Deep Learning CNN-LSTM Approach for Identifying Twitter Users Suffering from Paranoid Personality Disorder

Mourad Ellouze, Seifeddine Mechti and Lamia Hadrich Belguith
ANLP Group MIRACL Laboratory, FSEGS, University of Sfax-Tunisia, Tunisia

Keywords: Paranoid Personality Disorder Detection, Deep Learning Architecture, Symptoms and Disease Detection, Text Mining, Twitter.

Abstract: In this paper, we propose an approach based on artificial intelligence (AI) and text mining techniques for measuring the degrees of appearance of symptoms related to paranoid disease in Twitter users. This operation will then help in the detection of people suffering from paranoid personality disorder in a manner that provides justifiable and explainable results by answering the question: What factors lead us to believe that this person suffers from paranoid personality disorder? These challenges were achieved using a deep neural approach, including: (i) CNN layers for features extraction step from the textual part, (ii) BiLSTM layer to classify the intensity of symptoms by preserving long-term dependencies, (iii) an SVM classifier to detect users with paranoid personality disorder based on the degree of symptoms obtained from the previous layer. According to this approach, we get an F-measure rate equivalent to 71% for the average measurement of the degree of each symptom and 65% for detecting paranoid people. The results achieved motivate and encourage researchers to improve them in view of the relevance and importance of this research area.

1 INTRODUCTION

Paranoid personality disorder (PPD) is a psychological disease marked by widespread and persistent interpersonal distrust, in which others' acts are misunderstood as spiteful and malicious (Bernstein and Useda, 2007). These side effects might cause inappropriate and unwanted behaviors (such as recklessness, social isolation, insecurity and moodiness), putting the patient in a situation of conflict with society. The worst thing is that many complications are associated with the therapy of this disease, as psychiatric ailments are diagnosed differently than other diseases. This distinction is due to the fact that symptoms are not tangible as the difficulty of breathing and the feeling of oppression for people having Coronavirus.

All these effects have contributed to the appearance of several dangerous consequences existing frequently in our era such as suicide, terrorism, etc. Despite the danger of these diseases, we notice that the number of people having psychological problems is increasing, especially in less-developed countries (Kõlves et al., 2006) since there is negligence about different problems such as economic, social, etc. In this context, the World Health Organization (WHO) has declared that one for every four adults in the world suffers from mental problems and in half of the countries of the world, there is one psychiatrist per 100,000 people. Furthermore, 40 percent of countries have fewer than one hospital bed for mental diseases per 10,000 people (Organization, 2001). As a result, new approaches based on artificial intelligence (AI) have been increasingly used in recent years to automate the work of identifying people with psychiatric issues from raw data.

In this context, (Baumgartl et al., 2020) worked on Electroencephalographic data, while other works based their processing on speech data (Wang et al., 2021). Despite their relevance, these works are limited since to ensure the proper functioning of their systems it is necessary to have sophisticated equipment (MRI, sensors, etc.) which makes the task of the detection extremely challenging.

In this era, social media represents one of the most conducive environments that allows their users to interact and express themselves freely about everything that happens in the world. In recent years, a significant number of researchers have based their works on data collected from social media. In fact the progress and impressive development of computer technologies and tools has made the processing of the huge

Ellouze, M., Mechti, S. and Hadrich Belguith, L.
Deep Learning CNN-LSTM Approach for Identifying Twitter Users Suffering from Paranoid Personality Disorder.
DOI: 10.5220/0011322300003266
In *Proceedings of the 17th International Conference on Software Technologies (ICSOFT 2022)*, pages 612-621
ISBN: 978-989-758-588-3; ISSN: 2184-2833

amount of this data more attainable.

For that, our challenge in this work is to detect people having a personality disorder by analyzing their textual production on social media. To achieve our goal, we divided this paper into two objectives:

1. Measure the appearance degree of each symptom of paranoid disease for each user profile from their textual data. That will ensure the reliability of our final results by providing an explanation and justification for our decision.

2. Detect people having paranoid disease by taking advantage of the result of (1).

This work offers to Twitter the possibility to diagnose the state of their users in order to ensure their well-being by detecting hidden information. In matter of fact, users may not be aware of their mental state. In addition, this work can allow for Twitter the possibility of tracking the progress states of their users in real time (for example Twitter can compare the degree of symptoms related to a specific person in different periods of time).

Starting with a state of the art in which we present some studies done in this field and their limitations. Then, we detail our methods with the different tools used. Next, we discuss the results achieved. Finally, we conclude our work with a conclusion and some perspectives.

