Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 179 (2021) 582–589

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on Computer Science and Computational Intelligence 2020

Text-based Depression Detection on Social Media Posts: A Systematic Literature Review

David William[a,] *, Derwin Suhartono[b]

[a]*Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*
[b]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia*

Abstract

Due to the huge increase of awareness of mental health well-being, the detection of mental illness itself is starting to become a huge concern. Many psychiatrists found difficulties in identifying the existence of mental illness in a patient because of the complicated nature of each mental disorder, thus making it hard to give the appropriate treatment to the patient before it's too late. However, due to the integration of social media into people's daily life, this create an environment that may provide additional information regarding the mental disorder a patient bear. This study has been undertaken as a Systematic Literature Review (SLR), which is defined as a process of identifying, assessing, and interpreting the available resources to provide answers for a set of research questions. Analysis is made to answer questions regarding text-based mental illness detection based on the social media activity of people with mental disorders, and reveals that it indeed is possible to do early detection of depression on social due to the existence of a particular characteristics in the way these subjects use their social media. This SLR found that from the small amount of research using text-based approach, most studies use deep learning models such as RNN on the early detection of depression cases due to the limitation of data availability. However, this study will look to find method that may prove to be more effective.

Keywords: : Linguistic analysis; Natural language processing; Depression detection; Social media

* Corresponding author. Tel.: +6287888020602
E-mail address: david.william@binus.ac.id

10.1016/j.procs.2021.01.043

1. Introduction

In our current society, depression is a common condition that many people have. According to Parekh, depression is a medical condition that may negatively affect a person way of thinking, feeling, and/or acting [1]. Based on the data provided by *World Health Organization*, it is recorded that there are around 322 million cases in the span of 2015, with around 788.000 cases ended with suicide [2]. However, serious as it may seem, there still is a stigma going around society where having mental disorder shows weakness and often may lead to exclusion from community. A study shows that though people see depression as a problem that indeed is serious, they tend to think it is less amenable to treatment [3]. This may cause people with mental disorder to feel reluctant to seek help from professional, thus making even less people to be exposed to proper treatment. Around 75% - 85% of people with depression does not receive proper help when fighting with their disorder [4].

With the current condition where people tend to go into social media to vent about their problems, this may not only provide psychiatrists and/or psychologists with additional information prior to making decisions, but also opens the possibilities of early detection by using data obtained from the subject's social media platform [5]. It is reinforced by a finding that shows students with depressive symptoms does use the internet much more than those without symptoms [6]. This drives researchers to find the best method for early depression of detection [7] [8] [9]. This SLR aims to identify and analyze text-based approaches that are implementable for early detection of depression based on social media's posts.

2. Methods

For the objective of this SLR, the author first decided on a set of research questions to act as the guideline for the entire process in this research. Since this research aims to answer the question of "what might influence the usage of social media as a method for early detection of depression", the following set of research questions are formulated:

- RQ1 – What are the challenges in doing text-based approach for depression detection?
- RQ2 – What are the most effective text-based approach for early depression detection?

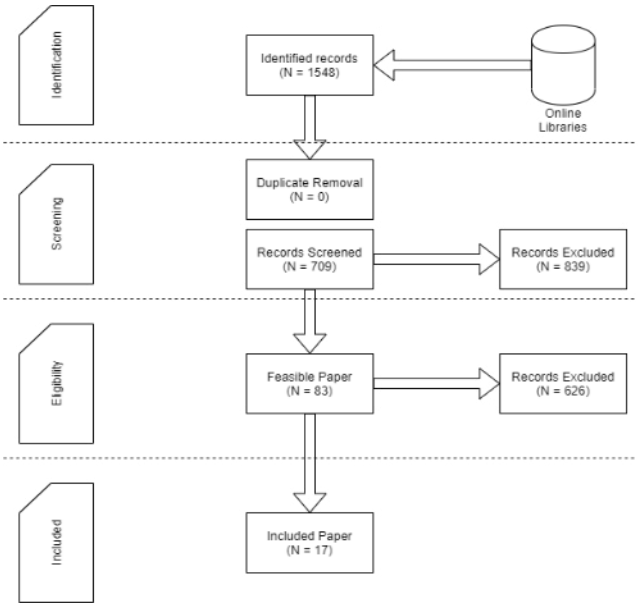

Figure 1. PRISMA Flowchart

This literature review will be using PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) as the main framework, to better organize and document the whole process and records used in this research. The steps of this research are shown in Figure 1.

