MentalBERT:
Publicly Available Pretrained Language Models for Mental Healthcare

Shaoxiong Ji[†], Tianlin Zhang[‡], Luna Ansari[†], Jie Fu[§], Prayag Tiwari[†], Erik Cambria[¶]

[†] Aalto University, Finland [‡] The University of Manchester, UK

[§] Mila, Québec AI Institute, Canada [¶] Nanyang Technological University, Singapore

{shaoxiong.ji; luna.ansari; prayag.tiwari}@aalto.fi

tianlin.zhang@postgrad.manchester.ac.uk

fujie@mila.quebec cambria@ntu.edu.sg

Introduction

- Mental health is a critical issue
- Mental disorders could sometimes turn to suicidal ideation
- Early detection of mental disorders and suicidal ideation from social content

- Natural language processing with deep learning
- Contextualized language models, e.g., BERT (Devlin et al., 2019)
- Domain-adaptive masked language models for mental health

Methods and Setup: Pretraining

- Standard pretraining protocols of BERT and RoBERTa
- Base network architecture
- Start form the checkpoints of original BERT and RoBERTa

- Set batch size to 16 per GPU, evaluate every 1,000 steps, and train for 624,000 iterations.
- Training with four GPUs takes around eight days

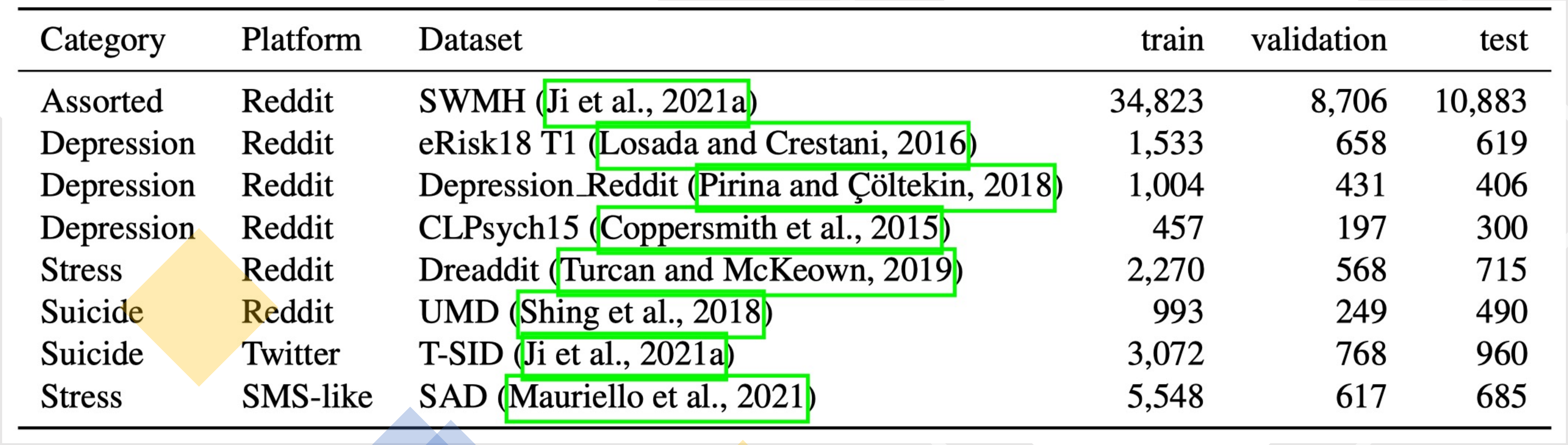
Methods and Setup: Corpus

- Reddit, an anonymous network of communities for discussion among people of similar interests
- Posts from mental health-related subreddits include:
  - ``r/depression’’
  - ``r/SuicideWatch’’
  - ``r/Anxiety’’
  - ``r/offmychest’’
  - ``r/bipolar’’
  - ``r/mentalillness/’’
  - ``r/mentalhealth''