2 RELATED WORKS

There are many obstacles related to the treatment of data obtained from social media, since there are certain criteria that may intervene and influence user-generated data. These criteria can include the age of the person, country, level of education, etc. In addition, many users did not consider social media as a formal framework for that, several users used in their writing style irony, sarcasm, etc., which can disrupt the treatment afterward. Moreover, we note a violation of the language rules (punctuation, capital letters, using terms that do not belong in a particular language's lexicon, composing a sentence in more than one language, etc.). However, many researchers opt for text extracted from social networks for their research works. In this context, several researchers have used the data of social networks to detect violent and extremist people (Rekik et al., 2019; Ahmad and Siddique, 2017). Despite the difficulty found in the processing of data extracted from social networks, the objective of these studies remains achievable because there is no concealed information, therefore we can detect the distinct classes using a lexical technique based on keyword searches. On the other hand, detecting hidden information such as age, personality traits and psychological problems is different.

For that in this case, we have to process a huge volume of varied data. In this context, (Varshney et al., 2017; Pramodh and Vijayalata, 2016; An et al., 2018) worked on data obtained from several sources and having different types to ensure the variation in the data. Other researchers worked on the diversity of the characteristics selected (Bleidorn and Hopwood, 2019; González-Gallardo et al., 2015; Celli and Lepri, 2018), for example some of them combined linguistic criteria as morphological analysis, etc., meta-data of Twitter such as number of friends and different information related to the tweet like number of words or number of hashtag. Generally, the result of hidden information detection system contains an important degree of uncertainty. For this reason, there are a lot of researchers who used the statistical approach (Pramodh and Vijayalata, 2016; Ellouze et al., 2020) instead of classical machine learning technique (Stankevich et al., 2019; Mbarek et al., 2019) and deep learning technique (Wang et al., 2019b; An et al., 2018; Wang et al., 2019a) in order to guarantee the notion of fuzzy logic. Among the drawbacks of different machine learning and statistical techniques is that their results are not explanatory, for this reason we found a lot of works that use the rule-based technique (Umar and Qamar, 2019; Muhammad et al., 2019). Although the results of these rules are explanatory since they are based on cause and effect links, the construction of these rules is very time consuming.

Other researchers focused on extracting useful knowledge for doctors by detecting linguistic specificities from the textual production of people having psychological problems (Hall and Caton, 2017),(Schwartz et al., 2013). Among the results found it by (Schwartz et al., 2013): (i) the extrovert people used more terms related to the lexicon of friends and family. Besides, they used terms showing positive feelings. Thus, (Baik et al., 2016) proposed an approach to extract the relevant writing style of each personality trait by associating for each trait some categories of the most used subjects. This work help authors to conclude that extrovert people are very interesting in sports, shopping, hotels. Whereas introverted people are more interesting in gaming.

After the analysis of the different mentioned papers, we note that most 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), and only a few researchers who worked on detecting personality disorders types (Haz et al., 2022; Ellouze et al., 2021b; Ellouze et al., 2021a). In fact, we did not find any paper which treats paranoid disease on social networks. Moreover, we note an excessive use of the English language, despite the existence of other languages with the same importance. Besides, there are some problems related to the use of the lexical approach (Salem et al., 2019). Generally, this technique is based on the research of keywords from the corpus (lexicons related to each class), so among problems related to this technique is the difficulty of finding a training corpus which includes all lexicons related to a specific class. We note also that results obtained (Stankevich et al., 2019; An et al., 2018; Lin et al., 2017) are very abstracts and difficult to interpret besides they need more explanation.

3 PROPOSED APPROACH

In this study, we propose an approach illustrated in figure 1 that allows Twitter to analyse in real time the textual production of their users in order to apply the process of diagnosis committed by the psychiatrist (listen to the patient, identify the different symptoms, detect the disease).

This was done using a novel deep learning model containing a set of convolution layers CNN for automatic features extraction task, since we do not know the criteria of distinction between classes. Next, BiLSTM to make the classification of the degree of each symptom of paranoid disease from the textual part since it highlights the long-distance dependencies of the textual part. Finally, SVM in order to detect paranoid disease based on the degree of each symptom since SVM is among the most configurable learning algorithms (see figure 2).

In addition, our approach treats other problems at the same time such as: (1) corpus imbalanced by using synthetic data generation step, (2) the lexical approach, by the use of sentence embedding technique in order to detect the meaning of the word in the sentence.

