Detecting Signs of Mental Disorders on Social Networks: a Systematic Literature Review

Ayrton Herculano, Glauco Gomes, Damires Souza, Alex Rêgo
Professional Master's in Information Technology (PPGTI)
Federal Institute of Education, Science and Technology of Paraiba (IFPB)
João Pessoa, Brazil
Email: {ayrton.herculano, glauco.gomes}@academico.ifpb.edu.br, Email: {damires, alex}@ifpb.edu.br

***Abstract*—Social networks have been used more and more by people to share feelings and emotions. This scenario helps to create an environment which may provide additional information regarding signs of mental disorders users carry. In this light, this study investigates the state of the art on how the literature has considered techniques or strategies related to identifying signs of mental disorders on social networks. Some analyses have been made to answer questions regarding methods, data, labelling and other aspects related to such identification. Challenges and gaps are also discussed in this work.**

Keywords–mental health; social media; pre-processing; labeling; systematic review.

I. Introduction

The use of social networks has increasingly gained more and more popularity. By joining communities on social networks, users have provided more information about themselves including, sometimes, feelings and opinions. Feelings or opinions are published on these platforms by means of texts, photos, emoticons or videos [1]. The emotion and language used on social media by means of such posts may, at times, indicate feelings of worthlessness, guilt, loneliness, helplessness, and self-hatred that may characterize, for example, a propensity for depression or for other kinds of mental disorders [2].

A mental disorder is characterized by a clinically significant confusion in an individual's cognition, emotional regulation, or behaviour [3]. Research on mental health in the areas of psychiatry, psychology, sociolinguistics and neuroscience, combined with computational support from specific tools of, for instance, sentiment analysis, can increase the understandings about the relationship between human behavior and the feelings expressed on social profiles. This is due to the fact that users with mental disorders tend to present different online behaviors from users who do not suffer from any disorder when using social networks [2] [4].

With this scenario in mind, this work presents a Systematic Literature Review (SLR) with the objective of investigating how works are identifying signs of mental disorders on social networks. In this sense, this work provides some findings on methods, data and features, data labeling strategies and other related aspects to the theme at hand. It also provides answers to some defined research questions and a list of 19 studies, which have been selected on the topic. As a result, it is able to discuss some identified impacts, challenges and critical success factors described by the selected works when working on identifying signs of mental disorders. An important finding concerns the still recurrent lack of more appropriate and effective strategies when labeling training data in this scenario.

The remaining parts of the paper are organized as follows: Section 2 introduces some theoretical background. Section 3 discusses some related work. Section 4 describes the methodology used in this study. Section 5 presents the results obtained and a discussion relating these findings to the research questions. Finally, Section 6 provides some concluding remarks and indicates future work.

II. Theoretical foundation

A mental disorder is conventionally defined as a syndrome with characteristics that cause significant clinical changes in a person's cognition, emotion or behavior, causing psychological and biological diseases [5]. The most common mental disorders are [6] [7]: depression, insomnia, anxiety disorders, eating disorders (e.g., anorexia), bipolar disorder, schizophrenia, self-harm, post-traumatic stress disorder and drug or alcohol use disorders. The term "disorder" is used to characterize symptoms and/or behaviors that can be clinically identified and, in most cases, are linked to an individual's suffering and cognitive disturbance [7]. The discovery of these signs in advance can help in clinical assistance and treatments, making it possible to prevent recurrences or even hospital admissions [8].

To help matters, sentiment analysis strategies have been used to identify signs of some mental disorders such as depression. Sentiment analysis seeks to establish and/or use techniques capable of extracting subjective information from data such as text, images, emoticons, emojis, or audios. Examples of these data include opinions or feelings that are usually provided in natural language, which implies the need of creating structured data from those. The idea is using such data to assist some kind of decision [9]–[11].

Through these strategies, it can be determined whether a piece of writing or image is positive, negative, or neutral. The definition of a sentiment is a combination of beliefs and emotions, which, in general, may be considered good or bad according to a given real scenario (e.g., depression). In the light of depression, for instance, "loss of interest or pleasure" may be a sign of negative sentiment. On the other

 ISBN: 978-1-61208-994-2

hand, a neutral sentiment occurs when there is a lack of emotion or opinion in such a way that it can not be detected. A sentiment analysis strategy for text, for instance, usually combines Natural Language Processing (NLP) and Machine Learning Techniques to assign weighted sentiment scores to the entities, topics, themes, and categories within a sentence or phrase [12].

