their treatment on speech data, this remains limited. This constraint arises is due that to ensure the good functioning of their systems, it is necessary to be in a specific environment with sophisticated equipment (sensors, MRI, etc.) which makes the task of the detection extremely difficult.

Nowadays, social media represents one of the most favorable environments, allowing its users to communicate and express themselves freely about everything happening in the world with total freedom. For this reason, social networks can represent a suitable climate for people with personality disorders to show their bad behavior, including aggressiveness, violence, etc. For that, we notice in the last years, an enormous increase in the degree of violence and harassment on social networks. Obviously, a person in a normal situation is not going to be violent and attempt to harass people. In this context "Statistica" states that 81% of women aged between 18 and 24 faced at least one form of harassment on a dating site in France in 2018[1]. This percentage presents a sense of danger that has moral and physical effects on people also it makes the browsing climate uncomfortable in this context, 64% of women have blocked someone to avoid being bothered by his messages[2]. The automation of detecting people with personality disorders is the major challenge for social networking services (SNS), but there are several factors related to the characteristics of data that make this assignment troublesome such as the volume, nonstructuring, general lexicon, an idea can be written in several ways even implicitly (do not contain bad words) as it shows this example: *Imagine you are beautiful*. For this reason, in the last years, there is a remarkable recourse to advanced techniques like AI in order to make the processing of the enormous amount of this data more feasible. For that, our challenge in this work is to ensure the monitoring of people with personality disorders on social media by making a deep analysis of their shared textual content. This is due that, detecting implicit information is difficult since it lacks lexical signals that reflect the true meaning. In this paper, we focus on two targets for achieving the stated objectives:

1. Detect people having personality disorders disease by classifying their tweets into a hierarchical tree representing the presence of the disease, the category and the type of the disease based on their textual production on social media. This task was done using a deep learning model composed of multilayers; unlike traditional machine learning techniques, deep learning techniques are able to automatically extract the relevant properties on which we will base our work from raw data (Ruiz et al. 2020).
2. Provide specialized monitoring of the state of each sick person by hiding shared tweets that may negatively affect their state. This was done by calculating the semantic similarity between their tweets that showed the presence of personality disorder disease (PD) and tweets shared by their following.
3. Detect the writing style of people with PD through the estimated effect of linguistic features on the classification class to validate our hypothesis that PD affects the writing style of its patients.

Our proposal may be used by experts and scientists who need to analyze social media data and keep track of the health of its sick users (those with personality disorders) in order to avoid a critical situation (suicide) and to not impact negatively others.

This paper is organized as follows. We start with an overview of various studies in this field. Then, we detail our methods for monitoring individuals suffering from personality disorders. Finally, we conclude this paper with a conclusion and some perspectives.

2 Related work

There are numerous obstacles related to the treatment of data coming from social media since there are several criteria that can intervene and influence the data supplied by users. These criteria can incorporate the age of the person, country, level of education, etc. Moreover, many users have not considered social media as an official framework for that they used in their writing style irony, sarcasm, etc., which may disrupt the treatment afterward. Thus, we note the breaking of linguistic rules (punctuation, capital letters, using terms that do not belong to the lexicon of a specific language, using more than one language to write a sentence, etc.).

However, social networks operate as a community mirror. There are a lot of researchers who choose the text of social networks as the principal source of data for their research works to ensure the monitoring of people's conditions by ensuring an attentive listening to their needs regardless of their type (social, health, economic, etc.). In this context, Comito (2021) presents a method for conducting a detailed study of Twitter data to verify how information about COVID-19 outbreaks has propagated across the USA while taking into account the evolution of the debate over time. Despite the diversity of themes in social media, it was not a barrier for researchers (Ellouze et al. 2021; Comito et al. 2016, 2019) to specify the context of speech in order to recognize key events and issues in the world. Similarly, many more studies have considered social networks as a useful

[1] https://fr.statista.com/statistiques/942385/femmes-harcelement-sites-de-rencontre-en-ligne-par-age-france/.

[2] https://fr.statista.com/statistiques/944602/part-femmes-ayant-deja-bloque-utilisateur-harceleur-sur-site-de-rencontre/.

resource for studying and monitoring individuals all around the world. For that, we have partitioned the distinctive works analyzed into two categories:

2.1 Explicit data processing

Explicit data processing is generally focused on identifying clear and obvious information. This can be accomplished simply by analyzing a single social media post, as this type of information is usually tied to a specific lexical topic, for example, *terrorism : {fanatic, extremist, hostages, enemies, crimes, weapons, war, attack, etc.}*

violence : {aggression, ostracism, bullying, mistreatment, racketeering, bullying, retaliation, etc.}

In this context, (Rekik et al. 2019) used a statistical approach to detect violent tweets based on the calculation of the degree of belonging of a tweet to each class presented with a set of n-gram words. Thus, Ahmad and Siddique (2017) worked to detect weird tweets by classifying them into four classes: compliance, dominance, submission and influence. The corpus used in this study is a collection of tweets that were gathered using certain keywords. Then, the classification step was done by the RapidMiner tool (Hofmann and Klinkenberg 2016). The result of this work is a visual representation in numerous shapes of output (graph, etc.) presenting the distinction of simple and compound words between classes. Detection of suicide on social media is an evolutionary axis of research. For that, Mbarek et al. (2019) recommended employing a classification algorithm to detect profiles of people with suicide intent on Twitter. The different features used in the classification step are linked to several information like: (1) linguistic features such as part of speech (POS), frequent word and n-gram, (2) emotional features such as emojis and depression terms, (3) facial features such as age, hair and mustache, which are extracted from the user's profile photograph, (4) chronology features such as the number of publications per day, per month, etc., (5) public information such as country. Emotion detection might have a negative impact on personality. That's why we found many works that were entitled in this field. We can cite (Wang et al. 2019) as an example among the several works found in this context. For the classification of emotions, this research presented a hierarchical tree structure of neural networks. The first part consists in modeling the document in an LSTM tree. This tree groups keywords with their weights and also the relationships between them as well as information about the subject's distribution. After that, a "softmax" output layer is used to classify social emotions.

2.2 Implicit data processing

Implicit information is frequently portrayed as concealed information such as age, personality traits and psychological problems where a particular treatment is required since in several cases, indeed with a human being, it cannot be recognized. This is owing to the fact that detecting them is a sensitive task that requires a larger volume of data compared to the detection of explicit information. For that, several researchers have worked on data obtained from multiple sources, with different types in order to ensure the data variation aspect (Varshney et al. 2017; Pramodh and Vijayalata 2016; An et al. 2018). Other researchers focused on the variety of characteristics chosen (Bleidorn and Hopwood 2019; González-Gallardo et al. 2015; Cellia and Lepri 2018); for example, some of them combined linguistic criteria (morphological analysis, etc.), meta-data such as the number of friends and different information related to the tweet like the number of words or number of hashtags.

In general, there is a large amount of uncertainty in the result of the hidden information detection system. For this reason, there are a lot of researchers who have turned to statistical approaches (Pramodh and Vijayalata 2016; Ellouze et al. 2020) rather than classical machine learning techniques (Stankevich et al. 2018; Mbarek et al. 2019) and deep learning techniques (Wang et al. 2019; An et al. 2018; Wang et al. 2019) in order to guarantee the notion of fuzzy logic.

