Literature Review in CLPsych

Heyuan Huang
Yale University
heyuan.huang@yale.edu

Abstract

This document is the collection of papers and their used datasets in the intersection of computer science and clinical psychology, published in 2022, and 2023. We searched the keywords 'mental', 'depression', and 'suicide' in the 4 main NLP conferences' proceedings to find these papers. Conferences include ACL 2023, EACL 2023, NAACL 2022, EMNLP 2022. This document can be used by students to find available datasets and quickly review recent research trends. [This is a draft version and will be revised later]

1 Dataset

1.1 Mental Disorder Detection

Social Media RSDD (Yates et al., 2017), Self-Reported Mental Health Diagnoses (SMHD) (Cohan et al., 2018), RSDD-Time and Mental Health Summarization (MentSum)(Sotudeh et al., 2022) datasets contain publicly available Reddit posts. SMHD contains 9 mental health conditions' (ADHD, bipolar, depression, anxiety, etc) diagnosed users' posts and control groups. Data request form. More resources link

Dreaddit dataset (Turcan and McKeown, 2019) is a publicly available dataset, containing 190K posts from 5 different domains (abuse, anxiety, financial, PTSD, social) of Reddit for stress identification (binary classification, stress or not). Download link

Hugging Face's cleaned Reddit dataset for depression: binary classification, is depression or not. Download link

ShreyaR depression detection dataset: binary classification, is depression or not. Download link

Jsfactory mental health Reddit posts: binary classification, 1 means diagnosed positive, and 0 means negative. Download link

CLPsych shared tasks have used various datasets with strict access restrictions. For the CLPsych2021 shared task, there are several public datasets using Twitter data for practice. Depression Detection Using Twitter Data is a binary classification dataset, consisting of depressive and non-depressive tweets evenly. All tweets have depressive hashtags so models trained on this dataset are more sensitive to the content rather than the tags. CLPsych2021 practice dataset also consists of train and test sets with twitter data.

Reddit Mental Health Dataset (RedditMH)(Low et al., 2020) contains posts from 15 mental health (depression, ADHD, suicidewatch, etc) and 11 non-mental health subreddits. Download link

Solomonk's Reddit Mental Health (SMK) contains 151K posts from 5 mental health subreddits (r/adhd, Asperger, depression, OCD, PTSD). Download link

Clinical Session Counseling and Psychotherapy Transcripts contains more than 2,000 anonymized clinical therapy session transcripts and 44,000 pages of client narratives, with class labels such as anxiety, depression, and schizophrenia. It can be used for classification tasks, for defining the patient-therapist relationship through their conversations, and for tracking the progress and setbacks of patients over multiple therapy sessions to develop a better treatment plan strategy. Data is available through Stanford Library membership (note: request by yale.edu account failed, requires Stanford University NetID).

Symptom Identification PsySym(Zhang et al., 2022) is annotated with 38 symptoms in DSM-5 for 7 mental diseases and has 1 status variable to show the certainty about symptom presence or not (sarcasm, metaphor, etc.)

Depression to Symptoms (D2S) dataset (Yadav et al., 2020) contains tweets with depression binary labels and 9 symptom labels in the PHQ-9.

1.2 Risk Prediction

Shared Tasks For CLPsych2019, the University of Maryland Reddit Suicidality Dataset Version 2(Zirikly et al., 2019; Shing et al., 2018) contains 11,129 users' longitudinal data posted on r/SuicideWatch and 11,129 control group users' posts. For suicide risk level (no risk, low, moderate, and severe risk) annotation, 621 users' data have crowdsourced annotation and 245 users' data are annotated by experts. It is of high quality but requires IRB exempt approval documentation for data access. Data Access Request Procedure.

eRisk tasks: Early risk prediction on the Internet. Organizers provide datasets and tasks every year, from 2017-2023 (eRisk Website). To use their text data collection for eRisk2023 shared tasks, signing the user agreement is enough. Its task 1, search for symptoms of depression, consists of ranking sentences from a collection of user writings according to their relevance to each depression symptom from 21 symptoms in the BDI Questionnaire. A sentence will be deemed relevant to a BDI symptom when it conveys information about the user's state concerning the symptom.

1.3 Mental Health Counseling

Counselor-Client Session Li et al. (2023) created a Mandarin text dataset from an online mental support platform, annotating each utterance with professional counselors' intentions (supporting, challenging, others) and 13 strategies or clients' reactions (positive, negative, others) and 9 behaviors. It is not publicly available and requires email to lianqi@westlake.edu.cn for access. Github link

1.4 Other Dataset

Garg et al. (2023) constructed a Reddit post dataset annotated with the binary presence (0,1) of Interpersonal Risk Factors (IRF): Thwarted Belongingness (TBe) and Perceived Burdensomeness (PBu). Download link

Zanwar et al. (2023b) constructed a German dataset for mental health detection, SMHD-GER, following the same approach proposed in SMHD(Cohan et al., 2018). It can be used to test the model's transferability to other languages.