Methods and Setup: Downstream Tasks

- binary and multi-class mental disorder classification
- mental disorder classification, e.g., stress, anxiety, depression
- suicidal ideation detection
- classification layers: MLP with the hyperbolic tangent activation function

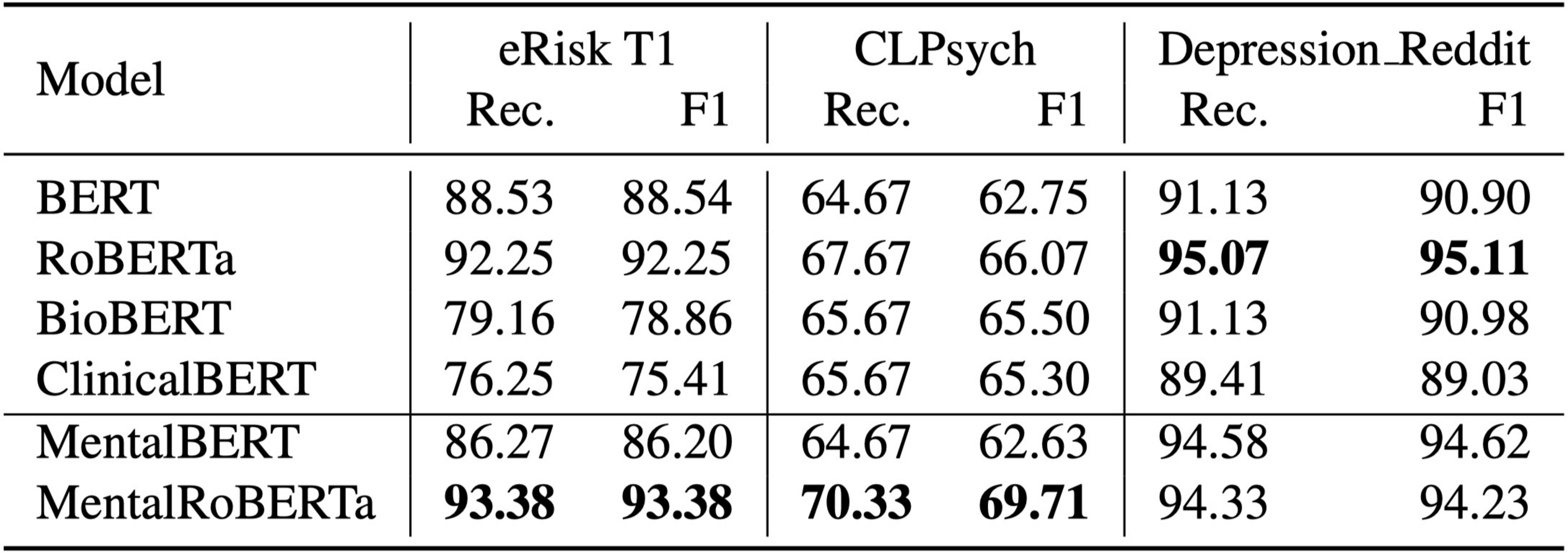
Datasets



Category	Platform	Dataset	train	validation	test
Assorted	Reddit	SWMH (Ji et al., 2021a)	34,823	8,706	10,883
Depression	Reddit	eRisk18 T1 (Losada and Crestani, 2016)	1,533	658	619
Depression	Reddit	Depression_Reddit (Pirina and Çöltekin, 2018)	1,004	431	406
Depression	Reddit	CLPsych15 (Coppersmith et al., 2015)	457	197	300
Stress	Reddit	Dreaddit (Turcan and McKeown, 2019)	2,270	568	715
Suicide	Reddit	UMD (Shing et al., 2018)	993	249	490
Suicide	Twitter	T-SID (Ji et al., 2021a)	3,072	768	960
Stress	SMS-like	SAD (Mauriello et al., 2021)	5,548	617	685