The processing of hidden information enables researchers to conduct further research in order to better understand hidden factors (in some cases unexpected) that lead to these results. In this context, many researchers have attempted to extract knowledge in the form of rules between the writing style and personality traits (Hall and Caton 2017; Schwartz et al. 2013). The results obtained by Schwartz et al. (2013) are that extroverted people used more expressions attached to the lexicon of friends, family, etc. Moreover, they have more positive feelings compared to the others. In the same context, Baik et al. (2016) proposed an approach for categorical data having the same topic (such as musician category and business category). The rules obtained from this approach are that extroverted people appear more intrigued in sports, shopping, hotels, whereas introverted people are drawn to video games. This extricated information can offer assistance in making expectations and can also enrich the knowledge base of therapists and psychologists. On the other hand, Holtzman et al. (2019) was interested in identifying linguistic markers used by narcissistic people. This approach began by identifying important LIWC[3] characteristics using the measure of "*Pearson weighted*"[4] to calculate the correlation between LIWC and narcissistic disorder. Then, using the metaphor package existing in language R, an estimate was computed for each extracted effect by computing the confidence intervals (CI) (Hoogman et al. 2017). This study

[3] Linguistic Inquiry and Word Count.

[4] https://en.wikipedia.org/wiki/Pearson_correlation_coefficient.

showed that there were positive correlations between LIWC and narcissism disorder, citing, for example, that the narcissistic person uses more words related to sport, as well as the pronoun of the second person. Furthermore, for negative correlations, there is frequent use of words related to anxiety and fear as well as words having multiple meanings.

2.3 Limits analysis

Following a review of the many studies listed, we discovered that the majority of authors have focused on the detection of the consequences of psychological diseases such as violence, terrorism or suicide (Rekik et al. 2019; Ahmad and Siddique 2017; Mbarek et al. 2019), rather than the detection of personality disorders types. Even researchers who worked on the illness centered on the discovery portion than the monitoring of health states. Moreover, despite the existence of other equally important languages, we observe an overwhelming use of the English language. Besides, there are several issues with the lexical approach's application (Salem et al. 2019). In general, this strategy is dependent on keyword research from the corpus (lexicons connected to each class), so one of the issues with this technique is the difficulty in finding a training corpus that includes all lexicons relevant to a certain class. Additionally, there are numerous issues related to the lexical approach such as the ability to recognize explicit but not implicit information. We also notice an overuse of machine learning algorithms in the classification process (Stankevich et al. 2018; An et al. 2018; Lin et al. 2017). However, one of the problems with the classification aspect is that an instance cannot belong to more than one class at once.

3 Proposed approach

In this study, we present an approach illustrated in Fig. 1 that allows Twitter to analyze the textual production of its users in real time in order to ensure the surveillance of people suffering from PD on social networks by filtering posts that may irritate them. To achieve this, we suggest a way for updating the list of Twitter observers for each post, starting with checking the psychological states of the default observers of this tweet by looking at for each of them their last 20 posts on Twitter.

This task was done using a novel deep learning model (see Fig. 2) containing a set of convolution layers CNN for the extraction of high-level features from raw textual data.

Besides, using two LSTM layers to highlight the long-distance dependencies between the different lexical units of the textual part, the output of the last layer is then passed to SVM, which is used to detect the presence of personality disorder disease, its category and finally its type (see algorithm 1). The choice of SVM is defended by the fact that SVM is among the foremost configurable classical machine learning algorithm that gives a better chance of achieving a good result. Besides, after reviewing the various related works, we discovered that in several cases, combining deep learning and SVM produces a good result (Chen and Zhang 2018; Ombabi et al. 2020). Following the detection of people with PD, our approach heads to screening tweets that may cause them distress while being cautious in the screening process. For this reason, we employed a method that took into account the aspect of similarity between the new tweet shared and each tweet found in the story of the person detected as sick in the previous step. This similarity is at the level of the topic and the semantic degree while considering a person's disposition toward this topic. Our approach addresses other problems at the same time such as: (1) unbalanced data via data duplication step and (2) the lexical approach, via sentence embedding technique for representing whole sentences and their semantic information as a set of numerical vectors. This technique assists the machine in understanding the context, intent and other nuances of the whole text.

Algorithm 1: Classification of people according to their personality disorder disease

```
Result: Detect the different types of personality disorders among users of social networks
for each newtweet in Twitter do
    observerList = The user's followers list that posted the new tweet;
    for each personX in ObserverList do
        last20tweets = Getting the last of each observer's 20 tweets;
        if personX == Person with PD then
            for each category of PD do
                if personX == category then
                    for each type of the category do
                        if personX == type then
                            tpoicX = Detect the topic of the new tweet;
                            for each tweetY of last20tweets do
                                tpoicY = Detect the topic of the tweetY;
                                moodY = Detect the mood expressed in the tweetY;
                                if (tpoicX == tpoicY and moodY == "bad mood") then
                                    if (sentence embedding (newtweet,tweetY)>thresshold) then
                                        observerList = observerList.remove(personX);
                                    end
                                end
                            end
                        end
                    end
                end
            end
        end
    end
end
```

3.1 Preprocessing

In this step, we focused on the preprocessing of our corpus by removing unnecessary elements (do not make a distinction between classes). This task was accomplished through the following steps: At the first time, we eliminated the different stopwords including articulator words such *as, and, also, therefore, etc*. These words are used by everyone, so it has not helped to distinguish between the various classes. Next, we eliminated from our corpus the different symbols used to express money, time, number, etc. Then, we normalized our corpus by transforming capital letters into lowercase letters and the abbreviation of words into a normal form using the resource Google Graph Knowledge[5] like *AI to artificial intelligence*. Finally, we transformed the inflectional forms of words to a common root to act in the same way with words having the same common root as *transform, transformation, transforming, etc*. This step was performed using the NLTK[6] library which enables automatic language processing (Bird 2006).

3.2 Features generation

This step consists of transforming the textual data (a set of 20 tweets of each person) into numerical vectors that can be processed by machine learning algorithms. According to our analysis of several works, there are several ways to make this transformation such as Word Embedding Bakarov (2018). However, the main problem with such a technique is that it does not retain the meaning of the entire sentence. This will not assist algorithms in deciphering the intent and nuance of the content. For that, we choose to work with sentence embedding techniques such as Sentence Bert Feng et al. (2020), InferSent Reimers and Gurevych (2019), Universal Sentence Encoder (USE) Cer et al. (2018), etc. Following an empirical study, we decided to work with LASER

[5] The Knowledge Graph is a knowledge base used by Google to compile the results of its search engine with semantic information from various sources.

[6] Natural Language Toolkit.

