

Detecting depression using natural language processing (NLP) techniques

Table of contents:

- Summary research paper's
- Critical Analysis
- Connections
- Conflicts between the findings
- Synthesis Matrix for Literature Review
- References

Summary research papers:

Paper1: Screening for Depression Using Natural Language Processing: Literature Review

This paper discusses a comprehensive literature review on the use of Natural Language Processing (NLP) techniques for screening and detecting depression from textual data, sourcing most of their data from social media posts, clinical interviews, and textual mental health records. These datasets require specific NLP techniques such as sentiment analysis, linguistic markers, word embedding, and advanced models like CNNs, RNNs, and large language models. This finding contributes to a powerful toolkit for detecting not only depression but also is a helpful tool to early detect mental illness disorders. In addition, the authors discuss the ethical concerns of the individual, lack of textual understanding, and bias and fairness issues. Overall, this paper concludes with how powerful and potential NLP is in improving mental illness detection, but further work is needed to address ethical, technical, and cross-cultural challenges (Teferra et al., 2024).

paper 2: Natural Language Processing as an Emerging Tool to Detect Late-Life Depression

In this paper, the authors discuss the use of natural language processing (NLP) to detect late-life depression (LLD) compared to younger adults using speech. Therefore, studying a speech dataset will need computational linguistics with statistical, machine learning, and deep learning models to take unstructured, free-form data (e.g., a voice recording or writing sample) and produce structured, quantitative acoustic and linguistic outputs. Nevertheless, using NLP in LLD will have its unique challenges in comprehending the identification of speech alterations, seeing older adults as it is more possible for them to have additional medical comorbid conditions that may affect their way of speech. To conclude, NLP can recognize the speech changes for depressed individuals (DeSouza et al., 2021).

Paper3: Natural language processing for structuring clinical text data on depression using UK-CRIS

This study demonstrates how Natural Language Processing (NLP) can elevate the medical field in diagnosing individuals by developing an algorithm that extracts clinical information from unstructured Electronic Health Records (EHRs) of patients with depression. The goal of the study is to support the Clinical Record Interactive Search (CRIS), which provides real-world data. The authors describe their methodology in two steps as follows: 1) Clinical experts identify the relevant information and then develop the training and testing datasets used for building statistical models. 2) They developed and refined the statistical models through active learning procedures to iteratively enhance performance while minimizing the annotation workload. This study demonstrates the success of the NLP models, which outperform human annotators in writing prescriptions. However, models can fail to achieve satisfactory results on more complex medical entities. Therefore, they introduce the model with a new large dataset into the training to adapt the model to a new pattern of textual data. Nevertheless, adding more information disrupted the weights of the categories, so the model underperformed at recalling subsequent information. In summary, the study shows how well-developed NLP models are in extracting complex medical information; the records show high accuracy results regarding drug-related information, but do not give the same accuracy on the auxiliary variables (Vaci et al., 2020).

Paper4: Predicting Depression Symptoms in an Arabic Psychological Forum

This study presents Natural Language Processing (NLP) and a machine learning approach to predict depression symptoms from Arabic text posted online. This study has been done before on English and Japanese, but not in Arabic. The authors' goal was to address the lack of research in the Arabic language in mental health prediction. Therefore, they developed both a lexicon-based approach and a machine learning-based approach, where they collected 20,000 posts from "Nafsany," which is an online platform that covers substantive subjects regarding diverse medical circumstances. Thus, they labeled the collected data into categories and further annotated them by psychologists using the "Diagnostic and Statistical Manual of Mental Disorders" (DSM-5) and "Patient Health Questionnaire" (PHQ-9). For the lexicon-based approach, they created the ArabDep Lexicon to collect depression-related terms using category-based and annotation-based methods. As for the machine-learning approach, six classifiers (AdaBoost, Decision Trees, KNN, Random Forest, SGD, and SVM) were used with many extraction methods such as n-grams and word embedding. Ultimately, they achieved promising performance with an accuracy that exceeds 80%, a recall of 82%, and a precision of 79% for the lexicon-based approach, and with an accuracy that exceeds 73%, a recall of 72%, and a precision of 74% for the machine-learning-based approach. Although the study shows reasonable results, further study is needed (Alghamdi et al., 2020).