3.1 Preprocessing

In this step, we focused on preparing our corpus by deleting unnecessary elements that do not distinguish between classes in order to avoid negatively affecting the subsequent processing, especially that our work is based on data extracted from Web 2.0 (Petz et al., 2015). This task was done by following these steps: For the first time, we eliminated the various stop words including articulatory words such as *and, also, therefore, etc*. These words are used by any person, so they do not help to make the distinction between the different classes. Next, we removed from our corpus the different symbols used for expressing money, time, number, etc. Then, we converted capital letters to lowercase letters and abbreviated terms to their ordinary form to normalize our corpus using the resource Google Graph Knowledge like *AI* to *Artificial Intelligence*. Finally, we converted the inflectional forms of words to a common root to behave similarly to words having the same common root as *transform, transformation, transforming, etc.*. This step was done using the library NLTK (Natural Language Toolkit).

3.2 Features Generation

This step involves converting the textual data into numerical vectors that can be handled by machine learning algorithms. Based on our review of several works, there are many ways to achieve this transformation such as Word Embedding (Bakarov, 2018). However, the major problem of this technique is that it does not preserve the meaning of the whole sentence which makes it difficult for the algorithm to measure the intention and the nuance existing in the text. For that, we choose to work with sentence embedding techniques as Universal Sentence Encoder (USE) (Cer et al., 2018), InferSent (Reimers and Gurevych, 2019), Sentence Bert (Feng et al., 2020), etc. After an empirical study, we choose to work with the Sentence Bert technique since this model is trained on a large amount of data also it has a specific architecture that allows it to learn deep bi-directional representations, it accepts a large number of parameters, making it more adaptable (Eke et al., 2021). This technique is based on the calculation of the similarity between sentences by applying pooling layers in order to keep only important descriptors. In addition, this technique provides as a result a set of standardized vectors while settling many recognized issues related to the size of the data set and the assortment of vocabularies in the corpus. In our work, we have only relied on textual data and we do not use other types of information such as the number of retweets per user, or the number of retweets per tweet since we work on data that is obtained in a streaming manner, therefore at the beginning, the value of these attributes is zero. In addition, our approach is based on a deep learning approach so its specific architecture offers assistance to distinguish automatically the relevant attributes from the raw data.

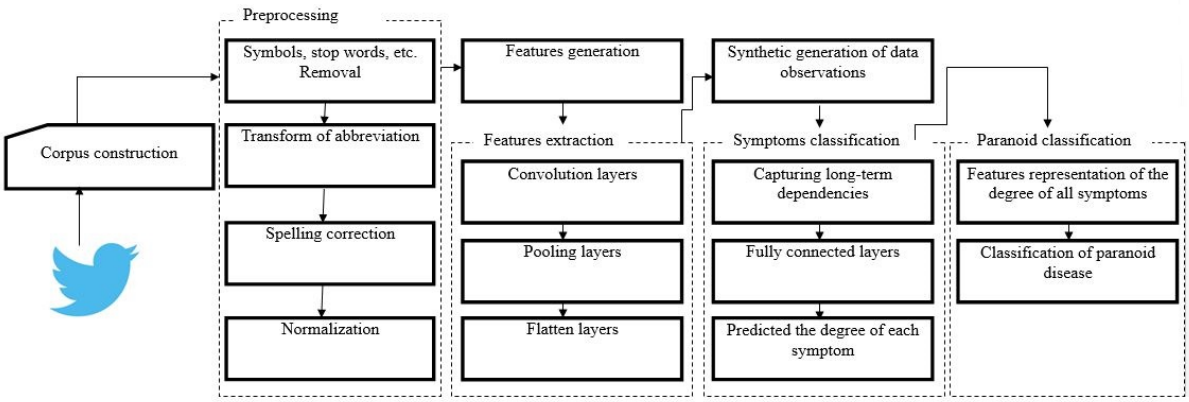

Figure 1: Proposed approach for symptoms and paranoid disease detection.