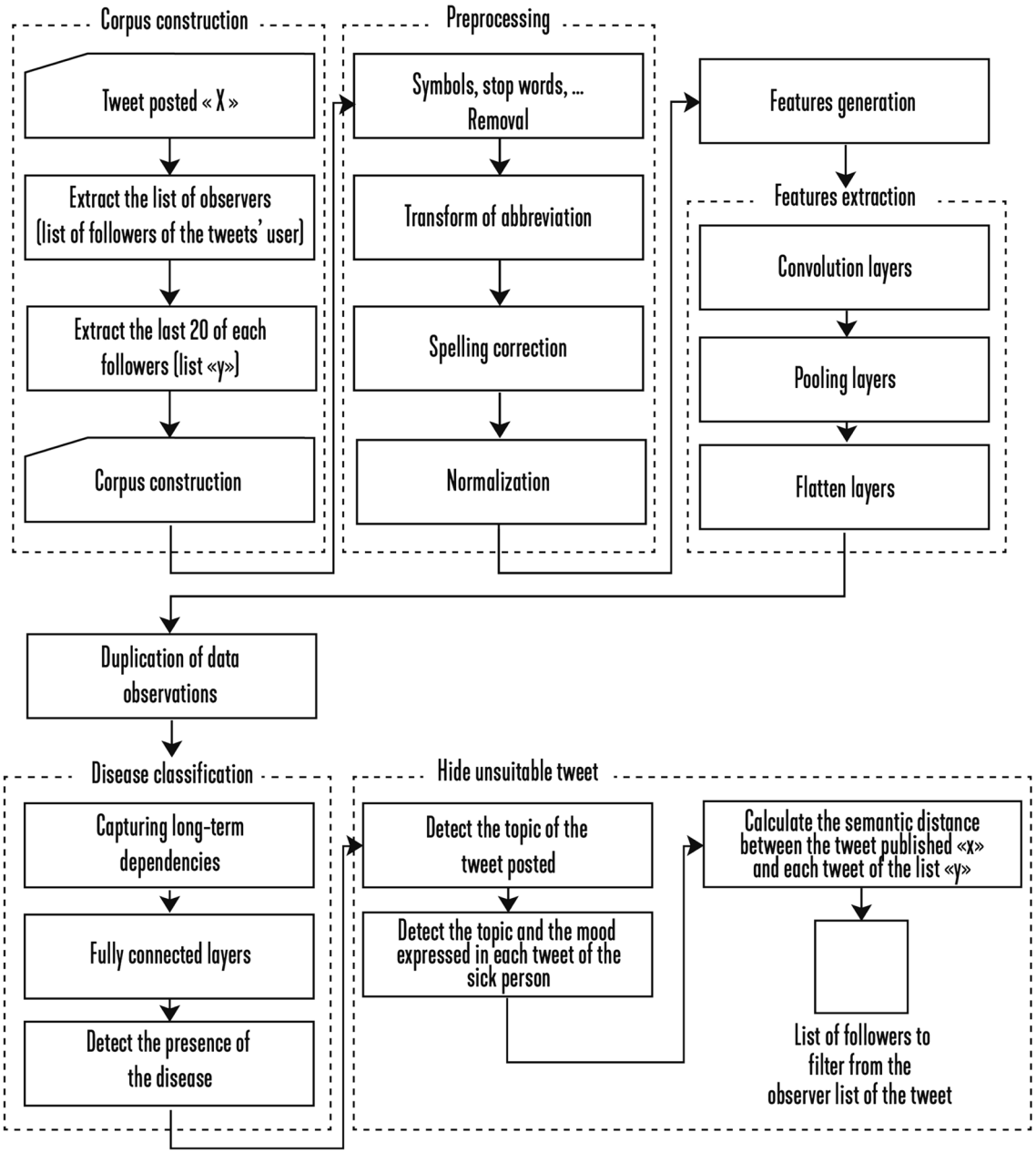


Fig. 1 The proposed approach for the detection and monitoring of people having personality disorders on social networks

(Language-Agnostic SEntence Representations) technique since this model handles multilingual text (93 languages). Moreover, it is trained on 223M parallel sentences from a variety of sources. Each sentence is represented as a 1,024-dimensional vector by the encoder, which is implemented as a 5-layer BiLSTM network Krasnowska-Kieraś and Wróblewska (2019). This method is based on calculating sentence similarity using pooling layers in order to maintain only the most essential descriptors. In addition, this technique provides as a result a set of standardized vectors while resolving an important number of well-known difficulties linked to the size of the data set and the diversity of vocabularies in the corpus.

3.3 Convolutional neural network for features extraction

The convolutional neural network (CNN) is a particular type of neural network whose architecture differs from the classic architecture of the MLP (multilayer perceptron) model. This difference mainly revolves around the convolutional part. The objective of this part is to reduce the raw size of

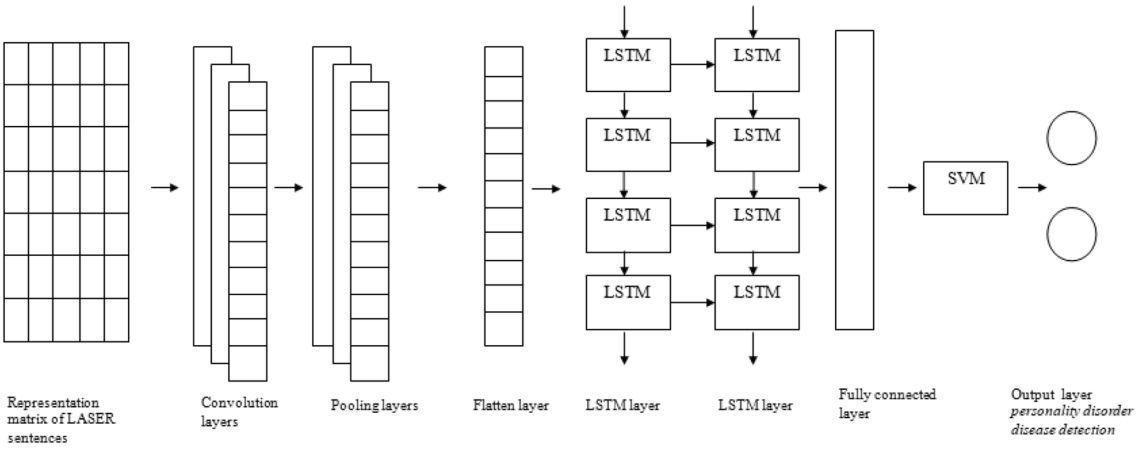


Fig. 2 Proposed deep CNN-LSTM model for existence, category and type of disease detected

the input form in order to highlight the relevant characteristics. Several studies have shown the importance of this technique compared to other techniques based on traditional handcrafted descriptors at the level of generalization and treatment of immanent noise issues (Kumar and Sundaram 2022; AlAjlan and Saudagar 2021). In addition, CNN architecture displayed a remarkable performance in different tasks of NLP (natural language processing) for capturing the syntactic and semantic elements (Ombabi et al. 2020) mentioning semantic parsing (Yih et al. 2014), sentence modeling (Kalchbrenner et al. 2014), search query retrieval (Shen et al. 2014) and other NLP challenges (Collobert et al. 2011); for this reason, we have used this technique in our work. In CNN model, an entry subset to its previous layers is connected through a convolutional layer, and hence, CNN layers are called feature map. Although these models worked well, they require stacking a large number of convolutional layers to capture long-term dependencies. For this reason, the CNN model uses a polling layer to reduce the output size of stack layers and this can help to reduce the computational complexity and to preserve only important information. The flatten layer is used to supply the output of polling layer and matches it with the following layers.

3.4 Duplication of data observations

In this step, we aim to maximize the number of instances in order to resolve the unbalanced data problem since in certain cases annotating a big amount of data or finding more data is difficult. For making this task, there are a lot of means such as Multiobjective Genetic Sampling for Imbalanced Classification (E-MOSAIC) (Fernandes et al. 2019), Exploratory Data Analysis (EDA) for handling duplicate records and Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al. 2002). After an empirical study, we decide to use the SMOTE technique, since it has demonstrated significant effectiveness in various applications and fields (Quan et al. 2021; Ishaq et al. 2021) and this is a perfect fit for our case, as our corpus is not related to a specific field. This technique uses the nearest neighbors algorithm to generate new and synthetic data until the minority and majority classes have the same share of the population.

3.5 Disease classification

LSTM (Graves 2012) is defined as RNN (Sherstinsky 2020) architecture with a supplementary cell of internal memory that was built to solve the problem of explosion and disappearance gradient faced by RNN since we may fall into a situation of delays of unknown length between different events in a time series. Subsequently, the basic advantage of utilizing LSTM is the capability to "keep in mind" past values for any length of time and for managing the information flow in the network, and LSTM applies recursive execution of the current cell block using the old hidden state and the current entry. In addition, LSTM uses other techniques such as : (i) Dropout Techniques: This strategy is employed to prevent the model from overfitting. It removes from the network, extraneous information that is not useful for further processing, and this can help also to improve the model's performance, and (ii) Dense Layer : It connects each entry with each exit by weight. In our work, we choose to use LSTM in order to maintain the links of dependence between the various lexical units. For this reason, we have concatenated the output of the convolutional layer to an LSTM layer. Then, the output of the first LSTM layer is transmitted to the second LSTM layer, which generates a deep representation of the original sentence. The final outputs of the LSTM layers are fused and transferred to a fully connected layer. After that, we passed the fully connected layer to an SVM classifier in order to make the classification of the disease. The last layer is composed of two neurons, whose goal is the

Table 1 Linguistic features extraction

Type	Description
Numeric features	Number of each punctuation
	Number of each sentence
	Number of words in a sentence
	Number of named entities
Morphological features	Number of each POS
	Tense of each sentence
	Number of entity gender (masculine/feminine)
	Number of entity forms (singular/plural)
Semantic features	Sentimental analysis
	Semantic relations

detection of the disease. Our decision to use SVM is justified by the fact that SVM is one of the most configurable learning algorithms that offers more opportunities to get good results. Furthermore, multiple studies have discovered that the best combination between deep learning and classical learning algorithms for text analysis is obtained by SVM (Chen and Zhang 2018; Ombabi et al. 2020; Thaiyalnayaki 2021). In addition, during the processing phase SVM takes into consideration the error and the complexity at once (Dilrukshi et al. 2013). The classification step has been repeated at least three times if we detect from the beginning (first classification) that the user has a PD. It is worth noting that we used binary classification for each time to ensure the multi-label aspect (one instance can have more than one class), as a person can have multiple diseases.

