Multi-Aspect Transfer Learning for Detecting Low Resource Mental Disorders on Social Media

Ana-Sabina Uban, Berta Chulvi, Paolo Rosso
University of Bucharest, Romania
Universitat Politècnica de València, Spain

LREC
June 2022

Mental health

Depression

Depression (major depressive disorder) is a common and serious medical illness that negatively affects how you feel, the way you think and how you act.

Depression causes feelings of sadness and/or a loss of interest in activities you once enjoyed. It can lead to a variety of emotional and physical problems and can decrease your ability to function at work and at home.

Source: *American Psychiatric Association website*

Mental health

Eating disorders

Eating disorders are illnesses in which the people experience severe disturbances in their eating behaviors and related **thoughts** and **emotions**. People with eating disorders typically become preoccupied with food and their body weight.

People with **anorexia** nervosa and bulimia nervosa tend to be **perfectionists** with **low self-esteem** and are **extremely critical** of themselves and their bodies.

Source: [American Psychiatric Association website]()

Mental health

PTSD

Post-traumatic stress disorder (PTSD) is a psychiatric disorder that may occur in people who have experienced or witnessed a traumatic event such as a natural disaster, a serious accident, a terrorist act, war/combat, or rape or who have been threatened with death, sexual violence or serious injury.

People with PTSD have **intense, disturbing thoughts and feelings** related to their experience that last long after the traumatic event has ended. They may relive the event through flashbacks or nightmares; they may feel **sadness**, **fear** or **anger**; and they may feel detached or estranged from other people.

Source: American Psychiatric Association website

Mental health

Suicide prevention

As the 10th leading cause of death in the United States and the **second leading cause of death** (after accidents) for people aged **10 to 34**, suicide is a serious public health problem.

Suicide is linked to **mental disorders**, particularly depression and alcohol use disorders.

Source: American Psychiatric Association website

Mental health disorders: Importance

- Affects quality of life (emotions, thoughts, activities, social)
- Affects physical health (sleep, eating, energy)
- Can lead to suicide
- COVID-19 pandemic affected mental health from multiple directions (health, social, economical, …)
- Social media engagement can further affect mental health
- Underdiagnosed, undertreated
  - Depression 50% diagnosed, 13–49% properly treated

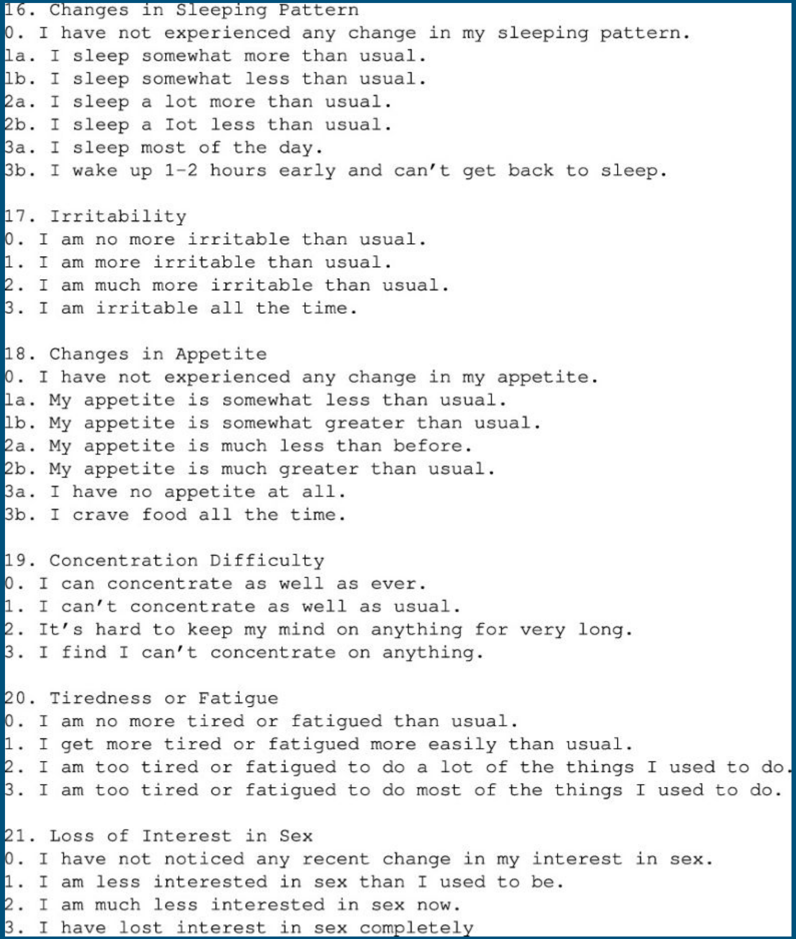
Mental disorders: automatic detection

Motivation and applicability

- **Alerting** users who show symptoms (recommend professional **help**); **suicide watch**, online counselling (chatbots) ...
- Preventing development of disorders (**early** detection)
- **Assisting clinicians** with new insights: building, developing diagnostic criteria (e.g. anorexia)
  - the diagnosis of certain disorders can also be a complicated issue, standards for diagnosis constantly evolving
  - evidence of co-morbidity between certain disorders

Data for mental disorders

- Medical records
- Questionnaires
- Therapy sessions



16. Changes in Sleeping Pattern

0. I have not experienced any change in my sleeping pattern.

1a. I sleep somewhat more than usual.