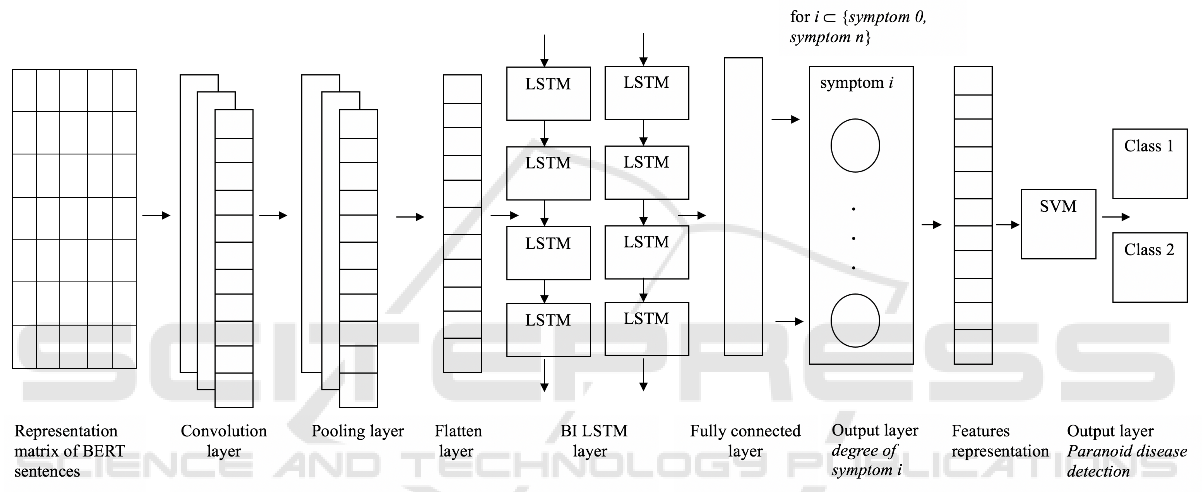

Figure 2: Proposed Deep CNN–BiLSTM for symptoms and paranoid disease detection.

3.3 Features Extraction

The architecture of the Convolutional Neural Network differs from the classic architecture of the MLP (Multi Layers Perceptron) model, this difference mainly revolves around the convolutional part. The purpose of this section is to automatically extract the characteristics and reduce the gross size of the entry form to highlight the relevant characteristics. For this reason, we used in our work the CNN architecture since it performed well on different tasks of natural language processing for capturing the syntactic and semantic aspects (Ombabi et al., 2020). The execution of this task was done by flowing the input (tweets) through a succession of filters, the output of these filters is called convolution maps. The resulting convolution maps are concatenated into a feature vector called CNN code.

3.4 Synthetic Generation of Data Observations

In this step, we aim to maximize the number of instances, considering the difficulty of annotating the data and the difficulty of getting balanced data. For this purpose, there are numerous ways like Multi-objective Genetic Sampling for Imbalanced Classification (E-MOSAIC) (Fernandes et al., 2019), Exploratory Data Analysis (EDA) for handling duplicate records, Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002), etc. After an empirical study, we choose to work with SMOTE technique, since it has shown a great deal of success in various applications and fields (Quan et al., 2021; Ishaq et al., 2021) and our corpus is not linked to a particular field. This technique uses the nearest neighbors algorithm to produce new and synthetic data.

3.5 Symptoms Classification

LSTM is an extension of RNN architecture that was created to address RNN's problem of explosion and vanishing gradient since we may find in a situation of lags of unknown duration between the different events of a time series. Therefore, the LSTM network is adapted to the classification, processing and the realization of predictions based on time series data. In our work, we choose to use BiLSTM in order to keep the dependency links between the lexicon. For that, we concatenated the output of the convolutional layer to two LSTM layers in order to measure the dependence between terms. The last layer is composed of 6 neurons (from 0 to 5) which measures the degree of each symptom of the disease such as exaggerated mistrust, negative interpretation of the gestures of others, incessant doubt, etc. The reason for considering the challenge as a classification rather than a regression is based on the low instances of certain degrees. This classification step has been repeated nine times (the most appeared number of symptoms of paranoid disease), given that in our case it is a multi-label issue, a person may simultaneously have more than one symptom.

3.6 Paranoid Classification

The task of disease detection is very sensitive, this sensitivity is due to the absence of specific rules that allow taking decisions. For example, it is not necessary to have all symptoms in a person to affirm that he has the disease. Moreover, it is difficult to find a person with incessant doubt and self-estimation at the same time. In addition, there are a huge number of combinations of symptoms degrees. For that, in this step, we took advantage of the degree of all symptoms detected in the previous step in order to build a vector (represent the list of symptoms related to an individual). Then, we passed this vector to the SVM layer in order to make the detection of the disease. In this part, the number of features is reduced for that we limited ourselves by the classical classification algorithms such as SVM, Naive Bayes, decision tree, etc. After an empirical study we choose to work with SVM algorithm.

4 EXPERIMENTS

This section presents the different details about our dataset, LSTM settings for "negative interpretation of others' actions" classification and an extract of our results. This work has been implemented using the python programming language which integrates the Tensorflow framework.

4.1 Corpus

We applied our approach to data composed of a set of tweets that included a vocabulary linked to the disease's negative effects "personality disorders" such as *"I congratulate myself", "I am wary", "I am in the confusion of", etc.*. This data was obtained using Apache Spark Streaming tool for tweets in French language from 01-03-2020 to 30-05-2020.