3.6 Hide unsuitable tweet

In this step, we applied the approach of Ellouze et al. (2021) in order to detect reasons that affect a PD among Twitter users. The choice of using this approach is justified by the fact that this work is applicable to all domains and it has already demonstrated its effectiveness while using the French language. Moreover, this approach is based on a hybrid approach in the form of a combination between linguistic method and numerical learning technique to ensure both semantic and statistical aspects. In addition, this work dealt with various problems related to text analysis tasks, including the growth of the vocabulary through time, as well as the lack of explanation and difficulties in interpreting machine learning algorithm results. In this state, we begin our work by determining whether a tweet has been shared by a user who has a personality disorder among their followers. In this case, we are going to compare the topic of the tweet shared by this user with the topic of each tweet from the list of the last 20 tweets of the sick person using the approach of Ellouze et al. (2021). Therefore, if we found a pair of tweets with the same topic and the sick person's mood in the tweet is negative, we move to measure the semantic distance between them using the technique of sentence embedding. If the value exceeds the threshold, we remove this user from the list of following of this tweet in order to not provoke him.

Note 1: The interest of the crossing by topic is to guarantee the conservation of the context example "I hate this food" and "I do not want the sport"; these two sentences have a certain degree of semantic resemblance, but they have not the same meaning.
Note 2: The choice of using the similarity distance calculation is justified by the fact that domains are very large. We take these two tweets as an example "I'm proud of our champion swimmer" and "I'm very angry to have my soccer team lose the match." These tweets are on the same topic, but they do not mean the same thing.
Note 3: We have applied a lot of filters in order to avoid falling into the problem of overfiltering, which can annoy users.

4 Data analysis

In this step, we aim to determine the language distinctiveness of people with personality disorders by looking for commonalities in their writing style characteristics. Thus, we attempt to validate our hypothesis that personality disorder sickness affects these patients' writing style, which makes it possible to discover persons with personality disorders from their textual production on social networks. This is accomplished by calculating the estimated effect of each feature among linguistics features criteria (see Table 1) extracted using NLTK library on the classification class, which enables us to detect the causal relations between them. Various statistical measures can be used to complete this task, such as: χ^2 *test*[7], *mutual information*[8], *the coefficient of likelihood*[9] *and measure of Bayes*[10]. Following an empirical investigation, we have decided to employ the χ^2 *test* for antisocial case. χ^2 *test* is a statistical measure that is used to test for independence among qualitative variables, while taking into account the number of occurrences and absences of the different elements together, likewise one among the others. After calculating χ^2 *test*, we calculate the P value to determine which features are dependent and independent in comparison with the classification classes using a decision threshold (Dahiru 2008).

[7] https://en.wikipedia.org/wiki/Chi-squared_test.
[8] https://fr.wikipedia.org/wiki/Information_mutuelle.
[9] https://en.wikipedia.org/wiki/Likelihood_function.
[10] https://en.wikipedia.org/wiki/Bayesian_probability.

Table 2 The distribution of instances per class

Classes	Number of instances per user	Number of instances per tweets
Person with PD	884 users	17680 tweets
Normal Person	531 users	10620 tweets
Person with Suspicious Category disease	422 users	8 440 tweets
Person with Emotional Category disease	580 users	11600 tweets
Person with Anxious Category disease	577 users	11540 tweets
Person with Paranoid disease	263 users	5260 tweets
Person with Schizoid disease	97 users	1940 tweets
Person with Schizotypal disease	174 users	3480 tweets
Person with Antisocial disease	154 users	9000 tweets
Person with Borderline disease	231 users	4620 tweets
Person with Histrionic disease	248 users	4960 tweets
Person with Narcissistic disease	83 users	1660 tweets
Person with Avoiding disease	79 users	1580 tweets
Person with Dependent disease	258 users	5160 tweets
Person with Obsessive compulsive disease	241 users	4820 tweets

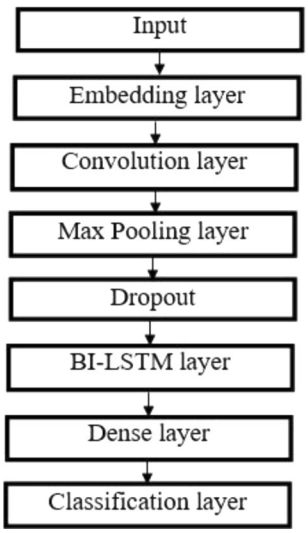


Fig. 3 Architecture of the proposed model

5 Experiments and results

This part presents information about our dataset, as well as information about our layers' configuration and an excerpt of the results obtained. This work has been implemented using the python programming language which integrates the Tensorflow framework.

5.1 Corpus

We applied our approach to data composed of a set of tweets including a vocabulary related to the side effects of the disease "personality disorders" such as *"I am wary," "I congratulate myself" and "I am in the confusion of."* This data was obtained in real time using Apache Spark Streaming tool for tweets in French language. This corpus was labeled by two psychiatrists who are asked to doubly annotate a set of tweets of our corpus according to their knowledge and experiences. The annotation process was started with an empirical study of 10% part of the corpus in order to better understand the specificities related to the language of social networks and to elaborate a manual annotation. Afterward, each annotator annotated the 90% part of the corpus separately. Annotations of different forms of classification are

Table 3 Model parameter structure

Layer type	Output shape	Param#
Input Layer	(700,1)	
conv1d (Conv1D)	(700, 320)	1280
Max_pooling1d	(233, 320)	0
Dropout (Dropout)	(233, 320)	0
conv1d_1 (Conv1D)	(233, 320)	307520
Max_pooling1d_1	(77, 320)	0
Dropout_1 (Dropout)	(77, 320)	0
conv1d_2 (Conv1D)	(77, 320)	307520
Max_pooling1d_2	(25, 320)	0
Dropout_2 (Dropout)	(25, 320)	0
Time_distributed	(1, 8000)	0
lstm (LSTM)	(250)	8251000
lstm_1 (LSTM)	(100)	140400
Dense (Dense)	(None, 80)	8080
Classification layer	2	82