1b. I sleep somewhat less than usual.

2a. I sleep a lot more than usual.

2b. I sleep a Iot less than usual.

3a. I sleep most of the day.

3b. I wake up 1-2 hours early and can't get back to sleep.

17. Irritability

0. I am no more irritable than usual.

1. I am more irritable than usual.

2. I am much more irritable than usual.

3. I am irritable all the time.

18. Changes in Appetite

0. I have not experienced any change in my appetite.

1a. My appetite is somewhat less than usual.

1b. My appetite is somewhat greater than usual.

2a. My appetite is much less than before.

2b. My appetite is much greater than usual.

3a. I have no appetite at all.

3b. I crave food all the time.

19. Concentration Difficulty

0. I can concentrate as well as ever.

1. I can't concentrate as well as usual.

2. It's hard to keep my mind on anything for very long.

3. I find I can't concentrate on anything.

20. Tiredness or Fatigue

0. I am no more tired or fatigued than usual.

1. I get more tired or fatigued more easily than usual.

2. I am too tired or fatigued to do a lot of the things I used to do.

3. I am too tired or fatigued to do most of the things I used to do.

21. Loss of Interest in Sex

0. I have not noticed any recent change in my interest in sex.

1. I am less interested in sex than I used to be.

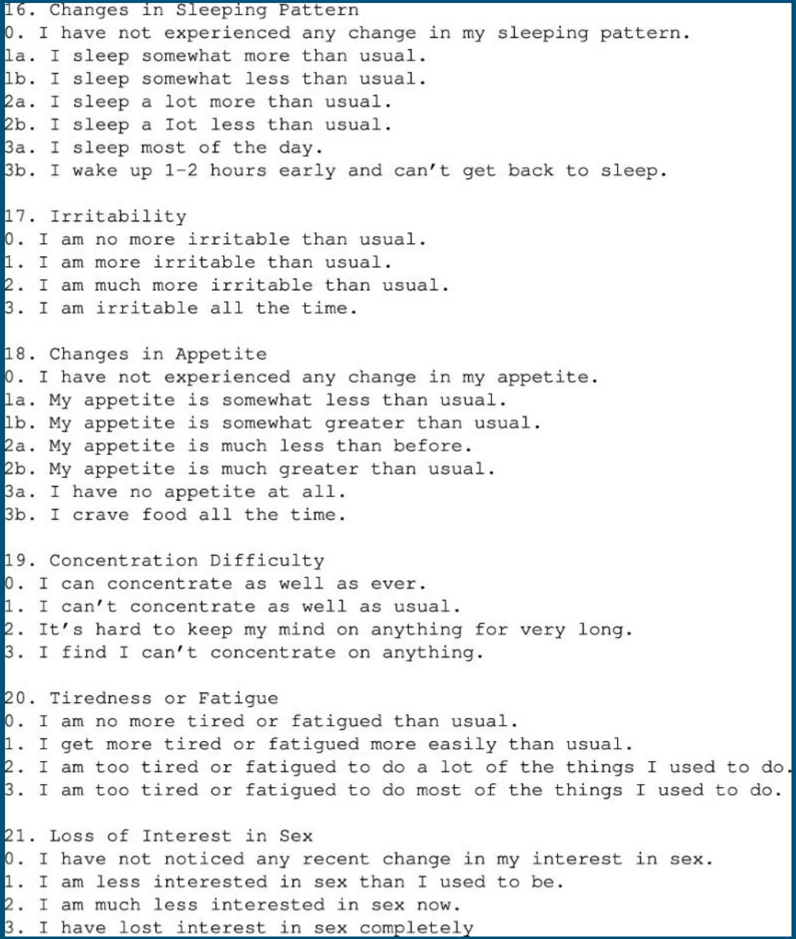
2. I am much less interested in sex now.

3. I have lost interest in sex completely

Data for mental disorders

- Medical records
- Questionnaires
- Therapy sessions

Costly to annotate



16. Changes in Sleeping Pattern
0. I have not experienced any change in my sleeping pattern.
1a. I sleep somewhat more than usual.
1b. I sleep somewhat less than usual.
2a. I sleep a lot more than usual.
2b. I sleep a Iot less than usual.
3a. I sleep most of the day.
3b. I wake up 1-2 hours early and can't get back to sleep.

17. Irritability
0. I am no more irritable than usual.
1. I am more irritable than usual.
2. I am much more irritable than usual.
3. I am irritable all the time.