The search for articles used in this study is done by looking through online libraries such as: 1) Science Direct (www.sciencedirect.com), 2) ACM Digital Library (dl.acm.org), 3) IEE Xplore Digital Library (ieeexplore.ieee.org), 4) Springer Link (link.springer.com), 5) Emerald Insight (www.emeraldinsight.com). Using ["depression" OR "depressed"] AND ["mental*" OR "disorder"] AND ["detection"] AND ["social media"] as the base search query among these libraries, and modified the query according to the supported format for each platform a set of records is found and further processed in the next step.

The collection of records will then be further filtered through a certain set of criteria, where records that are not research article written in English will be excluded in this phase. Some other criteria that may exclude a search result from being included for this review are:

- Written before 2015
- Duplicate publication of the same research
- Other depression related issues

Articles that does not meet any of the exclusion criteria will then be further filtered in the next phase by abstract screening. First, the full-text availability will of each records will be checked, and if it is available, author will read through the abstract of each records, then will classify whether the content of record in context align with the purpose of this review, and whether the content may be able to answer the research question that is predefined in this research. A record will be included in this review if it fulfils those two stated criteria. Table 1 shows the detail of achieved result in the process of record seeking.

Table 1. Selected Records

Source	Found	Candidate	Selected
Science Direct	449	35	6
IEEE Xplore	1	0	-
ACM Digital Library	181	24	6
Emerald Insight	146	3	-
Springer Link	771	21	5
Total	1548	83	17

3. Result

3.1 List of Paper Publications

As seen in Table 1, there are 17 research publications that are selected to be studied for this research. Table 2 shows that 17 publications.

Table 2. List of Paper Publications

Source	Year	Title
Science Direct	2015	Detecting Suicidality on Twitter [10]
Science Direct	2017	Psychiatric Symptom Recognition Without Labeled Data Using Distributional Representations of Phrases and on-line Knowledge [11]
Science Direct	2018	A Large-Scale Social Media Corpus for the Detection of Youth Depression (Project Note [12])
Science Direct	2018	An Automated Psychometric Analyzer based on Sentiment Analysis and Emotion Recognition for Healthcare [13]
Science Direct	2019	Detecting Arabic Depressed Users from Twitter Data [14]
Science Direct	2020	Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms [15]
ACM Digital Library	2018	Beyond the Coded Gaze: Analyzing Expression of Mental Health Illness on Instagram [16]

ACM Digital Library	2018	Personal Informatics in Interpersonal Contexts: Towards the Design of Technology that Supports the Social Ecologies of Long-Term Mental Health Management [17]
ACM Digital Library	2019	Exploring Indicators of Digital Self-Harm with Eating Disorder Patients: A Case Study [18]
ACM Digital Library	2019	Leveraging Routing Behavior and Contextually-Filtered Features for Depression Detection among College Students [19]
ACM Digital Library	2019	Prediction of Mood Instability with Passive Sensing [20]
ACM Digital Library	2019	Who is the "Human" in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media [21]
Springer Link	2015	Teenager's Stress Detection Based on Time-Sensitive Micro-blog Comment/Response Actions [22]
Springer Link	2016	A Systematic Exploration of the Micro-blog Feature Space for Teens Stress Detection [23]
Springer Link	2017	Latent Sentiment Topic Modelling and Nonparametric Discovery of Online Mental Health-related Communities [24]
Springer Link	2020	Understanding Depression from Psycholinguistic Patterns in Social Media Texts [25]
Springer Link	2020	Utilizing Temporal Psycholinguistic Cues for Suicidal Intent Estimation [26]

3.2 Result of Individual Studies

In this stage of the research, author will study each individual record and try to extract data from every record listed in Table 2. Data extraction is done to answer the predefined research questions (RQ) stated in early chapter 2.

3.2.1. What are the challenges in doing text-based approach for depression detection?

Table 3. Difficulties in text-based depression detection

No	Issue	Number of papers	Study identifiers
1	No means to verify the data validity	1	[10]
2	Data unable to provide sample characteristics	2	[10], [19]
3	Ethical concerns	6	[10], [16], [17], [18], [20], [21]
4	Insufficient data	4	[11], [12], [18], [19]
5	Stigma and/or lack of awareness	3	[14], [20], [21]
6	Severity of error	2	[16], [17]
7	Oversimplification	1	[21]

Table 3 shows the difficulty stated in the selected records. Though not every article stated the difficulty that occurs in the research process, it shows that most concerns lies in ethical concerns. This covers every issue regarding privacy protection and data accessibility. Even though most data used is made public on social networking sites, data related to mental well-being and disorders are regarded as sensitive data.