NLP is a subfield of computing dedicated to the natural understanding of human language by computational machines [13]. It includes techniques capable of analyzing and characterizing texts in order to perform language processing similar to the way human beings perform [14]. NLP strategies include a set of pre-processing tasks to obtain a structured representation of a less sparse set of words in the text for computational processing. Another approach used in NLP is the use of lexicons. Lexicons provide a vocabulary with preceding information about the type and intensity of each phrase or word and correlate it to a degree of polarity [15]. Words or phrases that have meaning in a lexicon are called lexemes. Such strategy is accomplished by the examination of a document or text looking for words that express a positive (e.g., good, perfect, nice, beautiful, etc.) or negative (e.g., bad, sad, etc.) feeling. In addition, automatically discovering topics from noisy and short texts posted on social networks is paramount. Due to this fact, topic modeling methods, based usually on classical probabilistic or recent neural approaches, have been considered [16]. As an illustration, the Latent Dirichlet Allocation (LDA) method is one of the most adopted in the literature. It is a generative and probability approach based on word co-occurrences [17].

The Machine Learning (ML) field relies on computational algorithms which is used to optimize a performance criterion using example data or past experience [18]. To this end, a model is produced up to some parameters, and learning is obtained by means of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both [18]. Supervised learning relies on using examples to provide predictions. Unsupervised learning does not use a target variable thus, instead of telling the machine to predict Y for specific data X, it asks what is possible to understand from data X [19]. Supervised learning is usually accomplished by classification or regression tasks [19]. For unsupervised ML, one of the most used tasks is clustering. According to [20], its objective is grouping objects that are similar to each other in clusters, based on the characteristics that these objects have. Considered as a subfield of ML, deep learning is a category of techniques that allows processing various levels of data abstractions using computational models, which can be applied in various tasks, such as associating posts and products with user preferences and also selecting content on social networks [21].

III. RELATED WORKS

In this section, some secondary works related to the SLR previously mentioned are discussed.

In [22], the authors carried out a SLR in order to discuss the state of the art on how studies used techniques for analyzing feelings and emotions to identify depressive mood disorders. The researchers analyzed which social networks, techniques, emotions and feelings were used most in discovering predecessors of depression. The SLR showed that text is the most used kind of data to identify depressive disorders. In addition, the research revealed that the selected works, for the most part, proposed strategies joining ML classifiers to lexical ones. Some gaps were also identified in the studies found, such as the little exploration of data as images, videos and emojis.

The systematic mapping produced by [23] aimed to provide insights on sentiment analysis and ML techniques used to identify profiles with a depressive tendency on social networks. This work identified which types of labeling techniques were proposed to the datasets of the studies found. The authors analyzed papers published from 2013 to 2019 and found that most of the works used supervised ML algorithms for the task of classifying depressive profiles on social networks. Lexical dictionaries, such as the Linguistic Inquiry and Word Count (LIWC), Affective Norms for English Words (ANEW) and NRC Emotion Lexicon (Emolex) were also used by the works found for the same purpose. The research showed that only one study used a dataset in Portuguese, while English was the most used language.

The research developed by [6] carried out a literature review seeking to analyze works that used social media texts for mental health surveillance, with greater attention to studies that approached depression and suicide. The researchers investigated which data collection techniques were used as well as features used in training ML models to predict population-level mental illness. They also considered which classifiers were adopted for training models. The results showed that the selected works collected data from questionnaires, such as the PHQ-9 (Patient Health Questionnaire), and scales, such as the Depression Scale Center for Epidemiological Studies (CES-D) used in psychiatry. Other forms of data collection highlighted by the review authors were the self-reports of people with depression or suicidal ideation on social networks, in addition to mental health support forums. The articles found indicated that most of the works used supervised ML algorithms combined with lexicons, exploring linguistic, semantic, behavioral and user profile features such as post volume and time.