18. Changes in Appetite
0. I have not experienced any change in my appetite.
1a. My appetite is somewhat less than usual.
1b. My appetite is somewhat greater than usual.
2a. My appetite is much less than before.
2b. My appetite is much greater than usual.
3a. I have no appetite at all.
3b. I crave food all the time.

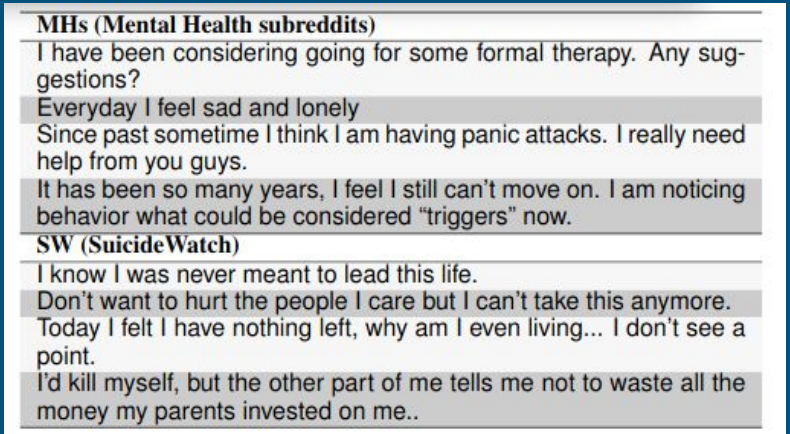
19. Concentration Difficulty
0. I can concentrate as well as ever.
1. I can't concentrate as well as usual.
2. It's hard to keep my mind on anything for very long.
3. I find I can't concentrate on anything.

20. Tiredness or Fatigue
0. I am no more tired or fatigued than usual.
1. I get more tired or fatigued more easily than usual.
2. I am too tired or fatigued to do a lot of the things I used to do.
3. I am too tired or fatigued to do most of the things I used to do.

21. Loss of Interest in Sex
0. I have not noticed any recent change in my interest in sex.
1. I am less interested in sex than I used to be.
2. I am much less interested in sex now.
3. I have lost interest in sex completely

Data for mental disorders

- Medical records
- Questionnaires
- Therapy sessions
- Social media

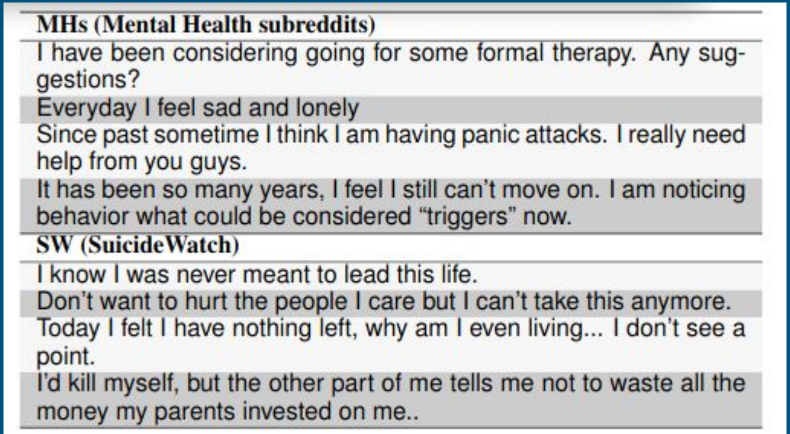


MHs (Mental Health subreddits)
I have been considering going for some formal therapy. Any suggestions?
Everyday I feel sad and lonely
Since past sometime I think I am having panic attacks. I really need help from you guys.
It has been so many years, I feel I still can't move on. I am noticing behavior what could be considered "triggers" now.
SW (SuicideWatch)
I know I was never meant to lead this life.
Don't want to hurt the people I care but I can't take this anymore.
Today I felt I have nothing left, why am I even living... I don't see a point.
I'd kill myself, but the other part of me tells me not to waste all the money my parents invested on me..

Table 1: Example titles of posts in the MHs and SW datasets; content has been carefully paraphrased to protect the privacy of the individuals.

Datasets for mental disorders

- Depression (mostly)
- Anorexia
- PTSD
- …



MHs (Mental Health subreddits)
I have been considering going for some formal therapy. Any suggestions?
Everyday I feel sad and lonely
Since past sometime I think I am having panic attacks. I really need help from you guys.
It has been so many years, I feel I still can't move on. I am noticing behavior what could be considered "triggers" now.
SW (SuicideWatch)
I know I was never meant to lead this life.
Don't want to hurt the people I care but I can't take this anymore.
Today I felt I have nothing left, why am I even living... I don't see a point.
I'd kill myself, but the other part of me tells me not to waste all the money my parents invested on me..

Table 1: Example titles of posts in the MHs and SW datasets; content has been carefully paraphrased to protect the privacy of the individuals.

Research questions

(RQ1) Can transfer learning be leveraged in order to improve the detection performance of automatic deep learning models for disorders where datasets are scarce, and be used across different social media platforms?