Table 4 Extract of results (translated to English) of disease detection

An excerpt from a user's history of tweets	Personality Disorder	Cluster A	Cluster B	Paranoid	Borderline
1. Du grand délire ! Une hystérie incommensurable. *(Great delirium! An immeasurable hysteria....)*	YES	YES	YES	YES	YES
2. Le Professeur Raoult confirme la fin de l'épidémie sur Radio Classique et règle ses comptes avec les sorciers prévisionnistes de la catastrophe. *(Professor Raoult confirms the end of the epidemic on Radio Classique and settles accounts with the wizards forecasting the catastrophe)*					
3. Ils sont payés au PV, dommage, les FDO perdent toute crédibilité ! *(They are paid by the ticket, too bad, the FDO lose all credibility!)*					
4. Second vague couplée á d'autres nouvelles pandémies donc confinement á perpétuité jusqu'á se que les populations ne se rappellent plus ce que signifie les mots liberté et contestation ... *(second wave coupled with other new pandemics therefore confinement in perpetuity until the populations no longer remember what the words freedom and protest mean...)*					
5. Très vindicative Ruth Elkrief envers Michel Onfray. Insupportables ces journalistes serviteurs du gouvernement. *(Very vindictive Ruth Elkrief toward Michel Onfray. Unbearable these journalists servants of the government.)*					
6. Nous avons beaucoup de variétés de penis de chiens. *(We have many varieties of dog penises.)*					
7. Attention : âmes sensibles, ne regardez pas ! Resto chinois, suite et fin (2/2) Question bonus : Qu'avons-nous de commun avec cette culture ? *(Warning: sensitive souls, do not watch! Chinese restaurant, continuation and end (2/2) Bonus question: What do we have in common with this culture?)*					
8. Aujourd'hui, pour les poulets, c'est un peu comme l'ouverture des soldes, faut remplir un maximum avant minuit, alors ils sont en mode racket intensif. Ils doivent toucher une prime pour être d'aussi "bons serviteurs de l'état." *(Today, for the chickens, it's a bit like the opening of the sales, they have to fill up as much as possible before midnight, so they are in intensive racket mode. They have to get a bonus to be such "good servants of the state.")*					

made independently, which means for each user profile (20 tweets) each annotator gives: (i) their decision about the state of the person *"person with PD"* or *"normal person."* We consider a person with a personality disorder if in his last 20 tweets there is a redundancy of linguistic indicators that reflect signs of this sickness such as semantic information expressing terrible disturbance and fear as, for example, the following expressions *"my hair is standing on the end," "I can hardly breathe," "my throat gets knotted," etc.*, (ii) if this person is classified as a person with PD, annotators move to classify their set of tweets into a binary classification for each category of PD (suspicious, emotional, anxious), for example, if the person has emotional and anxiety problems they accord "YES" for each label of the class in order to ensure the multi-label aspect (a person can have more than one category of PD), (iii) for each category of PD that a user possesses, annotators move to check each type of PD related to this category. After the annotation phase of the disease detection, they move to make the annotation of a set of combination of two tweets. This means that if they have the same meaning, the tweet may influence and exacerbate the sick person's case. In this case, we should remove the user from the list of following of this tweet. After the annotation phase initiated by the 2 experts to mark up the presence of diseases, we proceed to the calculation of the rate of agreement between these two experts using Cohen's Kappa measure[11]. In this context, we got an average value of 76% for disease classification and 87% for similarity calculation. Conflicting cases are primarily related to the misinterpretation of cases (error in measuring the degree of the intensity of symptoms as well as between missing information or diligence). For that, we invited our experts to reconvene and choose between reaching a consensus or eliminating cases that were in contention. Table 2 shows in more detail the distribution of annotated tweets per class. Besides our corpus, we used another open source corpus[12] that was elaborated in the work (Astuti 2021). This corpus is composed of 1251 Tweets written in Indonesian with 20 features and three different types; however, we have only focused our study on the text and class portion. These tweets were accessed through the Twitter API V2 service from April 10,

[11] https://fr.wikipedia.org/wiki/Kappa_de_Cohen.

[12] https://www.kaggle.com/fitriandri/antisocial-behaviour-public-twitter-indonesia.

Table 5 Extract of results (translated to English) of filtering inappropriate content

Tweet shared by a sick person	Tweet shared by a person among the list of following of the sick person	Filtering
#urgent ! 146 nouveaux cas de #coronavirus et un nouveau décès lié á la maladie sont enregistrés en #Haïti. (*#urgent ! 146 new cases of #coronavirus and one new death related to the disease are recorded in #Haiti.*)	Déjá plus de 520 000 décès par coronavirus dans le monde. (*Already more than 520,000 coronavirus deaths worldwide.*)	YES
Chine, 03 Juillet : Étude sur 15000 patients . "Nous avons constaté que le taux de COVID-19 symptomatique était plus fort. (*China, July 03: Study of 15,000 patients." We found that the level of symptomatic COVID-19 was higher.*)	NEW YORK NEW STUDY - Étude Rétrospective de 6493 patients ambulatoires et hospitalisés avec COVID-19. (*NEW YORK NEW STUDY - Retrospective study of 6493 outpatients and hospitalized patients with COVID-19.*)	YES
Covid-19 : le rapport choc des pompiers sur la gestion de la pandémie. (*Covid-19: the firefighters' shock report on the management of the pandemic.*)	Une étude chinoise met en garde, contre la possibilité d'un nouveau "#virus #pandémique" provenant des #porcs. (*A Chinese study warns of the possibility of a new "#pandemic #virus" from #pigs.*)	YES

2021, to April 16, 2021. The tweets mentioned are divided into five classes *Failure to conform social norms of lawful behavior, Reckless Disregard for Safety, Irritability and aggressive, Reckless Disregard for Safety, Lack of Remorse, Non-Antisocial or General Class*. In our case, we aggregated antisocial people's tweets, regardless of their type, to obtain a slightly balanced corpus (465 antisocial people's tweets and 786 regular people's tweets).

5.2 Proposed approach results

For the different parameters applied to each layer in our model, we used three convolution layers with 320 feature maps and ReLU as an activation function (AF) for each layer, followed by three pooling layers with a pool size equal to the number of feature maps in each layer (1,9). Next, we used two LSTM layers composed of 250 neurons for the first layer and 100 neurons for the second layer associated with one hidden layer with "softmax" as an activation function, associated with an output layer composed of 2 neurons (representing the presence or the absence of this disease). The model of CNN input and output with multiple parameters is presented in Table 3 and Fig. 3. We repeated the execution of this task, while the result of the detection is true, which means as long as the presence of PD is confirmed, we move to detect the category of the disease; in case the detection was true, the next step is to detect the disease's type.

Next, we advance to detect the reason of this disease by detecting the topic and the mood expressed in the tweet. After that, we move to measure the semantic similarity between the combination of tweets to keep track of the sick person's state and prevent him from entering into a critical situation. We employed the Python programming language to manage these different layers with their parameters. Tables 4 and 5 show an excerpt of our results:

5.3 Data analysis results

As mentioned previously, we took advantage of the Chi-square and the P value measures to calculate the dependence of all features derived from the data analysis section in relation to our corpus and the corpus used for the monitoring services of Indonesian public on Twitter (Astuti 2021). The choice of using two different corpora with different characteristics is supported by the objective of achieving generalized results on the temporal aspect (the tweets of the two corpora are not extracted during the same period) and the spatial aspect. (The tweets of the two corpora are not written in the same language.) As shown in Table 6, we discovered a plethora of related results.