In [7], the authors conducted a survey looking for works that proposed computational methods in the assessment of mental state from posts on social media. Most of the studies found by the researchers analyzed depressive, eating and post-traumatic stress disorders. The authors highlighted that the discussed articles performed resource extraction using topic models, such as the Latent Dirichlet Allocation (LDA) and lexical approaches such as LIWC or ANEW. They also employed techniques such as bag-of-words, word embeddings, writing behavior (frequent use of 1st person pronouns), social engagement (frequency and time of posts) and syntactic structures such as the use of negative words. The research also revealed that supervised ML algorithms, such as Support Vector

Machines (SVM), Naïve Bayes, Logistic/Linear Regression, Random Forest, Ada Boost and K-Nearest Neighbours (KNN), were the most applied ones by the works to online mental state assessment. Some articles proposed the use of deep learning models as the ones which are currently expanding. Regarding the usage of deep learning, some works used simple neural networks such as the Multilayer Perceptron (MLP) to identify mental disorders, while others adopted more complex networks, such as the Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN). In addition to ML models, some articles used rule-based approaches such as Character Language Models (CLMs), Sequential Incremental Classification (SIC), Decision List, and Temporal Mood Variation.

The current work extends some of the aspects discussed in previous related works. This work investigates various kinds of mental disorders on social networks. It also studies issues and aspects related to data, pre-processing methods, NLP techniques, labeling strategies and languages usually used to build corpora. Another concern regards the identification of linguistic terms or pronouns which may help identifying a mental disorder (e.g., depressive terms, first person pronouns). Likewise, behavioral (e.g., engagement, interval between posts) and social media features (e.g., number of reposts, comment tree) which may also be used are discussed in this work.

IV. Research Method

This work complies with the guidelines and practices established by [24] for conducting SLR. Thus, the research process accomplishes the following steps: (i) research planning, (ii) data search, and (iii) data selection, data extraction and data synthesis.

A. Research planning

The main research question that motivates this work is RQ: What is the state of the art on identifying signs of mental disorders on social networks profiles? Given the broad scope of such question, the following specific research questions may provide evidence to help answering RQ:

RQ1. What sentiment analysis strategies have been proposed/applied to detect signs of mental disorders on social media posts?

RQ2. Which behavioral characteristics or features have been used most in discovering predecessors of mental disorders?

RQ3. Which languages are most commonly used to build corpora?

RQ4. What data pre-processing strategies have been proposed/employed?

RQ5. What kind of techniques have been proposed/used for post training data labeling? What types of labels have commonly been used?

RQ6. What evaluation metrics have been used to assess the quality of results?

RQ7. What challenges and gaps have been found?

B. Data search

From the research questions presented along with some synonym adaptations, the authors extracted the constructs to be used to enable the data search, as follows:

(”sentiment analysis” OR ”text mining” OR ”emotion”) AND (”social networks” OR ” social media” OR ”social posts”) AND (”depression” OR ”depressive disorder” OR ”mental disorder”).

The strategy of this research used the following scientific web databases (libraries or proceedings): ACM Digital Library, IEEE Xplore Digital Library, Science Direct, Brazilian Journal of Information Systems (iSys), Journal of Information and Data Management (JIDM), Brazilian Conference on Intelligent Systems (Bracis), Symposium on Knowledge Discovery, Mining and Learning (KDmile), Brazilian Symposium on Multimedia and Web Systems (WebMedia), Brazilian Workshop on Social NetWork Analysis and Mining (Brasnam) and Brazilian Database Symposium (SBBD).

Two inclusion criteria were applied to filter the articles: (I1) works that answer at least one of the research questions and (I2) primary studies. The exclusion criteria used were: (E1) studies without scientific relevance; (E2) secondary or tertiary works; and (E3) articles published prior to 2016.

C. Data selection, data extraction and data synthesis

The studies collected by the search string went through a filtering process set in three phases. In Phase 1, the protocol analyzed the studies’ title, abstract, and keywords. The articles selected in this first phase went to Phase 2, in which researchers read the studies’ introduction and conclusion. In the same manner as Phase 1, this phase eliminated studies that did not answer at least one of the research questions (RQ1 to RQ7), i.e., studies that did not address the subject of this systematic review. In Phase 3, the authors read the papers completely and selected the ones which complied with the inclusion and exclusion criteria.

To assess the level of agreement among the authors and in order to rank the works which would be the most relevant to this research, an adaptation of the Likert scale was employed. Thus, three levels of evaluation were defined along with a respective value, as follows: (i) works that contributed little - value 5; (ii) works that contributed reasonably - value 10; (iii) works that contributed a lot - value 15. In addition, some quality dimensions were added to score articles based on the following criteria: 1 point, if only one of the research questions was answered, 2 points, if the work answered two or more research questions, 3 points, if it answered two or more questions and used machine learning techniques, lexical dictionaries (important to answer the RQs), and 4 points, if in addition to the criteria already mentioned, the work used the Portuguese language to build corpora.

