

LREC-COLING 2024



MentalHelp: A Multi-Task Dataset for Mental Health in Social Media

Nishat Raihan

Sadiya Sayara Chowdhury Puspo

Shafkat Farabi

Ana-Maria Bucur

Tharindu Ranasinghe

Marcos Zampieri

Motivation

- Mental health disorders involve significant disturbances in cognition, emotion regulation, or behavior. These disturbances lead to distress or disability in social, occupational, or other important activities.

"More than 700 000 people die due to suicide every year. Suicide is the fourth leading cause of death in 15–29-year-olds." - WHO

- Social media is a valuable source for data on mental health, as users frequently discuss their mental and emotional states.
- The need for professional annotation limits the size of usable datasets, posing challenges in developing robust detection models.
- Introduction of a large-scale, semi-supervised dataset to enhance the modeling and detection of mental disorders.

Previous Work

- Transformer-based models effectively analyze social media for detecting mental disorders, including suicide risk and conditions like depression, self-harm.
- Specialized transformer models like Bio-BERT, Clinical-BERT, and Multitask-Clinical BERT have been developed for mental health tasks, pre-trained on biomedical and clinical notes.
- Models such as Mental-BERT, Disor-BERT, and Suicidal-BERT are pre-trained on social media data specific to the mental health domain, improving their relevance and accuracy for mental health applications
- Many mental health datasets have limited annotations and public availability due to the sensitive nature of the data and confidentiality issues.

Benchmark Dataset

Dataset	Reference	Category	Source	Instances
depsev	(Naseem et al., 2022)	Depression	Reddit	3,553
datd	(Owen et al., 2020)	Depression, Anxiety	Twitter	5,550
signdep	(Kayalvizhi and Thenmozhi, 2022)	Depression	Reddit	16,632
sdcnl	(Haque et al., 2021)	Suicide	Reddit	1,895
sid	(Mirza Ibtihaj et al., 2020)	Suicide	Twitter	1,385
dreaddit	(Turcan and McKeown, 2019)	Stress	Reddit	3,553
cds	(Hung Chia Yu, 2022)	Depression, PTSD, Anxiety	Reddit	18,006
swmh	(Ji et al., 2022a)	Depression, Anxiety, Bipolar, Suicide	Reddit	54,412
dr	(Pirina and Çöltekin, 2018)	Depression	Reddit, Blogs	5,984
sad	(Mauriello et al., 2021)	Stress	SMS	6,849
depTweet	(Yeow Zi Qin et al., 2022)	Depression	Twitter	46,020
merged	–	–	–	163,839

Table 1: Benchmark datasets used in MentalHelp.

Models

➤ Base Models

- BERT (devlin, 2017)
- RoBERTa (liu, 2019)

➤ Task Pretrained Models

- DAPT
 - Bio BERT (Lee, 2020)
 - Clinical BERT (Alsentzer, 2019)
 - Bio-Clinical BERT (Ling, 2022)
 - Disor BERT (Aragon, 2024)
- TAPT
 - Mental BERT (Ji, 2023)

➤ Task Finetuned Models

- Suicide BERT (Tanaka, 2022)

➤ Large Language Models

- Flan-T5 (Chung, 2022)
- GPT3.5 (Brown, 2023)
 - Zero Shot
 - Few Shot (5)

Curating MentalHelp.....(1)

Outline

- Data Collection from Reddit
- Binarizing the Benchmark datasets
- Merging the Benchmark datasets
- Model Evaluation on the Merged dataset
- Picking Top-3
 - Flan-T5
 - Disor BERT
 - Mental BERT
- Generating the Labels with Democratic Cotraining

Curating MentalHelp.....(2)