After further observation, it seems that most of the difficulties stated in Table 3 are related to one another. Records number [10] for example, stated that it is difficult to verify whether the Twitter user is indeed suffering from depression due to ethical concerns. Hence, data collection is done by using API and may not provide a good sample characteristic and/or data distribution.

3.2.2. What is the most effective text-based approach for early depression detection?

Table 4. Methods used for text-based depression detection

No	Method	Study identifiers
1	Scikit-Learn Toolkit machine classification	[10]
2	Classifiers	[11], [14], [19], [23]
3	Support Vector Machine	[13], [14], [15], [22]
4	Random Forest	[14], [15]
5	Probabilistic classifier	[14], [15], [22], [24]
6	Clustering Approach	[15], [24]
7	Association rule mining	[19]
8	Logistic	[22]
9	Gaussian Process	[22]
10	Term Frequency – Inverse Document Frequency	[25]
11	BiLSTM + Attention	[26]

Table 4 records the methods that is used when doing text-based approach for depression detection. The most popular methods used in the set of selected articles are the usage of classifiers, support vector machine (SVM), and probabilistic approach (Bayesian, Hierarchical Dirichlet Process, etc.), while the best result is obtained by using BiLSTM + Attention model . However, this might be overgeneralizing due to difference in dataset used and difference in the tackled issue.

4. Initial Experiment

Results obtained from the literature review stated that BiLSTM + Attention model performs well on depression related textual data. Even though the achieved result may be satisfactory, there are certain issues with the model implemented in that research.

Recurrent Neural Network Model (RNN) based model suffers from a problem where they tend to have difficulties finding proper context from a word in a long sentence. This is caused by the nature of RNNs where they process data according to the order of words in a sequence. Aside from difficulty in finding context, this nature of RNN also makes the model take longer to train.

Another issue with this model is regarding its way of approaching a sequence. Even though having the term "Bidirectional" in its name, BiLSTM actually implements a LSTM model that processes a sequence from both directions making this model not genuinely bidirectional. The fact that it approaches a text from each direction makes it even harder for this model to obtain the context of a word, since there is bound to be more information.

This entice author to further improve the accuracy by diving into an experiment using a model that is based on BERT (Bidirectional Encoder Representations from Transformers). BERT itself have been proven to perform not only faster, but also solves both issues stated above. However, the experiment done will serves only as the initial experiment to support the findings of this literature review.

This experiment used a BERT-based model for classification which is already pretrained by Google [27]. The model is made to support transfer learning, which is why it underwent pre-training process by using both the BookCorpus and the English Wikipedia to help the model obtain an understanding of the English language. This training process takes a lot of resource and time, which make it more efficient to fine-tune a pre-trained model to a specific downstream. In this experiment, the model will be fine-tuned toward depression detection using a dataset crawled from Reddit [28].

BERT is built by stacking a set of encoders on top of each other, meaning it have self-attention layer in it, which improve the time needed to train a model and the performance of said model, at the expense of limiting capabilities in processing sequence which is longer than 512 tokens. It is found that the most optimal solution to this issue is by having the sequence truncated to the desired length [29]. However, this may cause the sequence to lose some critical information. This experiment will take a new approach by applying extractive summarization on sequences longer

than 512 tokens to adjust the length of said sequence. Table 5 shows the summarization result by displaying both the raw data, and the result obtained by summarizing said data.

Table 5. Raw data and summarized data

Before	*Need to Change Depression meds....but SCARED..HELP!!! I have been on Celexa 40mg for over 18 YEARS. It's NOT working and I really need to change but I'm so scared that I'll get even WORSE!! My Dr. has given me a month of Prozac but I never took it because of fear. This month he gave me a new drug Trintellix but I haven't started it yet either. I'm SO ON THE EDGE that I'm too scared to try something new because if it doesn't work (or God forbid, makes me WORSE.....) then I just know I for sure won't make it. What should I do? Has anyone else been too scared to try other new or different meds??*
After	*I'm SO ON THE EDGE that I'm too scared to try something new because if it doesn't work (or God forbid, makes me WORSE.....) then I just know I for sure won't make it*

For summarization, this experiment opted to use extractive method over the abstractive method, to maintain as much information without altering any part of the sequence. The summarization process uses a BERT-based model which is BERT-for-extractive-summarization [30], which is a model that uses BERT's text embedding, then taking it into K-Means algorithm to cluster these embedding, resulting an extractive summarized sequence. Data preprocessing is done by configuring the text summarization, which is set to a ratio of 0.2, minimum length of 20 characters, and a max length of 500. This means the model will first summarize the data to 20% of its raw form, while removing all sequences shorter than 20 characters or longer than 500 characters.