Table 6 Writing style analysis (general features) for the two corpus

	Our Corpus		Corpus Astuti (2021)	
	χ^2 test	P value	χ^2 test	P value
Punctuation : semicolon	4.17	0.38	3.91	0.27
POS : conjunction, pronoun, determining and preposition	104.66	0.04	81.66	2.32e-05
Punctuation : exclamation and question marks	21.74	0.17	10.70	0.05
POS : singular and named entity	66.03	0.05	102.37	0.5e-09
POS : adjective and adverb grammatical categories	99.51	0.01	57.1	0.00045
Sentiment analysis : negative feeling	4.71	0.58	0.54	0.46
semantic relation : consequence and explanation	10.02	0.05	11.07	0.006
Semantic relation : linking and addition	22.07	0.01	12.81	0.05
Semantic relation : opposition	4.86	0.08	0.07	0.78

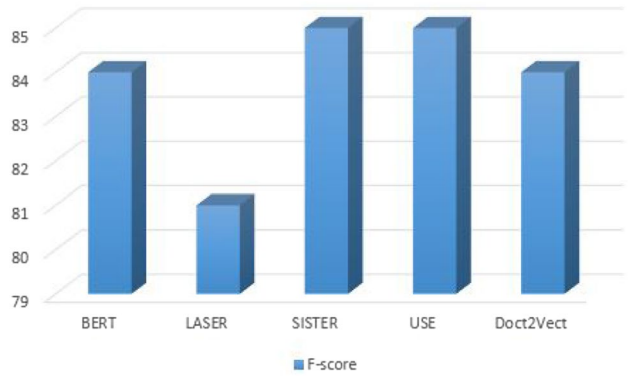
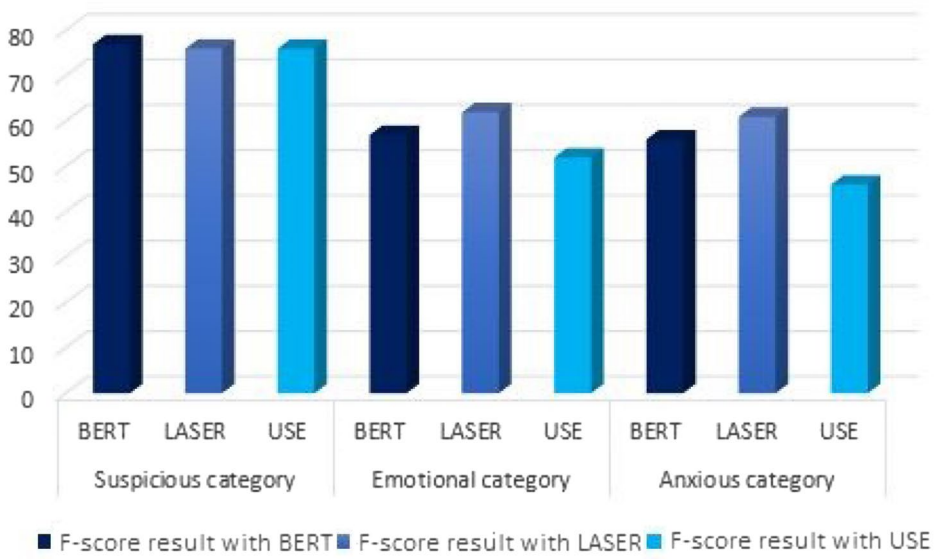


Fig. 4 F-score comparison result of different embedding models for personality disorder classification using model composed of CNN-BILSTM-SVM

5.4 Evaluation

We tested the performance of the different tasks of our work by applying the classical criteria recall, precision and F-measure to each classification type (personality disorder, category, type). Tables 8, 9, 10, 11, 12 , 13 and Figs. 4, 5, 6 display with more detail the evaluation of our approach for the task of disease detection with the different techniques of sentence embedding. For the result of the evaluation of the semantic analysis task to ensure the surveillance of people having PD, we have obtained an F-measure rate equivalent to 70% with a precision rate equivalent to 64% and 76% for the recall rate. Finally, for the assessment of writing style analysis, we compared our results with the results of the American Speech–Language–Hearing Association (ASHA)[13] (see Table 7).

Fig. 5 F-score comparison result of different embedding models for different categories of personality disorder classification using model composed of CNN-BILSTM-SVM



[13] https://www.asha.org/Practice-Portal/Clinical-Topics/Written-Language-Disorders/Signs-and-Symptoms-of-Written-Language-Disorders.

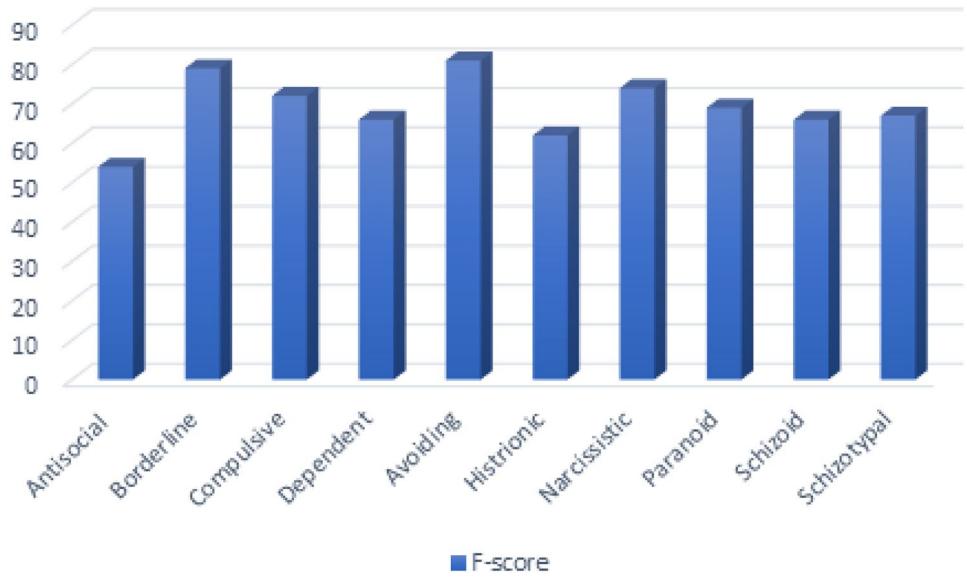


Fig. 6 F-score comparison result of LASER model embedding technique for different types of personality disorder classification using model composed of CNN-BILSTM-SVM

Table 7 Comparison of our analysis with ASHA results for antisocial writing style analysis.

ASHA	Analysis	Our corpus results	Corpus Astuti (2021) results
Tends to employ shorter T-unit of sentence fragments and has a less intricate sentence structure	The utilization rate of semicolon punctuation is notable	✓	✓
The narrative and explanatory discourse are poorly organized; finds it tough to bring his ideas to fruition	The usage of semantic relations to demonstrate explanations, connections, additions, and consequences have decreased significantly	✓	✓
Is not able to write from different viewpoints	Decrease in the rate of use of the opposition relationship	✓	✗
Suffers from inflexible morphological difficulties which contrarily influences the structure of sentences	The POS determinant, conjunction, and preposition are used less frequently	✓	✓
Difficulty in understanding and interpreting texts written by others	Exclamation points and question marks are frequently used, as are sentences having a negative tone	✓	✓
Uses less abstract language	The employment of singular versus plural forms is very impressive. High utilization of adjectives and named entities	✗	✗

Table 8 Variation of F-measure according to the different sentence embedding and classifiers used for PD detection

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	80	81	81	81	**87**
LSTM	78	82	**86**	85	84
LSTM+SVM	83	**85**	81	76	84
BILSTM+SVM	84	81	**85**	**85**	84

Bold values represent the best result related to each classification case

Table 9 Variation of F-measure according to the different sentence embedding and classifiers used for suspicious category (cluster A) of PD detection

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	**74**	65	66	67	51
LSTM	**77**	64	68	59	48
LSTM+SVM	**72**	71	65	62	58
BILSTM+SVM	**77**	76	63	76	51

Bold values represent the best result related to each classification case

Note 1: For the evaluation task of our approach, we applied the cross-validation technique. It should be noted that each time we switch between the training folds and the test fold. This is because the SMOTE technique was only used to the training folds to avoid influencing the evaluation results of our approach.