Two psychiatrists were requested to doubly annotate this corpus based on their knowledge and experiences. The annotation process began with an empirical study of a 10% part of the corpus in order to better grasp the nuances related to the language of social networks and to develop a manual of annotation. After that, each annotator separately annotated the 90% of the corpus. The annotations of both types of classification are done independently, which means for each user profile (20 tweets) each annotator gives: (i) the degree of each symptom (a number between 0 and 5 where 0 indicates the absence of the symptom and 5 indicates the high degree of the presence of a symptom), (ii) their decision about the state of the person *"paranoid person"* or *"normal person"*. We consider a person with a paranoid personality disorder if in their last 20 tweets there is a redundancy of linguistic indicators that show the symptoms of this disease such as the semantic information indicating 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.* We set a limit of 20 tweets per user because we aim to develop an approach able of recognizing people with PD by the fewest number of tweets possible in order to ensure early prevention while guaranteeing the credibility of the results.

The selection of paranoid symptoms is based on the most well-known symptoms of paranoid disorder that we have chosen to present at levels 1-5 to be precise. After the annotation phase committed by the 2 experts in order to annotate the degree of presence of symptoms as well as the existence of the disease, we proceed to calculate the rate of agreement between these 2 experts by using Cohen's Kappa measure. In this context, we obtained a value of 0.9 for the detection of the disease and 0.73 for the detection of the degree of symptoms. Conflicting cases are mainly related to the misinterpretation of cases (misinterpretation in measuring the degree of the intensity of symptoms as well as between missing information or diligence). For that, we asked our experts to meet again and choose between (agreement or removing) conflicting cases. Tables 1 and 2 show in more detail the distribution of tweets per class.

Note 1: In table 2 we present the number of paranoid symptoms of each user presented in table 1, which means one person that exists in table 1 can appear up to 9 times in table 2.

Note 2: It should be noted that in some cases we encountered difficulties in the collection of data. For example, for people who believe that they are always right or that they are isolated, they do not need to talk to others and try to persuade them. Particularly for the first case, which have a tendency to be self-centered. While in the case of ”reading hidden meanings in the innocent remarks” and ”recurrent suspicions”, anything can be a trigger and an incentive for these people to write and show what is not expressed (hidden ideas).

Table 1: The distribution of instances for paranoid classification.

Paranoid	YES	NO
Number of instances	280 users (5600 tweets)	450 users (9000 tweets)

Table 2: The distribution of instances for symptoms classification.

Symptoms	Number of instances
aggressiveness	121 users (2420 tweets)
perceives attacks	163 users (3260 tweets)
recurrent suspicions	282 users (5640 tweets)
isolation	46 users (920 tweets)
believing they are always right	76 users (1520 tweets)
read hidden meanings in the innocent remarks	227 users (4540 tweets)
poor relationships with others	273 users (5460 tweets)
doubt the commitment	193 users (3860 tweets)
unforgiving and hold grudges	187 users (3740 tweets)

4.2 Results

For the various settings applied to each layer in our model, we applied three convolution layers, we accord for each of them 320 feature maps and an activation function ”Relu”. Moreover, three pooling layers with a pool size of (1,9). Then, we employed two LSTM layers composed of 250 neurons for the first layer and 150 neurons for the second layer combined with a hidden layer using ”softmax” as an activation function and next with an output layer composed of 6 neurons (representing the degree of this symptom). We repeated the execution of this task 9 times (number of symptoms) and in each case we predicted for each symptom a value which represents the degree of this symptom. The model of CNN input and output with multiple parameters is presented in the table 3.

Table 3: Model parameter structure.

Layer type	Output shape	Param#
Input Layer	(768,1)	
conv1d (Conv1D)	(768, 320)	3200
max_pooling1d	(233, 320)	0
dropout (Dropout)	(233, 320)	0
conv1d_1 (Conv1D)	(233, 320)	921920
max_pooling1d_1	(85, 320)	0
dropout_1 (Dropout)	(85, 320)	0
conv1d_2 (Conv1D)	(85, 320)	921920
max_pooling1d_2	(9, 320)	0
dropout_2 (Dropout)	(9, 320)	0
time_distributed	(1, 8000)	0
lstm (LSTM)	(250)	1211000
lstm_1 (LSTM)	(100)	240600
classification layer	6	906

Next, we passed the vector composed of the degree of the nine symptoms obtained from the previous step to a SVM layer in order to make the classification of paranoid disease. We used SVM layer with a linear kernel and scale gamma since our instances are linearly separable. We employ the Python programming language to handle these various layers with their settings. The following table 4 shows an excerpt of our results for paranoid's symptoms degrees detection.