Results

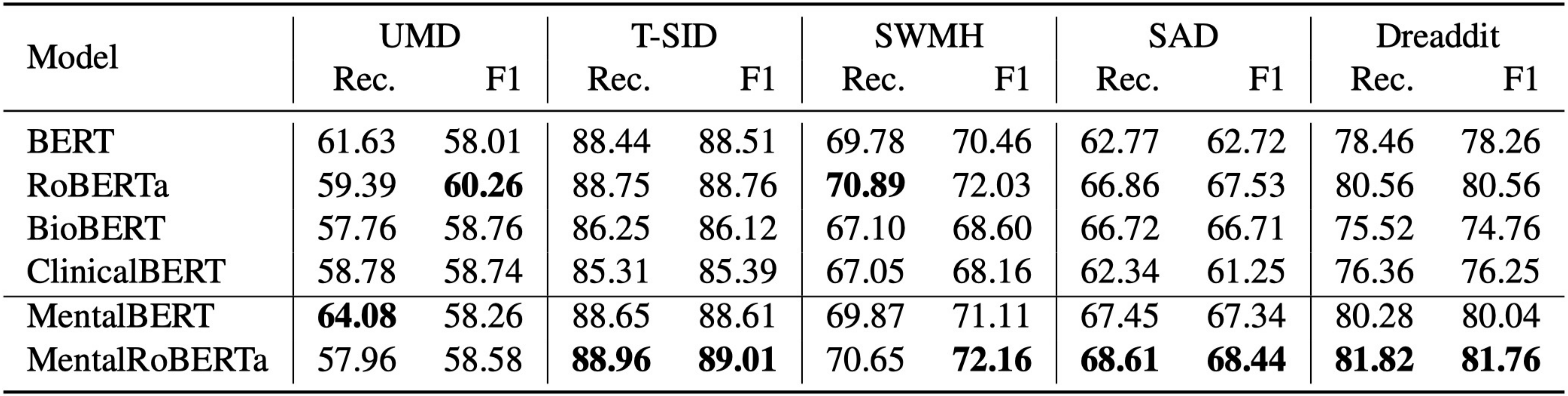
Results of depression detection



Model	eRisk T1		CLPsych		Depression_Reddit	
	Rec.	F1	Rec.	F1	Rec.	F1
BERT	88.53	88.54	64.67	62.75	91.13	90.90
RoBERTa	92.25	92.25	67.67	66.07	**95.07**	**95.11**
BioBERT	79.16	78.86	65.67	65.50	91.13	90.98
ClinicalBERT	76.25	75.41	65.67	65.30	89.41	89.03
MentalBERT	86.27	86.20	64.67	62.63	94.58	94.62
MentalRoBERTa	**93.38**	**93.38**	**70.33**	**69.71**	94.33	94.23

Results

Results of classifying other mental disorders including stress, anorexia, suicidal ideation.



Model	UMD		T-SID		SWMH		SAD		Dreaddit	
	Rec.	F1	Rec.	F1	Rec.	F1	Rec.	F1	Rec.	F1
BERT	61.63	58.01	88.44	88.51	69.78	70.46	62.77	62.72	78.46	78.26
RoBERTa	59.39	**60.26**	88.75	88.76	**70.89**	72.03	66.86	67.53	80.56	80.56
BioBERT	57.76	58.76	86.25	86.12	67.10	68.60	66.72	66.71	75.52	74.76
ClinicalBERT	58.78	58.74	85.31	85.39	67.05	68.16	62.34	61.25	76.36	76.25
MentalBERT	**64.08**	58.26	88.65	88.61	69.87	71.11	67.45	67.34	80.28	80.04
MentalRoBERTa	57.96	58.58	**88.96**	**89.01**	70.65	**72.16**	**68.61**	**68.44**	**81.82**	**81.76**

Conclusion

Two masked language models (https://huggingface.co/mental)

- Mental BERT
- Mental RoBERTa

Comprehensive evaluation on downstream classification

- depression
- stress
- suicidal ideation detection