2 Related Work

2.1 Mental Disorder Detection

Mapping Clinical Questionnaires Song et al. (2023) sampled depression, bipolar, anxiety, BPD subreddit posts with control groups from Reddit (dataset unavailable) and tested their trained model's Transferability on RSDD (Yates et al., 2017) and eRisk2018 (Losada et al., 2018). They compared the captured semantic meanings from the text to symptom-related descriptions to leverage expert domain knowledge for transferability and interpretability.

Adjusting Training Paradigm To improve model classification performance, Aragon et al. (2023) adapted BERT first to social media language and then adapted it to the mental health domain datasets, incorporating a depression lexicon to guide masking process.

Adding Features Zanwar et al. (2023a) tried 3 information fusion methods (psycholinguistic feature fusion, model fusion and task fusion) to improve mental health detection performance. They used GoEmotions dataset(Demszky et al., 2020) and Kaggle MBTI dataset(Li et al., 2018) to fuse Emotion and Personality information for better performance on SMHD and Dreaddit datasets.

Shifting Task Chhikara et al. (2023) changed the text classification task to a seq2seq Question-Answering task by converting original text-label pairs into Q-A format input. For binary classification, the question prompt is "Based on the following post by a social media user, are they at risk of suffering from any serious mental illness? <post> (a) Yes (b) No." For multi-class classification, the question prompt contains choices like "(a) OCD (b) ADHD (c) Depression (d) Aspergers (e) PTSD".

Toleubay et al. (2023) applied Logical Neural Network to clinical session transcript classification and proposed different predicate pruning methods for scalability and higher performance.

2.2 Mental Health Counseling

Li et al. (2023) trained classifiers to predict counselors' intentions, strategies, and clients' reactions and behaviors, based on a pre-trained Chinese RoBERTa-large model with an auxiliary masked language modeling (MLM) task.

Saha et al. (2022) designed a conversational virtual assistant, comprised of one mental disorder classification BERT model and one conversational response generator based on sentiment, rouge-L and BLEU score driven reinforcement learning, conditioned on the classified mental disorder.

2.3 Mental Support Resource Recommendation

Dang et al. (2023) integrated Reddit post discourse embedding for recommendation system development to provide social media users with the correct mental health support groups (subreddit classification) they need.

2.4 Suicide Risk Detection

Izmaylov et al. (2023) integrated the Suicide-Risk Factors (SRF) Lexicon in Hebrew in the pre-training task of a hierarchical BERT model for binary suicide risk prediction. The dataset is Sahar corpus, containing 40,000 sessions between users and counselors, with either positive or negative suicide risk labels. In the first layer, They use Aleph-BERT to encode each sentence in the session to a vector and concatenate a role embedding (user or counselor). In the second layer, a Context Encoder Transformer takes all sentence vectors as input and then the CLS classification head outputs the suicide risk prediction. The SRF lexicon contains 25 categories, such as "perceived burdensomeness", and "explicit suicide mentions". In the pre-training phase, they represent each session as a 25-dimension vector, with each dimension counting the number of sentences belonging to that category. Also, after using XGBoost feature selection, the top 5-dimension representation outperforms the 25-dimension vector. To integrate the lexicon knowledge, they designed **a new pre-training task**, Self Supervised Knowledge (SSK) by masking a sentence in a session with a probability of 80% and then letting the model predict the session's representation in the SRF space. The loss is the Mean Squared Error between the session's true representation and the predicted representation. The other 3 pre-training tasks are Next Utterance Generation, Masked Utterance Regression, Distributed Order Ranking Network, and the overall loss is the weighted sum of the 4 loss functions.

References

Mario Aragon, Adrian Pastor Lopez Monroy, Luis Gonzalez, David E. Losada, and Manuel Montes. 2023. DisorBERT: A double domain adaptation model for detecting signs of mental disorders in social media. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15305–15318, Toronto, Canada. Association for Computational Linguistics.

Prateek Chhikara, Ujjwal Pasupulety, John Marshall, Dhiraj Chaurasia, and Shweta Kumari. 2023. Privacy aware question-answering system for online mental health risk assessment. In *The 22nd Workshop on Biomedical Natural Language Processing and BioNLP Shared Tasks*, pages 215–222, Toronto, Canada. Association for Computational Linguistics.