4.3 Evaluation

We used the classical metrics of recall, precision, and F-measure to assess the performance of each type of classification (symptoms, disease). For the classification of symptoms degrees, we calculated the stated mentioned criteria for the results of the first output layer of our model (degree of symptom i) which means for each symptom of paranoid disease. For the evaluation of the classification of paranoid disease, we applied the mentioned criteria to the results of the last layer of our model.

Note 1: The evaluation of the classification of paranoid disease contains also the error rate figured on the detection of symptoms degrees.

The two tables 5 and 6 display with more details the evaluation of our approach. We compared the results of our work with a Baseline architecture composed of CNN, BiLSTM, SVM which consists in predicting the disease directly without going through the symptom prediction stage. This Baseline is inspired

Table 4: Extract of results (translate to English) of paranoid's symptoms degrees detection.

user's tweets	doubt the commitment	unforgiving and hold grudges	isolation
1. Il y a des gens vous êtes vrmt pourri jusqu'à l'âme. (There are people you are really rotten to the core.) 2. Imagine t'aimes pas provoquer ? quelle vie fade. (Imagine you don't like to provoke? what a bland life.) 3. Depuis que je suis ma seule priorité jsuis beaucoup plus heureuse, ça paye pas de faire passer les autres avant soi. (Since I'm my only priority I'm much happier, it doesn't pay to put others before yourself.) 4. Cette nuit j'aurais préféré ne pas rêver (Tonight I would have preferred not to dream.) 5. Y'a d comportement que jsupporte plus (There is a behavior that I can't stand anymore.) 6. C'est quoi cette nouvelle mode des meufs de se vanter de parler à 1939101 mecs c'est pas une fierté et ça le sera jamais. (What is this new trend for girls to brag about talking to 1939101 guys? It's not a pride and it will never be.) 7. Maintenant être fidèle c'est devenue une qualité alors que ça devrait être normal. (Now being faithful has become a quality when it should be normal.) 8. Y'en a vous êtes culotté c'est incroyable. (There are some of you who are cheeky, it's incredible.) 9. T'as qu'a imaginer que c'était un rêve et que tous ce qu'on a vécu c'était pas vrai. (You just have to imagine that it was a dream and that everything we lived was not true.) 10. L'ingratitude est la pire des choses sah tu donnes tout à des gens qui se foutent de ta gueule.(Ingratitude is the worst thing sah you give everything to people who make fun of you.) 11. Rencontrer quelqu'un avec le même état d'esprit que vous est rare. (Meeting someone with the same mindset as you is rare.) 12. Le dégoût sa aide beaucoup a oublier. (Disgust helps a lot to forget.) 13. Arrête de croire que tout le monde te considère comme tu les considérer ! (Stop believing that everyone considers you the way you consider them!) 14. C'était des grandes paroles en l'air. (It was all talk and no action.)	3	2	2

by (Ombabi et al., 2020) in which the authors addressed the issue of classifying textual data from social networks. The purpose of this comparison is to demonstrate the impact of the layer allowing the detection of symptoms before the disease on the results. The results of this comparison are illustrated in table 7.

Note 2: For the evaluation task of our approach, we applied K-fold cross-validation technique. It should be noted that each time we switch between the training folds and the test fold. This is due that, we applied the SMOTE technique only to the training folds in order to not influence the evaluation results of our approach.

Note 3: We were tolerated in the evaluation of symptoms degrees classification results, this is at the level of accepting the difference of +1 between the real value and the predicted value (the opposite direction is not accepted). However, this remains valid except for cases where the value is between 1 and 5 which means our system has committed an error in the choice of the degree and not in the existence of the symptom.

5 DISCUSSION

This paper presented an intelligent approach based on machine learning and text mining techniques. The objective of this approach is to measure the presence degrees of symptoms in order to detect afterwards paranoid disease among people using social networks. This work meets the limitations presented in the related work section at the level that we respected the logical passage to detect the disease (detect the symptoms then the disease) which makes our results precise, reliable and interpretable. In addition, we used the full process of the deep learning approach (features extraction and classification techniques) since we do not know precisely what are relevant features

Table 5: Variation of Recall, Precision and F-measure according to the model CNN+BiLSTM for symptoms classification.