After data extraction, some synthesis tasks were performed. The synthesis was then carried out for each RQ following specific methods adapted to each question’s proposal.

V. Results and Discussion

In total, 19 papers were found likely to contribute to this review. Table I shows the number of works collected or selected along the review process, grouped according to the scientic data source and phases employed. Figure 1 shows the publication timeline of the final selected papers. There has been an increase in works published in recent years on this subject. This may be due to the fact that the theme "metal disorders" has been increasingly highlighted on social networks in recent times, specially in the context of pandemics.

Some general findings of this SLR are pointed out in the following. Then, each one of the research questions is discussed. Table II shows the list with the references and titles of the selected works. References are used as identifiers of the selected papers in this section.

TABLE I
Number of selected studies: digital library vs phase.

Digital library	Phase 1	Phase 2	Phase 3	Final result
ACM Digital Library	477	100	24	8
IEEE Xplore	34	19	9	6
Science Direct	11	9	2	1
iSys	4	0	0	0
JIDM	7	1	0	0
Bracis	3	2	1	0
Brasnam	7	7	3	2
KDMile	2	2	1	1
SBBD	1	2	1	1
Webmedia	3	2	0	0
Total	**549**	**147**	**44**	**19**

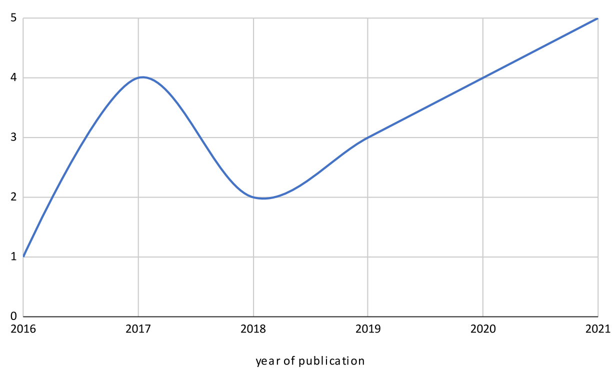

Fig. 1. Timeline of publications.

Regarding mental disorders, the selected articles revealed that depression (60%) is the most studied one. Articles that addressed anxiety (4%), bipolar disorder (4%), borderline personality disorder (4%) and stress (4%) were also found. Some studies have tried to identify positive or negative feelings, excessive sadness and behavioral change through posts on social networks. The research developed by [26] deserves to be highlighted for using data from Reddit (specific subcommunities) to analyze indicators of depression, self-harm and anorexia. In the work of [28], the authors proposed to use

TABLE II
Selected studies.

	Title
[25]	A knowledge-based recommendation system that includes sentiment analysis and deep learning
[26]	An emotion and cognitive based analysis of mental health disorders from social media data
[27]	Sentiment analysis in brazilian portuguese tweets
[28]	Characterization of anxiety, depression, and their comorbidity from texts of social networks
[29]	Depression detection using machine learning techniques on Twitter data
[30]	DepressionNet: Learning multi-modalities with user post summarization for depression detection on social media
[31]	Detecting stress based on social interactions in social networks
[32]	Early detection of depression: social network analysis and random forest techniques
[33]	Emotional and linguistic cues of depression from social media
[34]	Identifying depression among Twitter users using sentiment analysis
[4]	Mining Twitter data for signs of depression in Brazil
[35]	Modeling and detecting change in user behavior through his social media posting using cluster analysis
[36]	User emotional tone prediction models in mental health communities on Reddit
[37]	Multimodal sentiment analysis to explore the structure of emotions
[38]	Semi-Supervised approach to monitoring clinical depressive symptoms in social media
[39]	Subconscious Crowdsourcing: a feasible data collection mechanism for mental disorder detection on docial media
[40]	Tracing the emotional roadmap of depressive users on social media through sequential pattern mining
[41]	Using Twitter social media for depression detection in the Canadian population
[42]	What about mood swings: identifying depression on Twitter with temporal measures of emotions

deep learning to identify depression and anxiety. Figure 2 shows the disorders found in the 19 works that stood out the most.

