Deep learning for mental health disorders

Ana-Sabina Uban
(+Paolo Rosso, Berta Chulvi, Petr Lorenc)

BioMedical NLP

Other tasks

- Detecting the **severity** of depression / suicide risk level
- Detecting specific symptoms (lack of sleep, loss of appetite, lack of energy...)
- Detecting **causes** of depression - helps with prevention, and with targeted management
- Detecting depression from video therapy sessions (based on video/audio signals)
- Analyze different disorders jointly (co-morbidities); transfer learning
- Profiling users suffering from a disorder: age, behavioral patterns, social media activity patterns (nocturnal, seasonal)
- Conversational data: therapy sessions, therapist chatbot (https://woebothealth.com/)
- Multimodal depression detection
- Social media: depression and **aggression**

In practice: eRisk 2021

Best results in overall level of depression prediction (some metrics) at Task 3:
http://ceur-ws.org/Vol-2936/paper-75.pdf

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 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

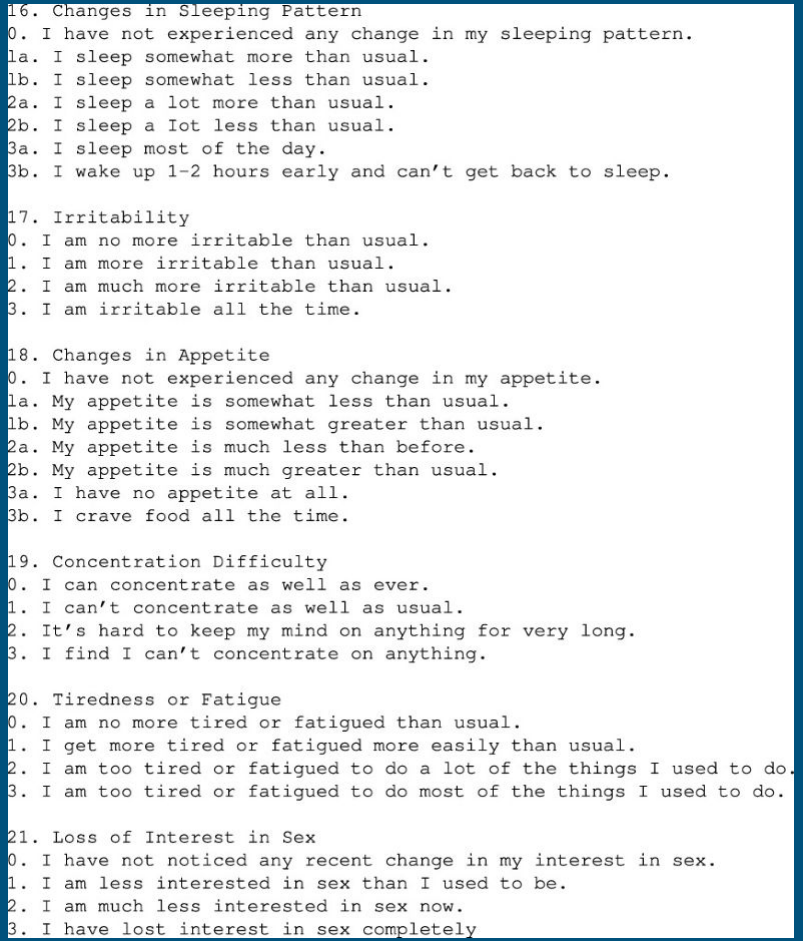
Mental disorders: automatic detection

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.