Symptoms	Recall (%)	Precision (%)	F-measure (%)
aggressiveness	80	73	76
perceives attacks	68	68	68
recurrent suspicions	61	59	60
isolation	86	77	81
believing they are always right	85	78	81
read hidden meanings in the innocent remarks	69	65	67
poor relationships with others	58	56	57
doubt the commitment	72	66	69
unforgiving and hold grudges	80	75	77

Table 6: Variation of Recall, Precision and F-measure according to the selected classifier for paranoid disease detection based on symptoms classification results.

	Recall (%)	Precision (%)	F-measure (%)
Softmax	60	61	60
Gradient Boosting	60	57	58
KNN	62	62	62
AdaBoosting	65	60	62
Random Forest	61	60	60
SVM	66	64	**65**

Table 7: Recall, Precision and F-measure comparison of our results with baseline results for Paranoid classification.

	Recall (%)	Precision (%)	F-measure (%)
Baseline (Ombabi et al., 2020)	59	53	56
Our architecture	66	64	65
Improvement	**7**	**11**	**9**

offering assistance in distinguishing between different classes. Moreover, we addressed issues associated with the size and the unbalanced corpus using the technique of data generation. Thus, problems linked to the lexical approach through the sentence embedding technique "BERT" that deals with the meaning of words in the sentence. We got the most satisfactory results (F-measure equal to 81%) for the following symptoms classification degrees: *believing they are always right* and *isolation*. We obtained the poorest results (F-measure equal to 57%) for the classification of *poor relationships with others symptom*. This difference is due to the language specificity linked to each symptom class and the way of reacting of the algorithm to each situation. In the same context, get the best result for the classification of symptom *believing they are always right* despite the lack of data compared to the symptom *poor relationships with others* for which we have enough data. This is justified by the fact that the second symptom is measured by four degrees with a high error rate, whereas in the first case the symptom is measured using only two degrees. Moreover, we can conclude that the task of data generation has helped us to overcome the problem of data reduction. For the classification of the disease, we obtained the best results using SVM algorithm since the instances of our corpus are linearly separable. Regarding the average results obtained for the detection of some symptoms compared to the results of disease detection, this is due to the: (i) linguistic phenomena such as irony, negations, etc, (ii) high number degree of symptoms (not like binary classification of the disease), (iii) issues with using features in lexicon format (general lexicon, an idea can be written in more than one way, etc). Finally, this work gives Twitter the opportunity to track the state of its users (if there are new symptoms that appear or disappear).

6 CONCLUSION

In this paper, we proposed a method to detect people with paranoid personality disorder. This method has an advantage compared to other works since it provides explanatory results by detecting symptoms of paranoid disease. Besides, it takes advantage of a deep learning approach that combines at the same time the extraction of features and the classification tasks. Furthermore, it addresses the problems of unbalanced data and reduced size of the corpus through the task of generating data. The proposed method was implemented and evaluated and the evaluation results obtained are encouraging, indeed, the F-measure is equal to 79%. As perspectives, we envisage testing our method on other types of personality disorders with a particular application field.

REFERENCES

Ahmad, N. and Siddique, J. (2017). Personality assessment using twitter tweets. *Procedia computer science*, 112:1964–1973.

An, G., Levitan, S. I., Hirschberg, J., and Levitan, R.

(2018). Deep personality recognition for deception detection. In *INTERSPEECH*, pages 421–425.

Baik, J., Lee, K., Lee, S., Kim, Y., and Choi, J. (2016). Predicting personality traits related to consumer behavior using sns analysis. *New Review of Hypermedia and Multimedia*, 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., and Buettner, R. (2020). Detecting antisocial personality disorder using a novel machine learning algorithm based on electroencephalographic data. In *PACIS*, page 48.

Bernstein, D. P. and Useda, J. D. (2007). Paranoid personality disorder.

Bleidorn, W. and Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, 23(2):190–203.

Celli, F. and 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, R. S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., et al. (2018). Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority oversampling technique. *Journal of artificial intelligence research*, 16:321–357.

Eke, C. I., Norman, A. A., and Shuib, L. (2021). Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and bert model. *IEEE Access*, 9:48501–48518.

Ellouze, M., Mechti, S., and Belguith, L. H. (2020). Automatic profile recognition of authors on social media based on hybrid approach. *Procedia Computer Science*, 176:1111–1120.

Ellouze, M., Mechti, S., and Belguith, L. H. (2021a). 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*, pages 659–669. Springer.

Ellouze, M., Mechti, S., Krichen, M., Belguith, L. H., et al. (2021b). A deep learning approach for detecting the behavior of people having personality disorders towards covid-19 from twitter.