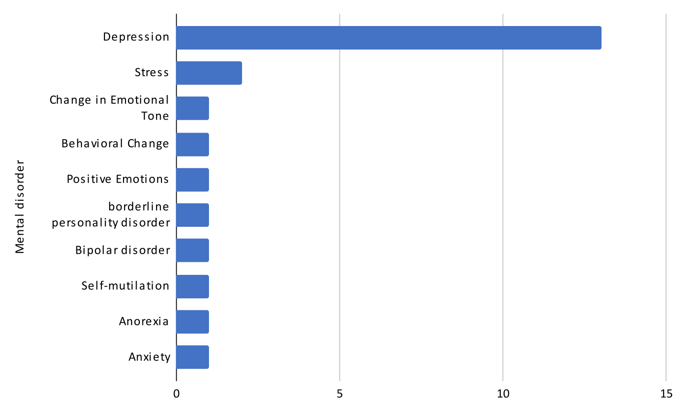

Fig. 2. Most studied mental disorders.

Most of the selected works used Twitter posts as data source (58%). The second most mentioned social network was Reddit (27%), followed by Facebook (5%). The study by [37] used the blogging platform Tumblr (5%). The authors of [31] obtained their data from Sina Weibo (5%), which is a kind of Chinese Twitter. In the following, we present the RQs and the results found by the teams in the data extraction and synthesis activities.

RQ1: For the task of predicting symptoms of mental disorders from social media posts, most researchers used combined ML approaches and lexical dictionaries, such as LIWC, VADER, ANEW and NRC Emolex. The vast majority of works used supervised ML algorithms, such as SVM, Naive Bayes, KNN and ensembles, such as Random Forest, Gradient Boosting and AdaBoost. Three works used unsupervised ML with the K-means algorithm. The work developed by [38] proposed a topic modeling using LDA to detect clinical depression from Twitter posts. Articles that applied deep learning through neural networks were also selected: CNN, LSTM and MLP. Figure 3 depicts some topic modeling and ML methods found, grouped by supervised, unsupervised and deep learning that were most used in the 19 highlighted works. It is also noticed that the use of deep learning models for this type of task has been increased. The reason underlying that may be linked to how these methods perform the analysis and learning.

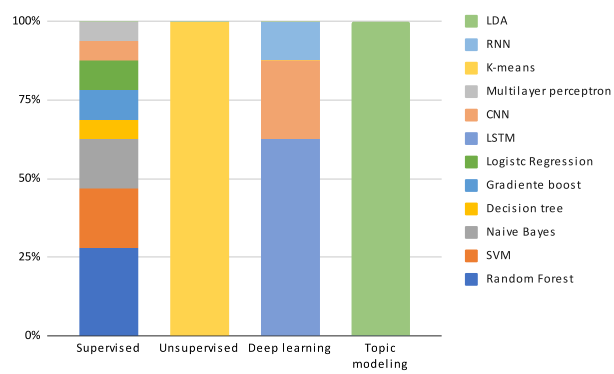

Fig. 3. Topic modeling and machine learning methods.

RQ2: Research has identified features to perform data analysis and explore them through some sentiment analysis strategy. The behavior patterns of users on social networks were essential in identifying some mental disorders. Table III presents the main features found that are grouped into 3 categories: Behavioral, that means how people act on networks; Linguistics, regarding the way they write their posts; and Features in the social network about information related to personal profiles. However, such properties were little explored over a certain period in the timeline of user posts on their profile. Interaction-related features, such as responses and comments between users with mental disorders and their friends, were also under-examined. The analysis of these characteristics taking into account the timeline of posts can bring new characteristics and discoveries of behavioral patterns to the study of mental disorders from social networks.

RQ3: The language most found in the data analyzed by the works was English (79%). For the Portuguese language, there were 3 articles (16%), highlighting the research by [4] that analyzed Twitter posts to predict signs of depression, and [25] that used facebook posts to predict signs of depression

TABLE III
FEATURES CATEGORIES.

Category	Feature
Behavioral	Time of posts, engagement, interval between posts, frequency on the networks, insomnia index, number of publications;
Linguistics	Antidepressant drug names, depressive terms, first person pronouns, syntactic features (verb, adverb);
Social network	Number of followers, number of reposts, comment tree, interaction with friends.

and stress. Only one study evaluated data by considering the Chinese language. There is a lack of research focused on studying mental disorders from social networks, exploring, in particular, datasets in Portuguese. This might be due to the difficulty in working with this language that has nuances, such as slang and regional dialects. According to the region of the country, the same word has different meanings, and can express different feelings according to the place where the user of the social network lives. For a more accurate computational diagnosis, this cultural diversity becomes a challenge, as the linguistic variety needs to be taken into account in steps such as pre-processing and labeling, when using ML algorithms. Otherwise, there is the possibility of erroneous analysis, and, consequently, false computational diagnoses. In addition, there is a shortage of lexical dictionaries in the Portuguese language, specifically to deal with mental disorders. Thus, here is an open field to improve existing solutions. The development of lexicons in the Portuguese language, for specific domains e.g. depression, anxiety, anorexia, presents itself as a promising line of research.