(RQ2) What can we learn about the similarity between the different disorders through studying the effectiveness of transfer learning?

(RQ3) How can we use interpretable multi-aspect deep learning models to reveal qualitative conclusions about the specific linguistic dimensions which are more similar across different disorders?

Experimental setup

Data: social media posts collected based on self-stated diagnoses

Text classification: **supervised binary classification** at **user level** (is a user depressed...?); cross-disorder classification (what is this user suffering from...?)

Deep learning model, hierarchical architecture (post-level attention + user-level attention); **features** from multiple **levels** of the text: content, style and emotion features

Transfer learning experiments:

- Cross disorders
- Cross platform
- Comparing strategies
- Analyzing errors and useful features

Datasets

Workshops and shared tasks on mental disorder detection

CLPsych: Computational Linguistics and Clinical Psychology (2014, 2015,...)

- ❖ Linguistic Twitter data to detect various mental disorders

eRisk: Early Risk Detection on Social Media (since 2017)

- ❖ Textual data from reddit forums: depression, anorexia, self-harm...

Datasets used:

- ❖ depression (CLPsych, eRisk, + additional Twitter depression dataset)
- ❖ self-harm (eRisk)
- ❖ anorexia (eRisk)
- ❖ PTSD (CLPsych)

Annotated based on self-stated diagnoses

Datasets statistics

Dataset	Users	Positive %	Posts	Words
eRisk self-harm (reddit)	763	19%	274,534	~ 6M
eRisk anorexia (reddit)	1287	10%	823,754	~ 23M
eRisk depression (reddit)	1304	16%	811,586	~ 25M
CLPsych depression (Twitter)	822	64%	1,919,353	~ 26M
CLPsych PTSD (Twitter)	1078	72%	2,541,214	~ 19M
Twitter depression dataset	519	50%	52,080	~500K

Classification experiments:
Features

Content:

❖ Word sequences + word embeddings (GloVe)

Style:

❖ Function words (as bag of words)

Emotion:

❖ NRC emotion lexicon (as proportion of each emotion in each post)

LIWC categories (topics, emotions, style) (as proportion of each category in each post)

Classification experiments
Features

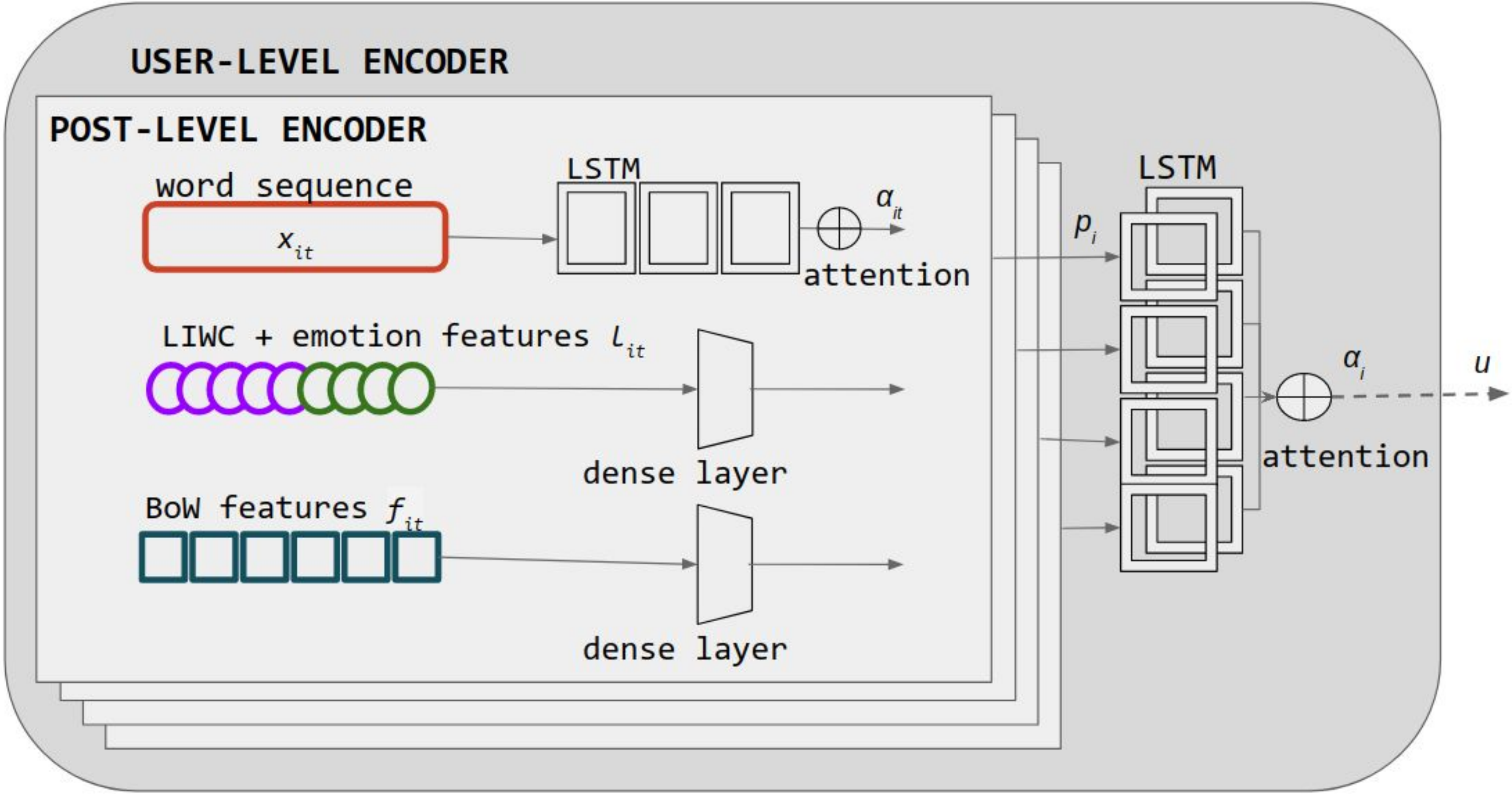
NRC emotions (Plutchik's 8 emotions + 2 sentiments):