Feng, F., Yang, Y., Cer, D., Arivazhagan, N., and Wang, W. (2020). Language-agnostic bert sentence embedding. *arXiv preprint arXiv:2007.01852*.

Fernandes, E. R., de Carvalho, A. C., and Yao, X. (2019). Ensemble of classifiers based on multiobjective genetic sampling for imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1104–1115.

González-Gallardo, C. E., Montes, A., Sierra, G., Núñez-Juárez, J. A., Salinas-López, A. J., and Ek, J. (2015). Tweets classification using corpus dependent tags, character and pos n-grams. In *CLEF (working notes)*.

Hall, M. and Caton, S. (2017). Am i who i say i am? unobtrusive self-representation and personality recognition on facebook. *PloS one*, 12(9):e0184417.

Haz, L., Rodríguez-García, M. Á., and Fernández, A. (2022). Detecting narcissist dark triad psychological traits from twitter.

Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., and Nappi, M. (2021). Improving the prediction of heart failure patients' survival using smote and effective data mining techniques. *IEEE access*, 9:39707–39716.

Kõlves, K., Värnik, A., Schneider, B., Fritze, J., and Allik, J. (2006). Recent life events and suicide: a case-control study in tallinn and frankfurt. *Social science & medicine*, 62(11):2887–2896.

Lin, H., Jia, J., Qiu, J., Zhang, Y., Shen, G., Xie, L., Tang, J., Feng, L., and Chua, T.-S. (2017). Detecting stress based on social interactions in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(9):1820–1833.

Mbarek, A., Jamoussi, S., Charfi, A., and Hamadou, A. B. (2019). Suicidal profiles detection in twitter. In *WEBIST*, pages 289–296.

Muhammad, A., Hendrik, B., and Iswara, R. (2019). Expert system application for diagnosing of bipolar disorder with certainty factor method based on web and android. In *Journal of Physics: Conference Series*, volume 1339, page 012020. IOP Publishing.

Ombabi, A. H., Ouarda, W., and Alimi, A. M. (2020). Deep learning cnn–lstm framework for arabic sentiment analysis using textual information shared in social networks. *Social Network Analysis and Mining*, 10(1):1–13.

Organization, W. H. (2001). The world health report 2001: Mental health: new understanding, new hope.

Petz, G., Karpowicz, M., Fürschuß, H., Auinger, A., Stříteský, V., and Holzinger, A. (2015). Reprint of: Computational approaches for mining user's opinions on the web 2.0. *Information Processing & Management*, 51(4):510–519.

Pramodh, K. C. and Vijayalata, Y. (2016). Automatic personality recognition of authors using big five factor model. In *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, pages 32–37. IEEE.

Quan, Y., Zhong, X., Feng, W., Chan, J. C.-W., Li, Q., and Xing, M. (2021). Smote-based weighted deep rotation forest for the imbalanced hyperspectral data classification. *Remote Sensing*, 13(3):464.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Rekik, A., Jamoussi, S., and Hamadou, A. B. (2019). Violent vocabulary extraction methodology: Application to the radicalism detection on social media. In *International Conference on Computational Collective Intelligence*, pages 97–109. Springer.

Salem, M. S., Ismail, S. S., and Aref, M. (2019). Personality traits for egyptian twitter users dataset. In *Proceedings* *of the 2019 8th International Conference on Software and Information Engineering*, pages 206–211.

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M. E., et al. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one*, 8(9):e73791.

Stankevich, M., Latyshev, A., Kuminskaya, E., Smirnov, I., and Grigoriev, O. (2019). Depression detection from social media texts. In *Elizarov, A., Novikov, B., Stupnikov., S (eds.) Data Analytics and Management in Data Intensive Domains: XXI International Conference DAMDID/RCDL*, page 352.

Umar, A. and Qamar, U. (2019). Detection and diagnosis of psychological disorders through decision rule set formation. In *2019 IEEE 17th International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 33–37. IEEE.

Varshney, V., Varshney, A., Ahmad, T., and Khan, A. M. (2017). Recognising personality traits using social media. In *2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI)*, pages 2876–2881. IEEE.

Wang, B., Wu, Y., Vaci, N., Liakata, M., Lyons, T., and Saunders, K. E. (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)*, pages 7243–7247. IEEE.

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

Wang, L., You, Z.-H., Chen, X., Li, Y.-M., Dong, Y.-N., Li, L.-P., and Zheng, K. (2019b). Lmtrda: Using logistic model tree to predict mirna-disease associations by fusing multi-source information of sequences and similarities. *PLoS computational biology*, 15(3):e1006865.