RQ4: Before being analyzed by models such as ML algorithms, the data collected from social media posts usually go through the pre-processing step. The selected works showed that techniques, such as stopword removal, tokenization, lemming, stemming and Term Frequency - Inverse Document Frequency (TF-IDF) are the ones that appear the most. The articles also showed that it was necessary to anonymize the posts by removing user names, mentions, URLs or any information that could identify the owner of the profile on the social network. Emojis and emoticons in most of the works were removed or transformed into text. The work of [34] stands out, using the demoji, a library developed in Python, and transformed the emoticons and emojis that appear in text. The exclusion or conversion of hashtags, emojis and emoticons in texts, or even exclusion of words with repeated letters, indicating intensity, could be better handled to provide a dataset with more subsidies and characteristics that would help in the evaluation of ML models.

RQ5: Labeling strategies used in the selected studies were largely done through self-reports of social network users, parsing and identifying sentences, such as: “I was diagnosed with depression” on their profiles. Some research carried out the labeling manually and had the help of experts to identify whether or not a particular post showed signs of a mental disorder. Only one article found focused on a proposal for automatic labeling using Textblob to punctuate the sentence and label it. We highlight the work of [29] who used Textblob, a library used by Python to calculate the sentiment score of a text according to subject and polarity, thus classifying sentences as -1 (negative), 0 (neutral) and 1 (positive). When the calculated value was less than 0, the sentences were labeled as depressive and when greater than 0, non-depressive. Regarding labeling training data, works have demonstrated that this is indeed the largest bottleneck in deploying machine learning models applied to sentiment analyses. Manual labeling can present subjectivity and different interpretations by human annotators. Experts are supposed to significantly enhance this process, although it demands a lot of time and effort to accomplish this task. The volume of data may be a restriction. In turn, the use of libraries such as Textblob can speed up the labeling step, however, it needs to be developed in such a way that may provide specificities to the data domain at hand. Thereby, there is a lack for more research and proposals to label training data to the mental disorders scenario. An automated or even semi automated strategy for labeling training data on mental disorders may bring two main benefits: (a) reducing the effort employed in the usual labeling step; and (b) contributing to find out new examples that could not be easily found by performing manual labeling. As a result, it may also increase the performance of the classifier used to predict signs of mental disorders.

RQ6: In order to measure the performance of ML models, evaluation metrics were used. Most articles used precision, recall, accuracy and F-measure. A possible explanation for this is that for the task of predicting signs of mental disorders, the classification problem is usually binary. Besides the mentioned metrics, in [26], the Area under the Receiver Operating Characteristic curve (AUC) was used to evaluate the performance of the models proposed in the research. The authors considered this metric to be less sensitive to the imbalance of the data set used in the work. Another highlight was the research in [33] which, in addition to accuracy, calculated the micro-F1 and macro-F1 scores to evaluate the Gradient Boosted Decision Trees classifier in the task of identifying emotional and linguistic signs of depression on Twitter.

RQ7: The main challenges are related to the development of approaches that can allow a more automated training data labeling, perhaps using lexical dictionaries specific to the mental disorder. Automatic training data labeling avoids the situation of depending on user reports on their networks or the need of performing manual annotations in large datasets. Likewise, there is lack of studies specifically focused on signs, emotions and linguistic, behavioral and associated with the timeline of user posts. Thus, it is worth studying some specific chronology on users to enhance such analysis on the propensity to mental disorders. The construction of a lexical dictionary in Portuguese specialized on a mental disorder such as depression as well as the investigation of behavioral resources and social interaction over a period of time using a database in Portuguese are gaps that still need to be filled.