anger, anticipation, disgust, fear, joy, sadness, surprise, trust; negative, positive

LIWC categories (64 categories):

- ➢ Sentiment polarity
- ➢ Emotions (*sadness, anxiety, affect*…)
- ➢ Syntactic categories (*pronouns, verbs, conjunctions*…)
- ➢ Topics (*health, money, religion, work*…)

Our solution: model architecture



Classification results: cross-disorder classification

Depression vs self-harm vs anorexia classification (Reddit): **0.44 F1**
Depression vs PTSD classification (Twitter): **0.72 F1**

Reddit

Predicted True	**Depr**	**Self-harm**	**Anorexia**
Depr	139	2	113
Self-harm	60	67	144
Anorexia	201	16	218

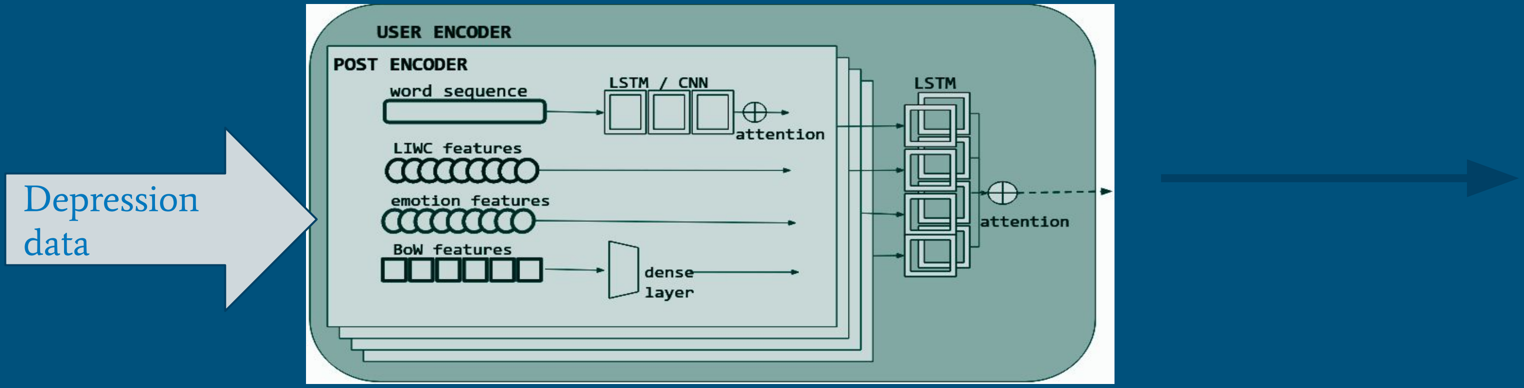
Twitter

Predicted True	**Depr**	**PTSD**
Depr	126	24
PTSD	65	95

Confusion matrices for classification between disorders

Transfer learning

Strategy 0. Zero-shot

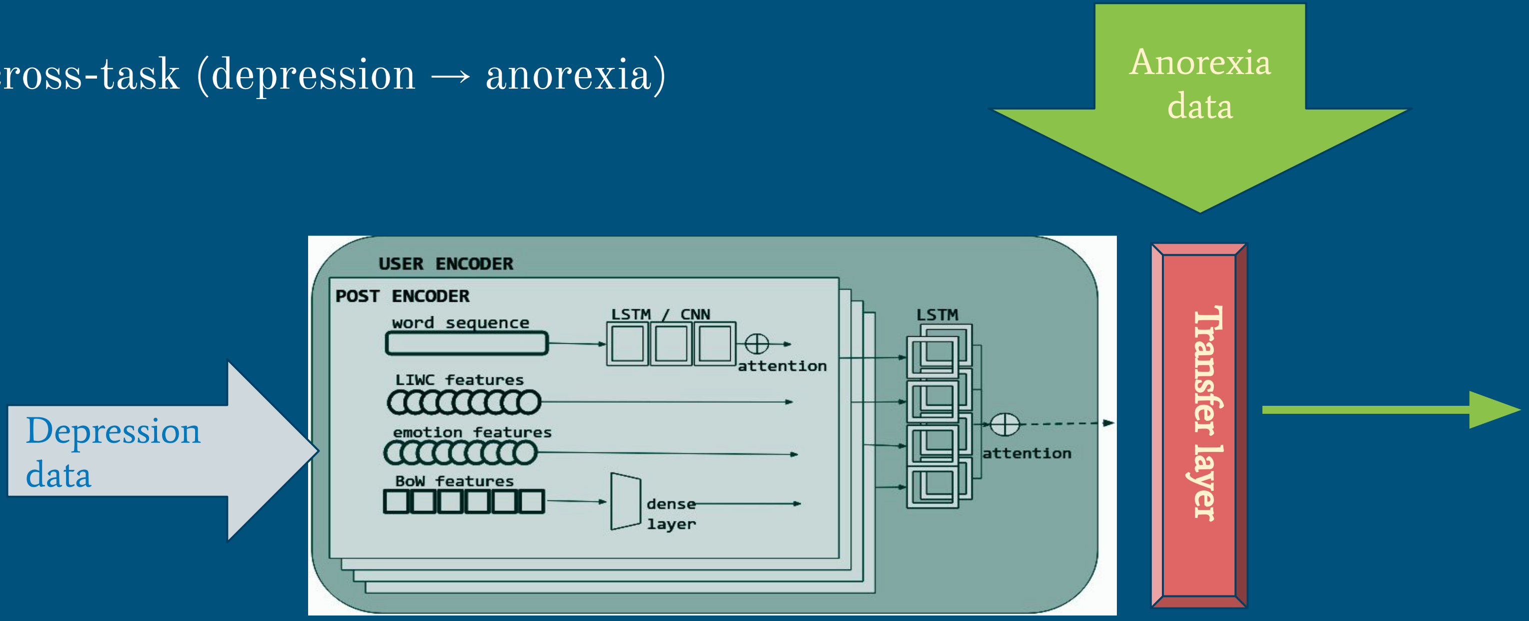


Depression data

Transfer learning

Strategy 1. Transfer layer

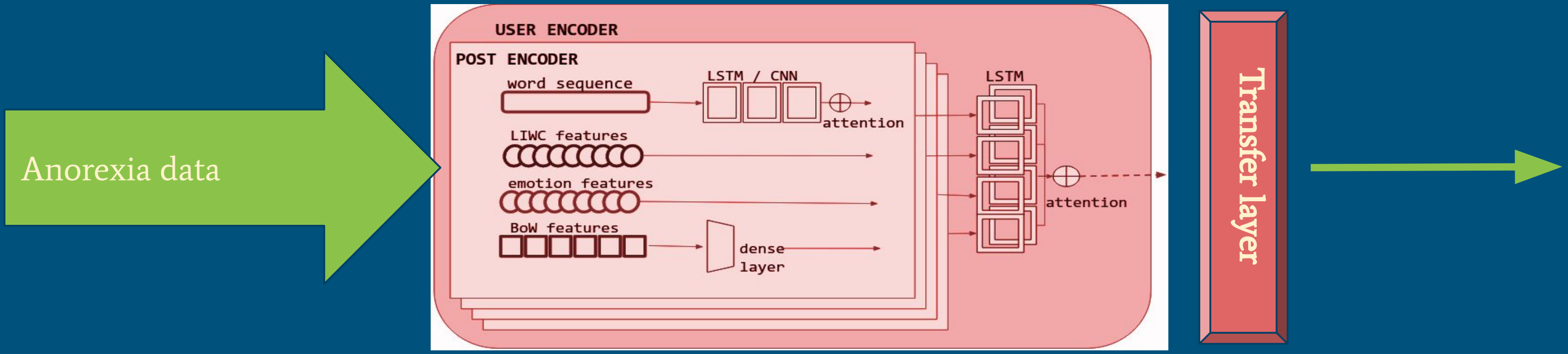
Example: cross-task (depression → anorexia)