Paper 5: Detection of Depression-Related Posts in Reddit Social Media Forum

The study showed in this paper presents a Natural Language model (NLP) and Machine learning (ML) approach to find depression words from “reddit” social media posts. They aimed to identify linguistic markers of depression and improve detection performance through proper feature selection and feature combinations. With data of 1,293 depression indicative posts and 548 standard posts from Reddit, which were labeled under the subreddit categories. The authors followed these three methodologies: 1) pre-processing the data, which includes removing URLs, punctuation, and stop words, and applying stemming to reduce words to their root forms. 2) extracted three types of features: N-grams (unigrams and bigrams using TF-IDF), LIWC (68 psycholinguistic categories), and LDA topic modeling (70 topics). 3) evaluated five classifiers: Logistic Regression, Support Vector Machine (SVM), Random Forest, AdaBoost, and Multilayer Perceptron (MLP). This resulted in 80% accuracy and 0.080 F1-score. However, the best overall performance was achieved by combining LIWC, LDA, and bigram features with the MLP classifier, reaching 91% accuracy and 0.93 F1-score. Finally, the authors state that the reasonable results are great for future work (Tadesse et al., 2019).

Paper 6: Deep learning for prediction of depressive symptoms in a large textual dataset

This research paper suggests a Deep Learning (DL) method to identify depression/depressive symptoms from textual data using Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM). Their aim is to develop an efficient system to identify texts with self-perceived symptoms of depression using symptom-based features rather than traditional word frequency techniques. 11,807 to 21,470 text entries classified as depression or non-depression by medical and psychological experts were gathered from a Norwegian public internet information channel (ung.no). The authors used these three approaches: 1) used a one-hot encoding process to extract strong features based on 189 unique phrases that reflect potential symptoms of depression as described by experts. 2) demonstrated superior clustering compared to conventional TF-IDF and one-hot techniques by using Linear Discriminant Analysis (LDA) to visualize and verify the robustness of the suggested features. 3) used the text data to train an LSTM-based RNN with an attention layer to simulate time-sequential emotional experiences. The authors concluded that their symptom-based feature approach with LSTM-RNN is highly effective for automatic depression detection in text after using Local Interpretable Model-Agnostic Explanations (LIME) to provide explainable AI decisions, demonstrating the significance of features in predicting depression (Uddin et al., 2022).

Critical Analysis:

The study on detecting depression using Natural Language Processing (NLP) models show a positive amount of well researched regarding this topic in mental health concerns. However, there is still significant limitations persist. Most research regard models as unexplainable black boxes, rely on tiny, language-specific datasets, and employ binary categorization, that it ignores the importance of addressing the severity of depression. And for most Lack research they lack of cross-cultural and cross-population validation impedes clinical usefulness. To close the gap between research, promise and practical mental health care, interpretability, severity assessment, and cross-cultural validation must be given top priority in future study.

Connections:

1. **Focus on NLP for depression detection:** Natural language processing (NLP) is a key tool in mental health research across various data sources and populations, as evidenced by the fact that all studies use NLP techniques to identify or predict depression.
2. **Diverse data sources:** The studies demonstrate that NLP may be used to both written and spoken language by using a variety of textual data sources, such as speech transcripts, electronic health records (UK-CRIS), and social media posts (Reddit, Arabic forums).
3. **Machine learning and lexicon-based approaches:** Researchers use lexicon-based techniques (LIWC, ArabDep Lexicon), deep learning models (CNNs, RNNs, LLMs), or machine learning classifiers (SVM, RF, MLP, SGD, etc.) to categorize depressive content throughout the studies, demonstrating a variety of methodological approaches.
4. **Performance evaluation:** To assess their models, all articles present quantitative performance indicators including accuracy, precision, recall, and F1-score. Depending on the dataset and methodology, reported accuracy ranges from 73% to 91%.