VI. CONCLUSION

The analysis of data from social networks focused on research to detect signs of mental disorders emerges as a promising topic in the field of sentiment analysis. This work carried out an SLR with the objective of investigating how the literature presents the sentiment analysis techniques proposed to identify mental disorders from posts on social networks. The characteristics and languages used were analyzed, revealing that few studies with data in Portuguese were found. The pre-processing methods were also analyzed, showing that features such as emoticons and emojis can be further explored. Training data labeling was accomplished manually or through user reports. Furthermore, it was revealed that the works are proposing the combination of ML methods and lexical dictionaries to detect signs of propensity to mental disorders. The most discussed disorder found in the studies was depression. The works identified behavioral and linguistic characteristics that helped in the early detection of depression. Thus, future investigations of this disorder are relevant. We also sought to find out which metrics evaluated the performance of the strategies proposed in the works. Finally, challenges and gaps to be filled according to the work obtained were identified. In future work, we can explore the development of automated or semi-automated training data labeling strategies, which are indeed important to detect mental disorders such as depression on social media.

REFERENCES

[1] J. W. G. Duque, A. L. Raymundo, and P. F. Neto, "An application of big data for twitter depressive sentence classification," *H-TEC humanities and technology magazine*, vol. 2, no. 1, pp. 82–95, 2018.

[2] M. De Choudhury, M. Gamon, S. Counts, and E. Horvitz, "Predicting depression via social media," in *Seventh international AAAI conference on weblogs and social media*, 2013, pp. 128–137.

[3] WHO, "World health organization," 2022, https://www.who.int/news-room/fact-sheets/detail/mental-disorders Last accessed 04 Sep 2022.

[4] O. V. Sperling and M. Ladeira, "Mining twitter data for signs of depression in brazil," in *Anais do VII Symposium on Knowledge Discovery, Mining and Learning*. SBC, 2019, pp. 25–32.

[5] American Psychiatric Association, *Diagnostic and statistical manual of mental disorders: DSM-5*. American psychiatric association Washington, DC, 2013, vol. 5.

[6] R. Skaik and D. Inkpen, "Using social media for mental health surveillance: a review," *ACM Computing Surveys (CSUR)*, vol. 53, no. 6, pp. 1–31, 2020.

[7] E. A. Ríssola, D. E. Losada, and F. Crestani, "A survey of computational methods for online mental state assessment on social media," *ACM Transactions on Computing for Healthcare*, vol. 2, no. 2, pp. 1–31, 2021.

[8] S. Abdullah and T. Choudhury, "Sensing technologies for monitoring serious mental illnesses," *IEEE MultiMedia*, vol. 25, no. 1, pp. 61–75, 2018.

[9] M. Huang, H. Xie, Y. Rao, J. Feng, and F. L. Wang, "Sentiment strength detection with a context-dependent lexicon-based convolutional neural network," *Information Sciences*, vol. 520, pp. 389–399, 2020.

[10] A. V. Tardelli, A. F. d. S. Dias, and J. B. d. S. França, "Introduction to sentiment analysis with word clouds," *Brazilian computing society*, pp. 38–67, 2019.

[11] F. Benevenuto, F. Ribeiro, and M. Araújo, "Methods for sentiment analysis in social media," *Brazilian computing society*, pp. 31–59, 2015.

[12] D. Sharma, M. Sabharwal, V. Goyal, and M. Vij, "Sentiment analysis techniques for social media data: A review," in *First international conference on sustainable technologies for computational intelligence*. Springer, 2020, pp. 75–90.