Transfer learning

Strategy 2. Fine-tuning

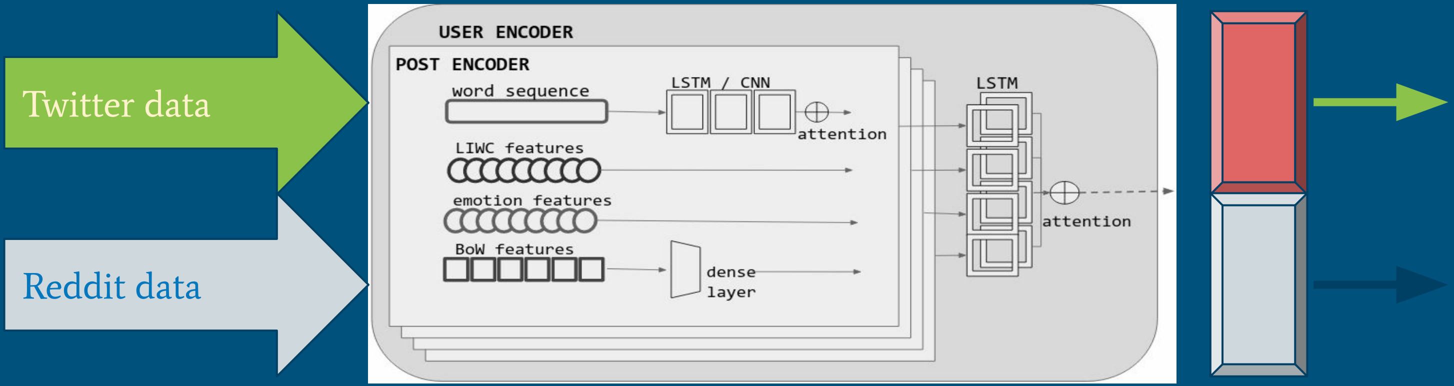
Example: cross-task (depression → anorexia)



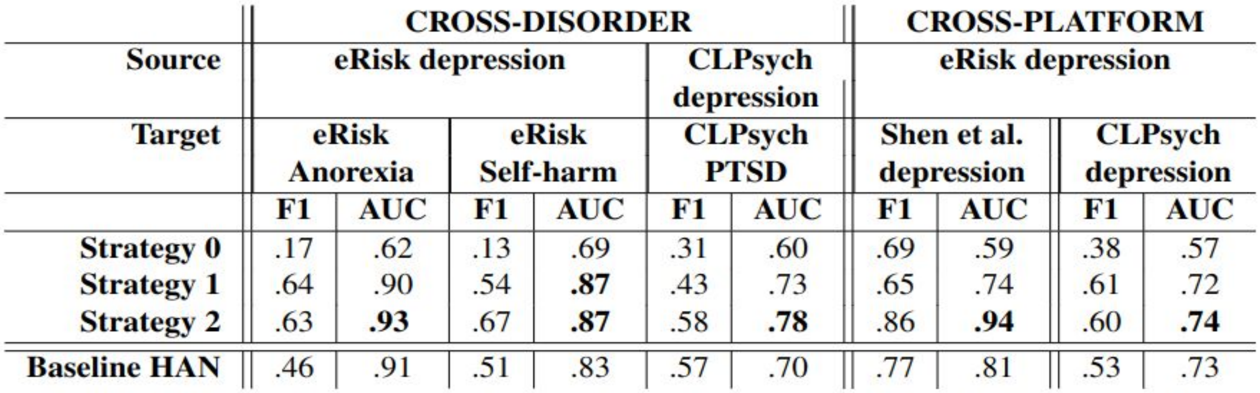
Transfer learning

Strategy 3. Multi-task learning

Example: cross-platform (reddit / Twitter)

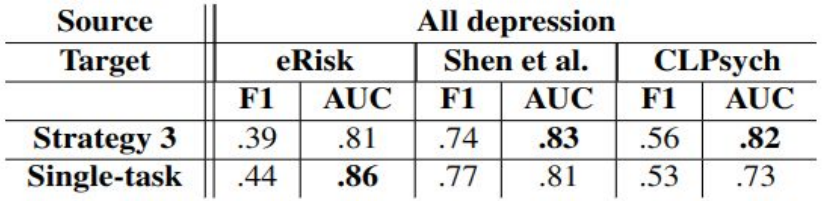


Transfer learning experiments. Results



	CROSS-DISORDER						CROSS-PLATFORM			
Source	**eRisk depression**				**CLPsych depression**		**eRisk depression**			
Target	**eRisk Anorexia**		**eRisk Self-harm**		**CLPsych PTSD**		**Shen et al. depression**		**CLPsych depression**	
	F1	**AUC**	**F1**	**AUC**	**F1**	**AUC**	**F1**	**AUC**	**F1**	**AUC**
Strategy 0	.17	.62	.13	.69	.31	.60	.69	.59	.38	.57
Strategy 1	.64	.90	.54	**.87**	.43	.73	.65	.74	.61	.72
Strategy 2	.63	**.93**	.67	**.87**	.58	**.78**	.86	**.94**	.60	**.74**
Baseline HAN	.46	.91	.51	.83	.57	.70	.77	.81	.53	.73