Table 10 Variation of F-measure according to the different sentence embedding and classifiers used for emotional category (cluster B) of PD detection

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	50	**61**	46	56	48
LSTM	77	64	68	**85**	59
LSTM+SVM	**60**	55	57	53	**60**
BILSTM+SVM	57	**62**	61	52	52

Bold values represent the best result related to each classification case

Table 11 Variation of F-measure according to the different sentence embedding and classifiers used for anxious category (cluster C) of PD detection

	BERT	LASER	SISTER	USE	Doc2Vect
BILSTM	57	46	**61**	59	51
LSTM	**55**	53	**55**	52	51
LSTM+SVM	55	46	**58**	50	**58**
BILSTM+SVM	56	**61**	56	46	60

Bold values represent the best result related to each classification case

Table 12 Variation of F-measure according to the different sentence embedding techniques for LSTM+SVM classifier combination used to detect PD type

	BERT	LASER	SISTER	USE	Doc2Vect
Antisocial	**55**	49	54	52	52
Borderline	74	**77**	64	69	60
Compulsive	62	**66**	56	62	55
Dependent	**76**	67	61	65	58
Avoiding	75	74	71	73	**80**
histrionic	55	52	48	**58**	48
Narcissistic	55	**70**	51	61	57
Paranoid	56	**62**	51	61	56
Schizoid	55	59	**71**	61	57
Schizotypal	56	**58**	57	50	50

Bold values represent the best result related to each classification case

Table 13 Variation of F-measure according to the different sentence embedding techniques for BILSTM+SVM classifier combination used to detect PD type

	BERT	LASER	SISTER	USE	Doc2Vect
Antisocial	**59**	54	57	57	58
Borderline	76	79	74	77	**83**
Compulsive	66	72	67	**85**	61
Dependent	65	66	72	**86**	67
Avoiding	73	81	79	**88**	71
Histrionic	60	**62**	52	54	60
Narcissistic	60	74	**76**	66	53
Paranoid	60	**69**	56	68	61
Schizoid	70	66	68	**75**	65
Schizotypal	53	**67**	54	58	52

Bold values represent the best result related to each classification case

6 Discussion

This paper proposed an intelligent approach combining machine learning and text mining techniques. The objective of this approach is to allow experts and scientists to make a deep analysis of the content provided by users on Twitter platform. With that, we can show different details about the status of Twitter users in relation to PD disease (category, type). This approach deals with the multi-label aspect at the level of whether the person has multiple types of PD at once, especially that symptoms of the different types of disease are very close. In addition, this work may provide Twitter the opportunity to ensure personalized monitoring of each sick user by filtering inappropriate Tweets according to their case. This work meets the limitations presented in previous works at the level that first our work follows the full process of the diagnosis of PD disease (detecting the PD and ensuring the surveillance of patients). Second, it takes advantage of the deep learning approach to extract relevant features using CNN layers. Third, it maintains linguistic links between the different lexical units using LSTM layers in order to obtain reliable results. In addition, our approach is based on the combination of: (i) different techniques of machine learning to resolve problems related to the difficulty of rules construction task, (ii) linguistic aspect at the level of using an ontology, etc., in order to make our results more interpretable. Moreover, we addressed problems related to the size and unbalanced data through the data generation technique and thus problems related to the lexical approach by using the sentence embedding technique "LASER" which treats the meaning of words in the sentence. Besides, we treat problems related to the detection of implicit information by treating an entire publication history related to each person.

We got the most satisfactory results (F-measure equal to 84%) for personality disorder disease compared to the category and type detection. This is due to the fact that there is less overlap between texts of people with PD and texts written by normal people. However, there is an important degree of similarity between expressions showing symptoms indicating the category and type of the disease.

In general, we found that combining the BILSTM and SVM algorithms yielded the best results for the majority of disease classification categories and types. That shows the validity of our hypothesis that LSTM is very efficient for text analysis task, especially in the detection of dependency links

between the different lexical units. In addition, SVM is very important and may improve the results since it is among the most configurable algorithm in comparison with the other classical algorithms.

For the different cases of category and type classification, we obtained various F-measure results (81% for avoiding classification and 54% for antisocial classification). This variety is due to the specificity of each type of disease class, such as the variety of the lexicon, the way of reacting of the algorithm to each situation. In our work, we can justify this variety by the existence and the combination of several criteria that may intervene and influence results such as: (i) the reduced size of the corpus used to classify this type of disease. For example, in the classification of antisocial people, we do not find enough instances. This is due to the fact that an antisocial person does not react frequently with social media users, (ii) the existence of some symptoms shared between these different diseases such as instability in antisocial and borderline diseases. Moreover, the reduced size of the number of instances of antisocial compared to borderline class may disrupt our algorithm especially that they belong to the same category); (iii) there are some symptoms that are very specific and vague, which may make the task of classification very difficult. For example, characters of histrionic disease are: fuzzy, vague and subjective. For this reason, we obtained 62% as F-measure result for the case of histrionic classification (less efficient compared to other results), and (iv) linguistic phenomena such as negations, irony and general lexicon (an idea can be written with several ways). For the result of the second evaluation (semantic analysis task) for filtering inadequate tweets, we notice that despite the use of several types of filters, we have obtained a recall value better than the precision value, and this is due to the generality of the different domains (vague lexicon). It should be noted that this result includes the evaluation of the topic and mood detection and thus the evaluation of sentence embedding. Therefore, the results obtained for the semantic analysis task grouped the error rate figured in the different used systems.

The results obtained during the data analysis phase were interpreted by our expert, stating that, in general, a person with PD admits to having a pessimistic style of thinking (rejection, people will assault him, etc.). For that, we notice from our result that an antisocial person does not regularly use the opposition as a semantic relation; moreover, their feeling is for the most part negative. That puts him in a position of not understanding others which justifies the use of exclamation and question marks. And the impulsivity and aggressiveness of antisocial people have blatantly affected their writing style, appearing in non-compliance with morphological criteria, and the lack of use of the semantic relations as the explanation is very noteworthy.

In our future work, we aim to integrate an analytical module to offer experts and scientists the opportunity to realize multiple dashboards that display the most geographic areas with more infected people by PD in relation with the most anxious topics.

7 Conclusion

In this paper, we proposed a method to detect types of personality disorders among social media users and to monitor their states in order to reduce the dangerous consequences of people with personality disorders such as suicide and violence. This method has several advantages compared to other works since it provides the full process of disease detection (detecting the personality disorder, the category and the type associated with this category). Thus, it provides the cause of this disease and the way to ensure the monitoring of the state of the sick person. Besides, it takes advantage of a deep learning approach that combines at the same time the extraction of features and highlights the links between the various lexical units and the classification tasks. Moreover, this method treats the problems of unbalanced data and the reduced size of the corpus via the task of data generation. The proposed method was implemented, and the obtained results for the evaluation task are encouraging. Indeed, the F-measure for the detection of personality disorder is equal to 84% and the accuracy rate for the filtering of inappropriate tweets task is equal to 70%. As perspectives, we plan to analyze our data and to test our method on a specific application domain.

References

Ahmad N, Siddique J (2017) Personality assessment using Twitter tweets. Proc Comput Sci 112:1964–1973

AlAjlan SA, Saudagar AKJ (2021) Machine learning approach for threat detection on social media posts containing Arabic text. Evolut Intell 14(2):811–822

An G, Levitan SI, Hirschberg J, Levitan R (2018) Deep personality recognition for deception detection. In: INTERSPEECH, pp 421–425

Astuti FA (2021) Antisocial behavior monitoring services of Indonesian public Twitter using machine learning. In: Proceedings of the international conference on data science and official statistics, pp 224–232

Baik J, Lee K, Lee S, Kim Y, Choi J (2016) Predicting personality traits related to consumer behavior using SNS analysis. New Rev Hypermedia Multimed 22(3):189–206

Bakarov A (2018) A survey of word embeddings evaluation methods, arXiv preprint arXiv:1801.09536

Baumgartl H, Dikici F, Sauter D, Buettner R (2020) Detecting antisocial personality disorder using a novel machine learning algorithm based on electroencephalographic data. In: PACIS, p 48

Bird S (2006) NLTK: the natural language toolkit. In: Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions, pp 69–72