[13] T. Beysolow II, *Applied Natural Language Processing with Python: Implementing Machine Learning and Deep Learning Algorithms for Natural Language Processing*. Apress, 2018.
[14] E. D. Liddy, *Natural language processing*. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc., 2001.
[15] A. Yadollahi, A. G. Shahraki, and O. R. Zaiane, "Current state of text sentiment analysis from opinion to emotion mining," *ACM Computing Surveys (CSUR)*, vol. 50, no. 2, pp. 1–33, 2017.
[16] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short-text data: A comparative analysis," *Frontiers in Artificial Intelligence*, vol. 3, p. 42, 2020.
[17] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
[18] E. Alpaydin, *Introduction to machine learning, 2nd edition*. MIT Press Cambridge, 2010.
[19] P. Harrington, *Machine learning in action*. Simon and Schuster, 2012.
[20] R. Linden, "Clustering techniques," *FSMA Information Systems Magazine*, vol. 4, no. 4, pp. 18–36, 2009.
[21] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[22] F. T. Giuntini *et al.*, "A review on recognizing depression in social networks: challenges and opportunities," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 11, pp. 4713–4729, 2020.
[23] V. Casani, R. G. Mantovani, A. C. C. Souza, and F. C. M. Souza, "Identification of depressive profiles in social networks using machine learning: a systematic mapping," *Anais do Computer on the Beach*, vol. 11, no. 1, pp. 183–190, 2020.
[24] B. Kitchenham, "Procedures for performing systematic reviews," *Keele, UK, Keele University*, vol. 33, no. 2004, pp. 1–26, 2004.
[25] R. L. Rosa, G. M. Schwartz, W. V. Ruggiero, and D. Z. Rodríguez, "A knowledge-based recommendation system that includes sentiment analysis and deep learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2124–2135, 2018.
[26] A.-S. Uban, B. Chulvi, and P. Rosso, "An emotion and cognitive based analysis of mental health disorders from social media data," *Future Generation Computer Systems*, vol. 124, pp. 480–494, 2021.
[27] D. P. Kansaon, M. A. Brandão, and S. A. de Paula Pinto, "Sentiment analysis in brazilian portuguese tweets," in *Anais do VII Brazilian Workshop on Social Network Analysis and Mining*. Porto Alegre, RS, Brasil: SBC, 2018, pp. 37–48. [Online]. Available: https://sol.sbc.org.br/index.php/brasnam/article/view/3578
[28] V. B. Souza, J. Nobre, and K. Becker, "Characterization of anxiety, depression, and their comorbidity from texts of social networks," *Anais do XXXV Simpósio Brasileiro de Banco de Dados. SBC, Porto Alegre, RS, Brasil*, pp. 121–132, 2020.
[29] K. A. Govindasamy and N. Palanichamy, "Depression detection using machine learning techniques on twitter data," in *2021 5th international conference on intelligent computing and control systems (ICICCS)*. IEEE, 2021, pp. 960–966.
[30] H. Zogan, I. Razzak, S. Jameel, and G. Xu, "Depressionnet: learning multi-modalities with user post summarization for depression detection on social media," in *proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 2021, pp. 133–142.
[31] H. Lin *et al.*, "Detecting stress based on social interactions in social networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1820–1833, 2017.
[32] F. Cacheda, D. Fernandez, F. J. Novoa, and V. Carneiro, "Early detection of depression: social network analysis and random forest techniques," *Journal of medical Internet research*, vol. 21, no. 6, p. e12554, 2019.
[33] N. Vedula and S. Parthasarathy, "Emotional and linguistic cues of depression from social media," in *Proceedings of the 2017 International Conference on Digital Health*, 2017, pp. 127–136.
[34] F. Azam, M. Agro, M. Sami, M. H. Abro, and A. Dewani, "Identifying depression among twitter users using sentiment analysis," in *2021 International Conference on Artificial Intelligence (ICAI)*. IEEE, 2021, pp. 44–49.
[35] D. J. Joshi, N. Supekar, R. Chauhan, and M. S. Patwardhan, "Modeling and detecting change in user behavior through his social media posting using cluster analysis," in *Proceedings of the Fourth ACM IKDD Conferences on Data Sciences*, 2017, pp. 1–9.
[36] B. Silveira, A. P. C. da Silva, and F. Murai, "Models for predicting the emotional tone of users in mental health communities on reddit," in *Anais do IX Brazilian Workshop on Social Network Analysis and Mining*. SBC, 2020, pp. 13–24.
[37] A. Hu and S. Flaxman, "Multimodal sentiment analysis to explore the structure of emotions," in *proceedings of the 24th ACM SIGKDD international conference on Knowledge Discovery & Data Mining*, 2018, pp. 350–358.
[38] A. H. Yazdavar *et al.*, "Semi-supervised approach to monitoring clinical depressive symptoms in social media," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 2017, pp. 1191–1198.
[39] C.-H. Chang, E. Saravia, and Y.-S. Chen, "Subconscious crowdsourcing: A feasible data collection mechanism for mental disorder detection on social media," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2016, pp. 374–379.
[40] F. T. Giuntini *et al.*, "Tracing the emotional roadmap of depressive users on social media through sequential pattern mining," *IEEE Access*, vol. 9, pp. 97 621–97 635, 2021.
[41] R. Skaik and D. Inkpen, "Using twitter social media for depression detection in the canadian population," in *2020 3rd Artificial Intelligence and Cloud Computing Conference*, 2020, pp. 109–114.
[42] X. Chen, M. D. Sykora, T. W. Jackson, and S. Elayan, "What about mood swings: Identifying depression on twitter with temporal measures of emotions," in *Companion Proceedings of the The Web Conference 2018*, 2018, pp. 1653–1660.