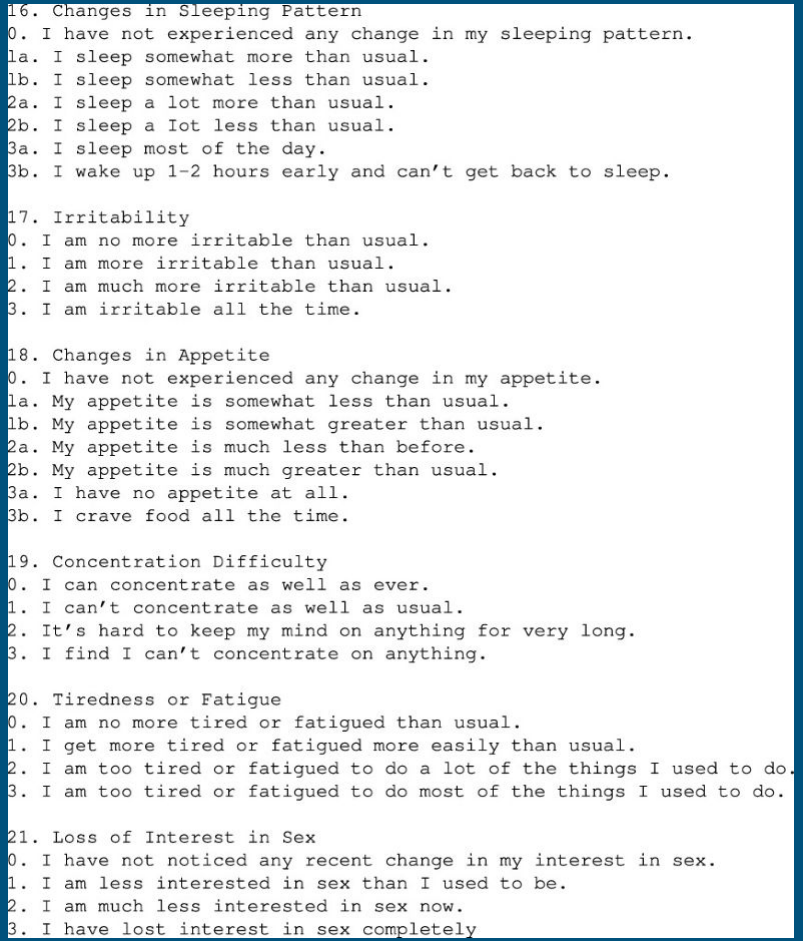
20. Tiredness or Fatigue
0. I am no more tired or fatigued than usual.
1. I get more tired or fatigued more easily than usual.
2. I am too tired or fatigued to do a lot of the things I used to do.
3. I am too tired or fatigued to do most of the things I used to do.

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

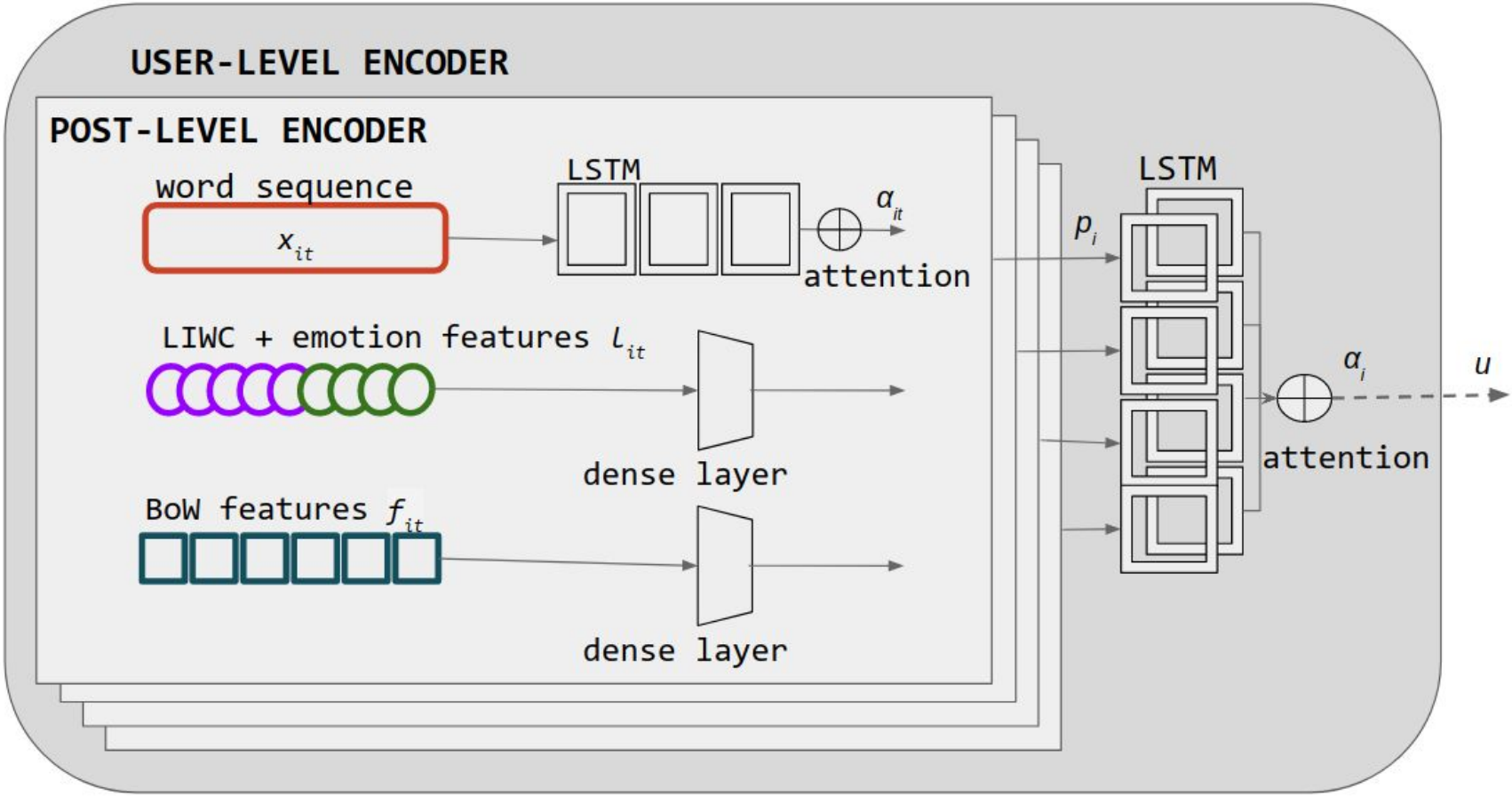
NRC emotions (Plutchik's 8 emotions + 2 sentiments):