Bleidorn W, Hopwood CJ (2019) Using machine learning to advance personality assessment and theory. Personal Soc Psychol Rev 23(2):190–203
Celli F, Lepri B (2018) Is big five better than MBTI? A personality computing challenge using Twitter data. In: CLiC-it
Cer D, Yang Y, Kong S-y, Hua N, Limtiaco N, John RS, Constant N, Guajardo-Cespedes M, Yuan S, Tar C, et al. (2018) Universal sentence encoder, arXiv preprint arXiv:1803.11175
Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357
Chen Y, Zhang Z (2018) Research on text sentiment analysis based on CNNs and SVM. In: 2018 13th IEEE conference on industrial electronics and applications (ICIEA), IEEE, pp 2731–2734
Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P (2011) Natural language processing (almost) from scratch. J Mach Learn Res 12:2493–2537
Comito C, Pizzuti C, Procopio N (2016) Online clustering for topic detection in social data streams. In: 2016 IEEE 28th international conference on tools with artificial intelligence (ICTAI), IEEE, pp 362–369
Comito C, Forestiero A, Pizzuti C (2019) Word embedding based clustering to detect topics in social media. In: 2019 IEEE/WIC/ACM international conference on web intelligence (WI), IEEE, pp 192–199
Comito C (2021) How COVID-19 information spread in us the role of Twitter as early indicator of epidemics. IEEE Trans Services Comput 15(3):1193–1205
Dahiru T (2008) P-value, a true test of statistical significance? A cautionary note. Ann Ib Postgrad Med 6(1):21–26
Dilrukshi I, De Zoysa K, Caldera A (2013) Twitter news classification using SVM. In: 2013 8th international conference on computer / science & education, IEEE, pp 287–291
Ellouze 2021, Mechti S, Belguith LH (2021) Approach based on ontology and machine learning for identifying causes affecting personality disorder disease on Twitter. In: International conference on knowledge science, engineering and management, Springer, pp. 659–669
Ellouze M, Mechti S, Belguith LH (2020) Automatic profile recognition of authors on social media based on hybrid approach. Procedia Comput Sci 176:1111–1120
Feng F, Yang Y, Cer D, Arivazhagan N, Wang W (2020) Language-agnostic bert sentence embedding, arXiv preprint arXiv:2007.01852
Fernandes ER, de Carvalho AC, Yao X (2019) Ensemble of classifiers based on multiobjective genetic sampling for imbalanced data. IEEE Trans Knowl Data Eng 32(6):1104–1115
González-Gallardo CE, Montes A, Sierra G, Núnez-Juárez JA, Salinas-López AJ, Ek J (2015) tweets classification using corpus dependent tags, character and POS N-grams. In: CLEF working notes
Graves A (2012) Long short-term memory. In: Supervised sequence labelling with recurrent neural networks. Springer, pp 37–45
Hall M, Caton S (2017) Am I who I say I am? Unobtrusive self-representation and personality recognition on Facebook. PloS One 12(9):e0184417
Hofmann M, Klinkenberg R (2016) RapidMiner: Data mining use cases and business analytics applications. CRC Press, Boca Raton
Holtzman NS, Tackman AM, Carey AL, Brucks MS, Küfner AC, Deters FG, Back MD, Donnellan MB, Pennebaker JW, Sherman RA et al (2019) Linguistic markers of grandiose narcissism: a LIWC analysis of 15 samples. J Lang Soc Psychol 38(5–6):773–786
Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LS, van Hulzen KJ, Medland SE, Shumskaya E, Jahanshad N et al (2017) Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. Lancet Psychiatry 4(4):310–319
Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, Nappi M (2021) Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE Access 9:39707–39716
Kalchbrenner N, Grefenstette E, Blunsom P (2014) A convolutional neural network for modelling sentences, arXiv preprint arXiv:1404.2188
Kõlves K, Värnik A, Schneider B, Fritze J, Allik J (2006) Recent life events and suicide: a case-control study in Tallinn and Frankfurt. Soc Sci Med 62(11):2887–2896
Krasnowska-Kieraś K, Wróblewska A (2019) Empirical linguistic study of sentence embeddings. In: Proceedings of the 57th annual meeting of the association for computational linguistics, pp. 5729–5739
Kumar V, Sundaram S (2022) Offline Text-independent writer Identification based on word level data, arXiv preprint arXiv:2202.10207
Lin H, Jia J, Qiu J, Zhang Y, Shen G, Xie L, Tang J, Feng L, Chua T-S (2017) Detecting stress based on social interactions in social networks. IEEE Trans Knowl Data Eng 29(9):1820–1833
Mbarek A, Jamoussi S, Charfi A, Hamadou AB (2019) Suicidal profiles detection in Twitter. In: WEBIST, pp 289–296
Ombabi AH, Ouarda W, Alimi AM (2020) Deep learning CNN-LSTM framework for Arabic sentiment analysis using textual information shared in social networks. Soc Netw Anal Min 10(1):1–13
Organization WH et al (2001) Atlas of mental health resources in the world 2001. World Health Organization, Technical Report
Pramodh KC, Vijayalata Y (2016) Automatic personality recognition of authors using big five factor model. In: 2016 IEEE international conference on advances in computer applications (ICACA), IEEE, pp 32–37
Quan Y, Zhong X, Feng W, Chan JC-W, Li Q, Xing M (2021) SMOTE-based weighted deep rotation forest for the imbalanced hyperspectral data classification. Remote Sens 13(3):464
Reimers N, Gurevych I (2019) Sentence-bert: Sentence embeddings using siamese bert-networks, arXiv preprint arXiv:1908.10084
Rekik A, Jamoussi S, Hamadou AB (2019) Violent vocabulary extraction methodology: application to the radicalism detection on social media. In: International conference on computational collective intelligence, Springer, pp. 97–109
Ruiz AP, Gila AA, Irusta U, Huguet JE (2020) Why deep learning performs better than classical machine learning? Dyna Ingenieria E Industria 95(1):119–122
Salem MS, Ismail SS, Aref M (2019) Personality traits for egyptian twitter users dataset. In: Proceedings of the 2019 8th international conference on software and information engineering, pp 206–211
Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Ramones SM, Agrawal M, Shah A, Kosinski M, Stillwell D, Seligman ME et al (2013) Personality, gender, and age in the language of social media: the open-vocabulary approach. PloS One 8(9):e73791
Shen Y, He X, Gao J, Deng L, Mesnil G (2014) Learning semantic representations using convolutional neural networks for web search. In: Proceedings of the 23rd international conference on world wide web, pp 373–374
Sherstinsky A (2020) Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Phys D Nonlinear Phenom 404:132306
Stankevich M, Smirnov I, Kiselnikova N, Ushakova A (2019) Depression detection from social media profiles. In: International conference on data analytics and management in data intensive domains. Springer, pp 181–194
Thaiyalnayaki K (2021) classification of diabetes using deep learning and SVM techniques. Int J Curr Res Rev 13(01):146

Varshney V, Varshney A, Ahmad T, Khan AM (2017) Recognising personality traits using social media. In: 2017 IEEE international conference on power, control, signals and instrumentation engineering (ICPCSI), IEEE, pp 2876–2881

Wang L, You Z-H, Chen X, Li Y-M, Dong Y-N, Li L-P, Zheng K (2019) LMTRDA: using logistic model tree to predict MiRNA-disease associations by fusing multi-source information of sequences and similarities. PLoS Comput Biol 15(3):e1006865

Wang C, Wang B, Xu M (2019) Tree-structured neural networks with topic attention for social emotion classification. IEEE Access 7:95505–95515

Wang B, Wu Y, Vaci N, Liakata M, Lyons T, Saunders KE (2021) Modelling paralinguistic properties in conversational speech to detect bipolar disorder and borderline personality disorder. In: ICASSP 2021-2021 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 7243–7247

Yih W-t, He X, Meek C (2014) Semantic parsing for single-relation question answering. In: Proceedings of the 52nd annual meeting of the association for computational linguistics, Vol 2: Short Papers, pp 643–648