As stated above, the model then undergoes the process of fine-tuning by post-training the model using the preprocessed data which consists of 3412 data with 1884 data labelled as depressed, while the other 1528 is labelled as control. The training is then done by using a learning rate of $2x10^{-5}$, epsilon of $1x10^{-8}$, and is repeated until it reaches 4 epochs with a batch size of 8, while it is evaluated by doing a train-test split of 90% - 10% ratio. Another 10% of the training set is then separated to act as the validation set, to keep the model from overfitting. As shown in Figure 2, the model achieves a training accuracy of 99% and validation accuracy of 92% on the fourth epoch, while Figure 3 shows that the training process had a loss value of 0.02 while the validation loss is at 0.40 on the fourth epoch. Each epoch takes around 5 minutes which totals to a 20 minutes of model fine-tuning. Even though upon evaluation on the test set, the model shows a satisfying result, the configuration of summarization is suspected to affect some aspect of this model; thus, it is necessary to find the optimal configuration for performance optimization.

Figure 3. Training and validation accuracy

Figure 2. Training and validation loss

5. Conclusion

This systematic literature review observes the current implementation of text-based approach for early depression detection. The study started from 1548 obtained records through online databases search, reach a total of 709 articles after further exclusion through a set of criteria, then filtered through title and abstract to 17 included articles which each of its content will be analyse to answer the research question defined in this study.

Based from the finding in this literature review, it is found that the 3 most concerning issues are regarding (1) Ethical concerns, (2) Lack of data, (3) Stigma and/or awareness of mental well-being. This study also gave an overview of the currently used methods in text-based depression detection. Even though it reported that using (1) Classifiers, (2) Support Vector Machine, and (3) Probabilistic Classifier proves to be the most popular approach, it is interesting that the BiLSTM + Attention method proves to yield the best result.

This study also made an experiment by taking a BERT-based model and not only fine-tuned it for depression detection, but also suggest a new method to deal with long sequences by summarizing the text before feeding it into the model. Current result is better than any text-based depression detection model, but still needs optimization such as summarizer configuration or hyperparameter tuning to further improve the performance of this model.

References

1. Parekh R. American Psychiatric Association. [Online].; 2017 [cited 2020 6 5. Available from: HYPERLINK "https://www.psychiatry.org/patients-families/depression/what-is-depression" https://www.psychiatry.org/patients-families/depression/what-is-depression .
2. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates Geneva: World Health Organization; 2017.
3. Edwards S, Tinning L, Brown JSL, Boardman J, Weinman J. Reluctance to Seek Help and the Perception of Anxiety and Depression in the United Kingdom. The Journal of Nervous and Mental Disease. 2007;: p. 258-261.
4. Philip S. Wang SAGJEW. Worldwide Use of Mental Health Services for Anxiety, Mood, and Substance Disorders: Results from 17 Countries in the WHO World Mental Heal (WMH) Surveys. The Lancet. 2007;: p. 841-850.
5. Choudhury MD. Role of Social Media in Tackling Challenges in. Proceedings of the 2nd International Workshop on Socially-Aware Multimedia (SAM '13). 2013;: p. 49-52.
6. Raghavendra Kotikalapudi SCFMDWKL. Associating Internet Usage with Depressive Behavior Among College Students: IEEE Tech & Society Magazine; 2012.
7. Glen Coppersmith MDCH. Quantifying Mental Health Signs in Twitter. Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014;: p. 51-60.
8. Glen Coppersmith MDCHKHMM. CLPsych 2015 Shared Task: Depression and PTSD on Twitter. Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015;: p. 31-39.
9. Judy Hanwen Shen FR. Detecting anxiety on Reddit. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology. 2017;: p. 58-65.
10. O'Dea B, Wan S, Batterham PJ, Calear AL, Paris C, Christensen H. Detecting Suicidality on Twitter. Internet Interventions 2. 2015 April.
11. Zhang Y, Zhang O, Wu Y, Lee HJ, Xu J, Xu H, et al. Psychiatric Symptom Recognition Without Labeled Data Using Distributional Representations of Phrases and on-line Knowledge. Journal of Biomedical Informatics. 2017 June.
12. Zaghouani W. A Large-Scale Social Media Corpus for the Detection of Youth Depression (Project Note). In The

4th International Conference on Arabic Computational Linguistics (ACLing 2018); 2018 November; Dubai. p. 347-351.