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

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

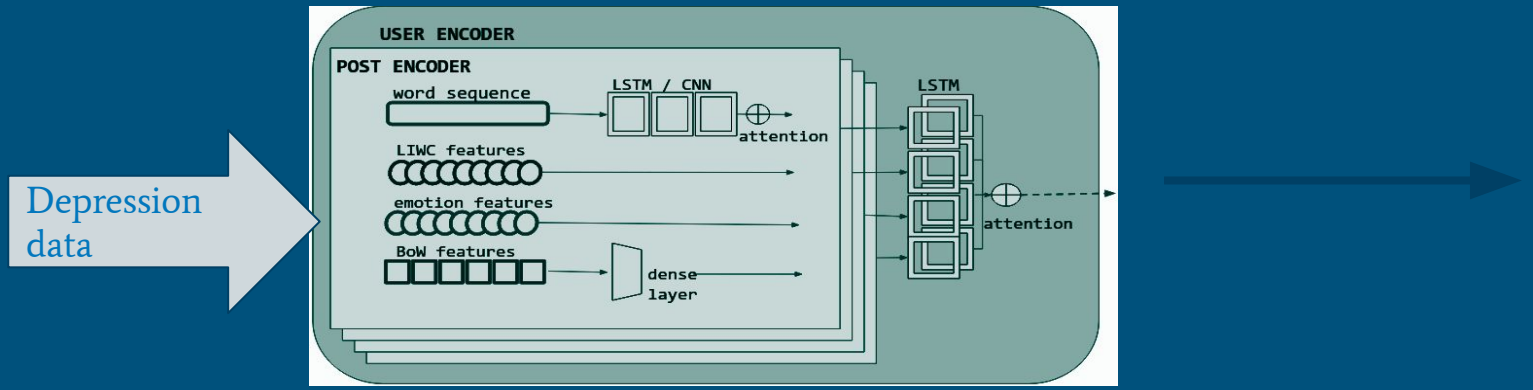
Twitter

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

Confusion matrices for classification between disorders