Conflicts between the findings:

1. **Symptom-based vs. word frequency features:** Some studies found that symptom-based features with LSTM-RNN achieved superior accuracy (Uddin et al., 2022), while others achieved strong results using traditional features like LIWC and n-grams (Tadesse et al., 2019), showing disagreement on the best feature approach.
2. **Deep learning vs. traditional machine learning:** LSTM-RNN outperformed traditional models in one study (Uddin et al., 2022), while others found lexicon-based approaches (Alghamdi et al., 2020) or MLP classifiers (Tadesse et al., 2019) to be more effective, indicating no consensus on the optimal model architecture.
3. **Explainability vs. performance:** Some studies incorporated explainable AI to provide transparency in predictions (Uddin et al., 2022), while others focused solely on maximizing accuracy without explaining model decisions, reflecting a divide in research priorities.
4. **Clinical vs. social media data:** Clinical records offer professionally documented information (Vaci et al., 2020), while social media data provides larger volumes but relies on self-disclosure or inferred labels (Tadesse et al., 2019; Alghamdi et al., 2020), creating tension between data quality and accessibility.
5. **Cross-cultural applicability:** Most models are developed for specific languages without validation across populations, conflicting with the recognized need for culturally sensitive and multilingual solutions (Teferra et al., 2024).

Synthesis Matrix for Literature Review:

Study	Authors	Year	Research Design	Key Findings/Results	Methodology	Limitations
Screening for Depression Using Natural Language Processing: Literature Review	Bazen Gashaw Teferra , BSc, MSc, PhD; Alice Rueda , PhD; Hilary Pang , MSc, MD; Richard Valenzano , PhD; Reza Samavi , PhD; Sridhar Krishnan , PhD; Venkat Bhat , MSc, MD	2024	Literature Review	The result shows a successful result to the NLP techniques such as sentiment analysis, linguistic markers, word embeddings, and deep learning models (CNNs, RNNs, and LLMs). Also, Machine learning models showed high accuracy in classifying depressive states.	The authors described their methods by employing terms like "depression screening," "depression detection," and "natural language processing" to search Semantic Scholar, PubMed, and Google Scholar. To find recurring themes and gaps in the literature, data were retrieved and synthesized.	The effectiveness of NLP techniques was not statistically evaluated because the literature search and review depended on qualitative synthesis without doing a meta-analysis. Furthermore, ethical and cross-cultural issues were raised but not thoroughly examined in all of the investigations , which may have resulted in gaps in generalizability and fairness.
Natural Language Processing as an Emerging Tool to Detect Late-Life Depression	Danielle D. DeSouza , Jessica Robin , Melisa Gumus , Anthony Yeung	2021	Experimental	Depressed people exhibit changes in speech characteristics such fundamental frequency (F0), speech pace, and pause duration. LLD	This review is narrative in nature. The authors used keywords associated with "late-life depression," "speech,"	The review may not have included all pertinent studies because it is narrative rather than systematic. It is challenging to distinguish

				categorization accuracy varied from 77 to 86%, while vocal measurements predicted depression levels in older persons with 86–92% accuracy. Jitter, shimmer, and MFCCs are examples of acoustic characteristics that successfully separated depressive patients from controls. However, speech analysis in elderly persons may be complicated by comorbidities such as dementia and drug side effects.	and "natural language processing" to search PubMed, Web of Science, and Google Scholar. They reviewed earlier studies on the use of linguistic and acoustic speech analysis to identify late-life depression (LLD), contrasted the results with those of younger persons and other clinical illnesses, and discussed potential future initiatives. There was no adherence to a systematic review process.	depression-specific alterations in older persons, as they frequently have medical comorbidities and take various medications that impact speech. Concerns about misdiagnosis are raised by the overlap between speech abnormalities and dementia and Parkinson's disease. The authors note that more studies with larger sample sizes and closely matched control groups are required.
Natural language processing for structuring clinical text data on depression	Nemanja Vaci , Qiang Liu, Andrey Kormilitzin,1 Franco De Crescenzo,1,2 Ayse Kurtulmus,1,3 Jade Harvey,2	2020	Experimental	The extraction of drug-related information is highly accurate, according to the results. On the other hand, with regard to auxiliary	The started with collecting data of a 13,000 patient records (1.8 million documents)	Even with 526 annotated papers, the model did not perform well on complicated medical