13. Vij A, Pruthi J. An Automated Psychometric Analyzer based on Sentiment Analysis and Emotion Recognition for Healthcare. In International Conference on Computational Intelligence and Data Science (ICCIDS 2018); 2018; Haryana. p. 1184-1191.
14. Almouzini S, Khemakhem M, Alageel A. Detecting Arabic Depressed Users from Twitter Data. In 16th International Learning & Technology Conference 2019; 2019. p. 257-265.
15. Priya A, Garg S, Tigga NP. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. In International Conference on Computational Intelligence and Data Science (ICCIDS 2019); 2019. p. 1258-1267.
16. Feuston JL, Piper AM. Beyond the Coded Gaze: Analyzing Expression of Mental Health Illness on Instagram. In Proceedings of the ACM on Human-Computer Interaction; 2018. p. 51-51:21.
17. Murnane EL, Walker TG, Tench B, Voida S, Snyder J. Personal Informatics in Interpersonal Contexts: Towards the Design of Technology that Supports the Social Ecologies of Long-Term Mental Health Management. In Proceedings of the ACM on Human-Computer Interaction; 2018. p. 127-127:27.
18. Pater JA, Farrington B, Brown A, Reining LE, Toscos T, Mynatt ED. Exploring Indicators of Digital Self-Harm with Eating Disorder Patients: A Case Study. In Proceedings of ACM Human-Computer Interaction; 2019. p. 84-84:26.
19. Xu X, Chikersal P, Doryab A, Villalba DK, Dutcher JM, Tumminia MJ, et al. Leveraging Routing Behavior and Contextually-Filtered Features for Depression Detection among College Students. In Proceedings of ACM Interact. Mon. Wearable Ubiquitous Technol.; 2019. p. 116-116:33.
20. Morshed MB, Saha K, Li R, D'Mello SK, Dhoudhury MD, Abowd GD, et al. Predictions of Mood Instability with Passive Sensing. In Proceedings of ACM Interact. Mob. Wearable Ubiquitous Technol.; 2019. p. 75-75:21.
21. Chancellor S, Baumer EPS, Choudhury MD. Who is the “Human” in Human-Centered Machine Learning: The Case of Predicting Mental Health from Social Media. In Proceedings of ACM Human-Computer Interaction; 2019. p. 147-147:32.
22. Zhao L, Jia J, Feng L. Teenager’s Stress Detection Based on Time-Sensitive Micro-blog Comment/Response Actions. In IFIP International Conference on Artificial Intelligence in Theory and Practice; 2015. p. 26-36.
23. Zhao L, Li Q, Xue Y, Jia J, Feng L. A Systematic Exploration of the Micro-blog Feature Space for Teens Stress Detection. Health Information Science and Systems. 2016; 4(3).
24. Dao B, Nguyen T, Venkatesh S, Phung D. Latent Sentiment Topic Modelling and Nonparametric Discovery of Online Mental Health-related Communities. International Journal of Data Science Analytics. 2017 September; 4(209).
25. Trifan A, Antunes R, Matos S, Oliveira JL. Understanding Depression from Psycholinguistic Patterns in Social Media Texts. In European Conference on Information Retrieval; 2020. p. 402-209.
26. Mathur P, Sawhney R, Chopra S, Leekha M, Shah RR. Utilizing Temporal Psycholinguistic Cues for Suicidal Intent Estimation. In European Conference of Information Retrieval; 2020. p. 265-271.
27. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT. 2019;: p. 4171-4186.
28. Partlow A, Chin J, Hai MS, Angeles R. Covial Media Emotion Analysis Machine Learning.; 2018.
29. Sun C, Qiu X, Xu Y, Huang X. How to Fine-Tune BERT for Text Classification. 2019.
30. Miller D. Leveraging BERT for Extractive Text Summarization on Lectures. arXiv preprint arXiv: 1908.08345v2. 2019.