Model Evaluation on the Benchmark dataset

Model Name	depsev	datd	signdep	sdcnl	sid	dread	cds	swmh	dr	sad	depTweet	avg
flan-T5	0.80	0.91	0.90	0.61	0.86	0.74	0.69	0.80	0.90	0.94	0.94	0.83±0.10
Disor-BERT	0.81	0.94	0.83	0.69	0.80	0.86	0.62	0.81	0.94	0.95	0.93	0.83±0.10
Mental-BERT	0.82	0.94	0.83	0.73	0.84	0.80	0.65	0.78	0.93	0.95	0.86	0.83±0.09
RoBERTa	0.81	0.94	0.83	0.71	0.80	0.81	0.61	0.78	0.96	0.95	0.83	0.82±0.10
BERT	0.81	0.94	0.83	0.68	0.81	0.78	0.62	0.78	0.92	0.94	0.82	0.81±0.10
Suicidal-BERT	0.81	0.93	0.83	0.69	0.84	0.72	0.61	0.77	0.92	0.95	0.92	0.82±0.11
Clinical-BERT	0.77	0.95	0.80	0.61	0.76	0.72	0.60	0.77	0.90	0.92	0.78	0.78±0.11
Bio-Clinical-BERT	0.79	0.94	0.82	0.62	0.75	0.76	0.60	0.76	0.90	0.94	0.78	0.79±0.11
Bio-BERT	0.81	0.93	0.82	0.64	0.79	0.75	0.60	0.77	0.92	0.94	0.81	0.80±0.10
GPT3.5-turbo (Few-Shot)	0.69	0.78	0.84	0.52	0.68	0.79	0.54	0.78	0.89	0.92	0.89	0.76±0.13
GPT3.5-turbo (Zero-Shot)	0.63	0.74	0.82	0.52	0.61	0.71	0.52	0.79	0.87	0.79	0.83	0.71±0.12

Table 2: Weighted F_1 -score comparison of all the models on the benchmark datasets. The highest score(s) for each dataset are marked in bold.

Curating MentalHelp.....(3)

Model Evaluation on the Merged dataset

Model Name	Weighted F_1
flan-T5	0.95
Disor-BERT	0.92
Mental-BERT	0.91
RoBERTa	0.89
Suicidal-BERT	0.88
Clinical-BERT	0.85
GPT3.5-turbo (Few-Shot)	0.85
BERT	0.84
Bio-Clinical-BERT	0.84
Bio-BERT	0.83
GPT3.5-turbo (Zero-Shot)	0.82

Table 4: Weighted F_1 -score comparison on the merged dataset.

MentalHelp Dataset

A Snapshot

Instance	flan-T5	Disor-BERT	Mental-BERT	Average	Label
Anyone want to vent? Im a good listener... Just want to have someone to speak to online, nobody cares enough irl	0.010	0.001	0.001	$0.004_{\pm 0.005}$	NO
I think I died a long time ago. So what exactly am I doing here? It isn't living. I've been called lifeless and told I am wasting away, told I am wasteful, that I am more or less a zombie.	0.998	0.999	0.945	$0.981_{\pm 0.027}$	YES
When you cant just be yourself and youre so plagued by your damn mental illness.	0.999	0.999	0.995	$0.998_{\pm 0.002}$	YES
Didn think I would have live this long to see 2020.. Dont even know if this is considered an accomplishment.	0.044	0.001	0.995	$0.347_{\pm 0.561}$	NO
I want to disappear and stop being the bur- den I am to people. (I am not in danger)	0.997	0.003	0.882	$0.627_{\pm 0.543}$	YES

Table 3: Selected examples from the MentalHelp dataset. We present the confidence scores from flan-T5, Disor-BERT, and Mental-BERT. The confidence scores represent how confident a model is that a particular text depicts symptoms of mental disorders.

MentalHelp Dataset

Class Distribution

Average confidence	# Instances	YES (%)	NO (%)
>0.9 or <0.1	7,522,421	57.52	42.47
>0.8 or <0.2	8,406,396	56.64	43.35
>0.7 or <0.3	9,514,001	55.04	44.95
>0.6 or <0.4	13,095,547	50.02	49.97
>0.5 or <0.5	14,097,946	49.97	50.02

Table 5: Class distribution in MentalHelp across different average confidence intervals.

Conclusion

- The absence of a public large-scale dataset for detecting depression from social media text data led to the development of MentalHelp.
- MentalHelp can be refined to use only high-quality data for further use.
- Future Avenue: Future work involves fine-tuning models like BERT and RoBERTa on MentalHelp for competitive performance and expanding into multilingual datasets to detect mental health issues across various languages.



Thank You!