using UK-CRIS	Bessie O'Dell,1 Simeon Innocent,1 Anneka Tomlinson,1 Andrea Cipriani ,1,2 Alejo Nevado-Holgado1,4,			variables, a significantly lower level of accuracy is <u>shown</u> . The model's performance significantly improves when combined with cutting-edge active learning methods.	from the UK-CRIS database. they developed an NLP pipeline using GloVe embedding and bi-directional neural network. For the fine-tuning used an active learning tool to rapidly annotate a vast amount of information to the NER model.	entities, such as symptoms and life events. While adding more data during active learning, the weight of existing data was disturbed which led to decreasing the accuracy of certain variables.
Predicting Depression Symptoms in an Arabic Psychological Forum	Norah Saleh Alghamdi; Hanan A. Hosni Mahmoud; Ajith Abraham; Samar Awadh Alanazi; Laura García-Hernández	2023	Experimental	The lexicon-based approach achieved the best performance with 80.45% accuracy, 79.36% precision, 82.30% recall, and 80.81% F1-score using the Annotated ArabDep Lexicon. It performed best on strong depression cases (95% accuracy). The machine-learning approach reached 73%	The authors started with collecting post from an online platform. Then, they labeled them into forum categories where they had psychologists into manually annotate posts using DSM-5 and PHQ-9 guidelines. As for lexicon-based	limitations appear to be: 1) The dataset size was relatively small (20,000 posts), which limited the performance of deep learning models such as word embeddings and GloVe; 2) The study treated depression as a single binary class rather than distinguishing between different

				accuracy with SGD and TF-IDF. Word embeddings performed poorly due to small dataset size. Overall, the lexicon-based approach outperformed both the machine-learning approach and previous Arabic studies	approach, they created the ArabDep Lexicon using category-based and annotation-based methods. They applied N-gram modeling with unigrams (single words), bigrams (word pairs), and trigrams (three-word groups) to extract features. A rule-based algorithm then predicted depression by counting depression-related words against a threshold. For the machine-learning approach, they tested six classifiers (AdaBoost, DT, KNN, RF, SGD, SVM) with features like n-grams,	types or severity levels (e.g., seasonal affective disorder, bipolar disorder, postpartum depression); and 3) The lexicon-based approach showed decreased accuracy when the psychologist's diagnostic certainty was lower, indicating sensitivity to ambiguous cases
--	--	--	--	--	---	---

					TF-IDF, and word embeddings .	
Detection of Depression-Related Posts in Reddit Social Media Forum	Michael M. Tadesse; Hongfei Lin; Bo Xu; Liang Yang	2023	Experimental	The MLP model achieved a very excellent results by hitting 91% accuracy and 0.93 F1-score, beating SVM (90%, 0.91), LR (89%, 0.89), RF (85%, 0.85), and AdaBoost (79%, 0.81). while also performing better than the combined feature set that included unigrams, which got 83% accuracy and 0.84 F1-score with RF as the top classifier. According to these results, the key to better performance lies in choosing the right features and combining them effectively.	The authors incorporate a technical description of approaches applied for depression identification using NLP and text classification techniques. For data pre-processing, they used NLP tools to pre-process the dataset. After data pre-processing, they feed the models with the features that reflect users' language habits in Reddit forums. N-gram modeling is used to examine the features from the posts to calculate the probability of co-occurrence of each	1)The data collect is relatively small, with only 1,293 depression-indicative posts and 548 standard posts. this limited sample size which may lead to overfitting. 2) The study only performed binary classification without getting into the verity of depression nuanced.