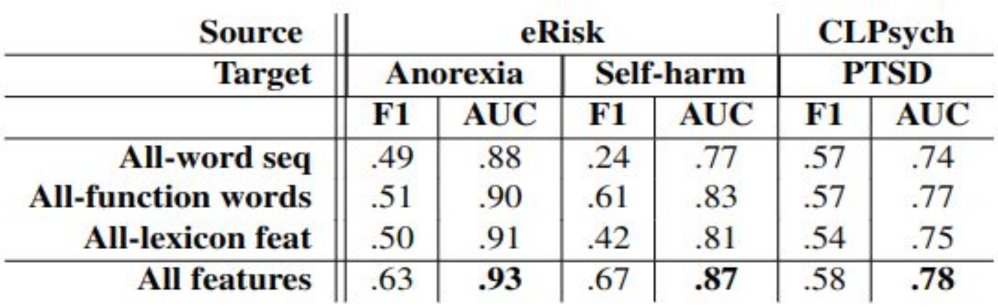
Cross-disorder and cross-platform transfer learning results, compared to individual disorder prediction



Source	**All depression**					
Target	**eRisk**		**Shen et al.**		**CLPsych**	
	F1	**AUC**	**F1**	**AUC**	**F1**	**AUC**
Strategy 3	.39	.81	.74	**.83**	.56	**.82**
Single-task	.44	**.86**	.77	.81	.53	.73

Cross-platform multi-task learning results

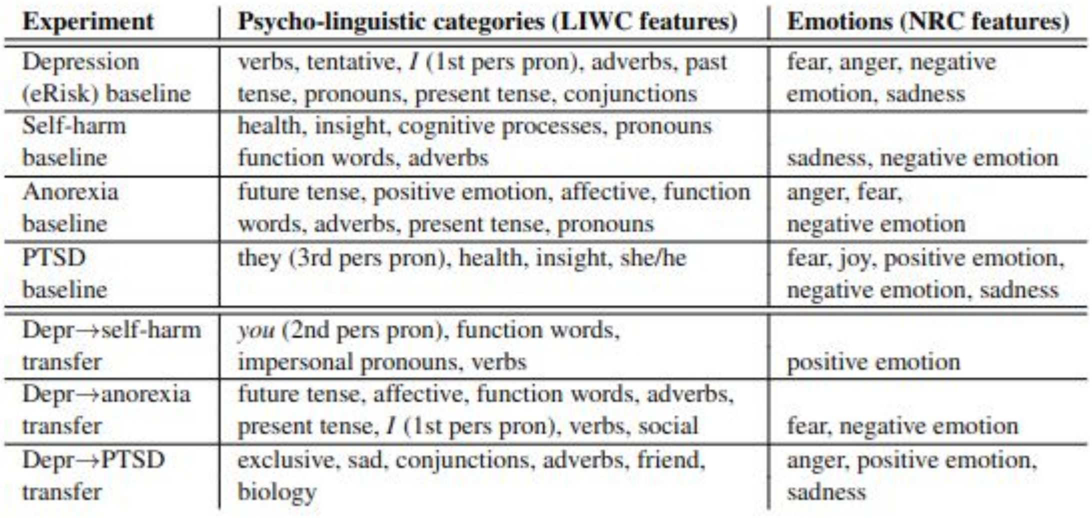
Transfer learning experiments. Ablation



Source	eRisk				CLPsych	
Target	Anorexia		Self-harm		PTSD	
	F1	AUC	F1	AUC	F1	AUC
All-word seq	.49	.88	.24	.77	.57	.74
All-function words	.51	.90	.61	.83	.57	.77
All-lexicon feat	.50	.91	.42	.81	.54	.75
All features	.63	**.93**	.67	**.87**	.58	**.78**

Ablation results for cross-disorder
transfer learning experiments (fine-tuning strategy)

Transfer learning experiments. Error analysis



Experiment	Psycho-linguistic categories (LIWC features)	Emotions (NRC features)
Depression (eRisk) baseline	verbs, tentative, *I* (1st pers pron), adverbs, past tense, pronouns, present tense, conjunctions	fear, anger, negative emotion, sadness
Self-harm baseline	health, insight, cognitive processes, pronouns function words, adverbs	sadness, negative emotion
Anorexia baseline	future tense, positive emotion, affective, function words, adverbs, present tense, pronouns	anger, fear, negative emotion
PTSD baseline	they (3rd pers pron), health, insight, she/he	fear, joy, positive emotion, negative emotion, sadness
Depr→self-harm transfer	*you* (2nd pers pron), function words, impersonal pronouns, verbs	positive emotion
Depr→anorexia transfer	future tense, affective, function words, adverbs, present tense, *I* (1st pers pron), verbs, social	fear, negative emotion
Depr→PTSD transfer	exclusive, sad, conjunctions, adverbs, friend, biology	anger, positive emotion, sadness

Features with highest differences between correctly classified and misclassified texts.

Conclusions & future work

Our experiments have shown that transfer learning could be leveraged to build detection models for disorders where annotated data is scarce.

We have investigated and demonstrated the similarity between manifestations of different disorders at different levels of language (some more than others).

Future: multi-modal solutions and sentence embeddings as models; additional disorders with known comorbidities.

Thank you!

¡Gracias!

Mersi!

Merci!