Arman Cohan, Bart Desmet, Andrew Yates, Luca Soldaini, Sean MacAvaney, and Nazli Goharian. 2018. SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1485–1497, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Hy Dang, Bang Nguyen, Noah Ziems, and Meng Jiang. 2023. Embedding mental health discourse for community recommendation. In *Proceedings of the 4th Workshop on Computational Approaches to Discourse (CODI 2023)*, pages 163–172, Toronto, Canada. Association for Computational Linguistics.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Muskan Garg, Amirmohammad Shahbandegan, Amrit Chadha, and Vijay Mago. 2023. An annotated dataset for explainable interpersonal risk factors of mental disturbance in social media posts. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 11960–11969, Toronto, Canada. Association for Computational Linguistics.

Daniel Izmaylov, Avi Segal, Kobi Gal, Meytal Grimland, and Yossi Levi-Belz. 2023. Combining psychological theory with language models for suicide risk detection. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2430–2438, Dubrovnik, Croatia. Association for Computational Linguistics.

Anqi Li, Lizhi Ma, Yaling Mei, Hongliang He, Shuai Zhang, Huachuan Qiu, and Zhenzhong Lan. 2023. Understanding client reactions in online mental health counseling. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10358–10376, Toronto, Canada. Association for Computational Linguistics.

Charles Li, Monte Hancock, Ben Bowles, Olivia Hancock, Lesley Perg, Payton Brown, Asher Burrell, Gianella Frank, Frankie Stiers, Shana Marshall, Gale Mercado, Alexis-Walid Ahmed, Phillip Beckelheimer, Samuel Williamson, and Rodney Wade. 2018. Feature extraction from social media posts for psychometric typing of participants. In *Augmented*

Cognition: Intelligent Technologies: 12th International Conference, AC 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I, page 267–286, Berlin, Heidelberg. Springer-Verlag.

David E. Losada, Fabio A. Crestani, and Javier Parapar. 2018. Overview of erisk: Early risk prediction on the internet (extended lab overview). In *Conference and Labs of the Evaluation Forum*.

Daniel M Low, Laurie Rumker, John Torous, Guillermo Cecchi, Satrajit S Ghosh, and Tanya Talkar. 2020. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. *Journal of medical Internet research*, 22(10):e22635.

Tulika Saha, Saichethan Reddy, Anindya Das, Sriparna Saha, and Pushpak Bhattacharyya. 2022. A shoulder to cry on: Towards a motivational virtual assistant for assuaging mental agony. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2436–2449, Seattle, United States. Association for Computational Linguistics.

Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36.

Hoyun Song, Jisu Shin, Huije Lee, and Jong Park. 2023. A simple and flexible modeling for mental disorder detection by learning from clinical questionnaires. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12190–12206, Toronto, Canada. Association for Computational Linguistics.

Sajad Sotudeh, Nazli Goharian, and Zachary Young. 2022. MentSum: A resource for exploring summarization of mental health online posts. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 2682–2692, Marseille, France. European Language Resources Association.

Yeldar Toleubay, Don Joven Agravante, Daiki Kimura, Baihan Lin, Djallel Bouneffouf, and Michiaki Tatsubori. 2023. Utterance classification with logical neural network: Explainable AI for mental disorder diagnosis. In *Proceedings of the 5th Clinical Natural Language Processing Workshop*, pages 439–446, Toronto, Canada. Association for Computational Linguistics.

Elsbeth Turcan and Kathy McKeown. 2019. Dreaddit: A Reddit dataset for stress analysis in social media. In *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, pages 97–107, Hong Kong. Association for Computational Linguistics.

Shweta Yadav, Jainish Chauhan, Joy Prakash Sain, Krishnaprasad Thirunarayan, Amit Sheth, and Jeremiah Schumm. 2020. Identifying depressive symptoms from tweets: Figurative language enabled multitask learning framework. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 696–709, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Andrew Yates, Arman Cohan, and Nazli Goharian. 2017. Depression and self-harm risk assessment in online forums. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2968–2978, Copenhagen, Denmark. Association for Computational Linguistics.

Sourabh Zanwar, Xiaofei Li, Daniel Wiechmann, Yu Qiao, and Elma Kerz. 2023a. What to fuse and how to fuse: Exploring emotion and personality fusion strategies for explainable mental disorder detection. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8926–8940, Toronto, Canada. Association for Computational Linguistics.

Sourabh Zanwar, Daniel Wiechmann, Yu Qiao, and Elma Kerz. 2023b. SMHD-GER: A large-scale benchmark dataset for automatic mental health detection from social media in German. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 1526–1541, Dubrovnik, Croatia. Association for Computational Linguistics.

Zhiling Zhang, Siyuan Chen, Mengyue Wu, and Kenny Zhu. 2022. Symptom identification for interpretable detection of multiple mental disorders on social media. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9970–9985, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Ayah Zirikly, Philip Resnik, Özlem Uzuner, and Kristy Hollingshead. 2019. CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*.

A Example Appendix

This is a section in the appendix.