					input sentence as a unigram and bigram. Also, use the term frequency-inverse document frequency (TF-IDF). And For text classification techniques, the proposed framework is developed by using Logistic Regression, Support Vector Machine, Random Forest, Adaptive Boosting, and Multilayer Perceptron classifiers.	
Deep learning for prediction of depressive symptoms in a large textual dataset	Md Zia Uddin, Kim Kristoffer Dysthe, Asbjørn Følstad & Petter Bae Brandtzaeg	2021	Experimental	The LSTM-RNN using symptom-based features performed better than traditional methods on two datasets, reaching a high of 91% with 98% and 99% accuracy. LDA visualization	From the Norwegian public information website <u>ung</u> , the writers gathered 11,807 and 21,470 text posts. <u>No</u> , <u>which</u> medical and psychologic	The dataset may not be as generalizable to other languages and cultures because it was gathered from a single Norwegian website. The method would need to be

				showed better feature clustering, and LIME generated explainable predictions.	al professional <u>s classified</u> <u>as</u> either depression or non-depression. They created binary features using one-hot encoding after extracting 189 distinct words associated with depression symptoms. Then, in order to simulate time-sequential emotional experiences, they trained an LSTM-based Recurrent Neural Network (RNN) with an attention layer. Tenfold and fivefold cross-validation were used to assess the model.	modified for other languages because the attributes were based on Norwegian language expressions of depression. The study did not evaluate the severity of depression; instead, it focused solely on binary classification (depression vs. non-depression). Furthermore, the paradigm requires a predetermine d list of symptom phrases, which might not encompass all potential manifestation s of depression.
--	--	--	--	--	--	--

References:

- Teferra BG, Rueda A, Pang H, Valenzano R, Samavi R, Krishnan S, Bhat V
Screening for Depression Using Natural Language Processing: Literature Review
Interact J Med Res 2024;13:e55067. <https://www.i-jmr.org/2024/1/e55067/>
- DeSouza DD, Robin J, Gumus M and Yeung A (2021) Natural Language Processing as an Emerging Tool to Detect Late-Life Depression. *Front. Psychiatry* 12:719125. doi: 10.3389/fpsyt.2021.719125.
<https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2021.719125/full>
- Vaci N, Liu Q, Kormilitzin A, De Crescenzo F, Kurtulmus A, Harvey J, et al. Natural language processing for structuring clinical text data on depression using UK-CRIS. *Evidence Based Mental Health*. 2020;23:.
<https://doi.org/10.1136/ebmental-2019-300134>
- N. S. Alghamdi, H. A. Hosni Mahmoud, A. Abraham, S. A. Alanazi and L. García-Hernández, "Predicting Depression Symptoms in an Arabic Psychological Forum," in *IEEE Access*, vol. 8, pp. 57317-57334, 2020, doi: 10.1109/ACCESS.2020.2981834. keywords: {Depression;Feature extraction;Natural language processing;Social networking (online);Mental health;Mental disorders;Supervised learning;semi-supervised learning;machine learning;predictive models;depression;lexicon;text analysis}
<https://ieeexplore.ieee.org/document/9040556/authors#authors>
- M. M. Tadesse, H. Lin, B. Xu and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," in *IEEE Access*, vol. 7, pp. 44883-44893, 2019, doi: 10.1109/ACCESS.2019.2909180. keywords: {Feature extraction;Task analysis;Linguistics;Twitter;Natural language processing;Support vector machines;Natural language processing;machine learning;Reddit;social networks;depression},
<https://ieeexplore.ieee.org/document/8681445/authors#authors>
- Uddin, M.Z., Dysthe, K.K., Følstad, A. *et al.* Deep learning for prediction of depressive symptoms in a large textual dataset. *Neural Comput & Applic* **34**, 721–744 (2022). <https://doi.org/10.1007/s00521-021-06426-4>