Transfer learning

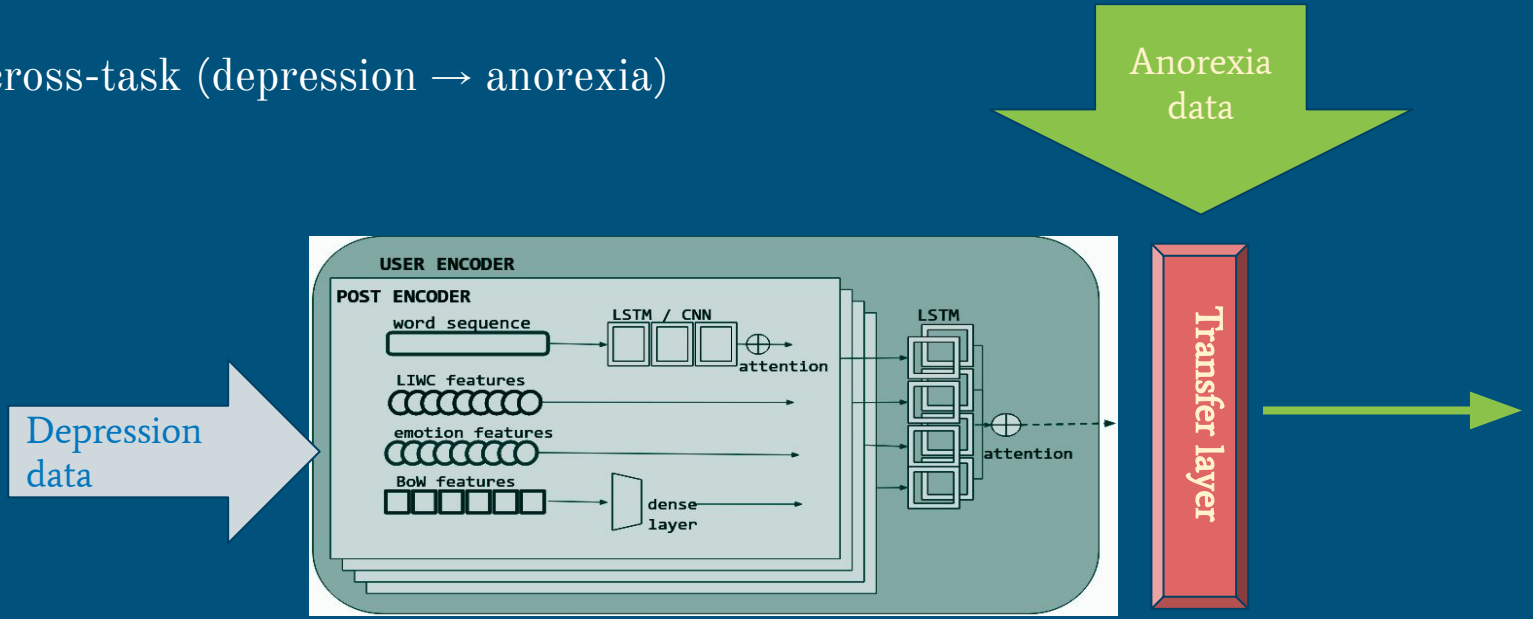
Strategy 0. Zero-shot



Transfer learning

Strategy 1. Transfer layer

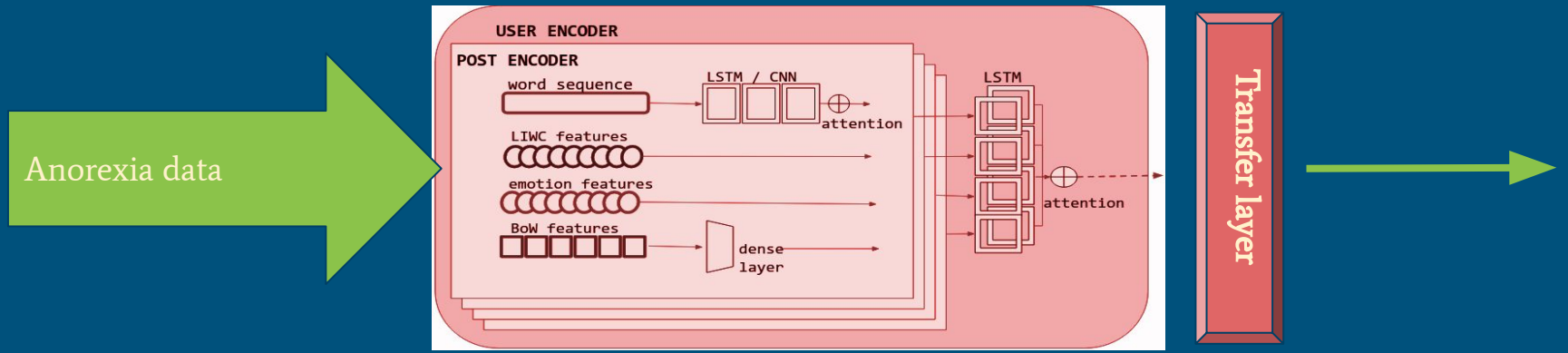
Example: cross-task (depression → anorexia)



Transfer learning

Strategy 2. Fine-tuning

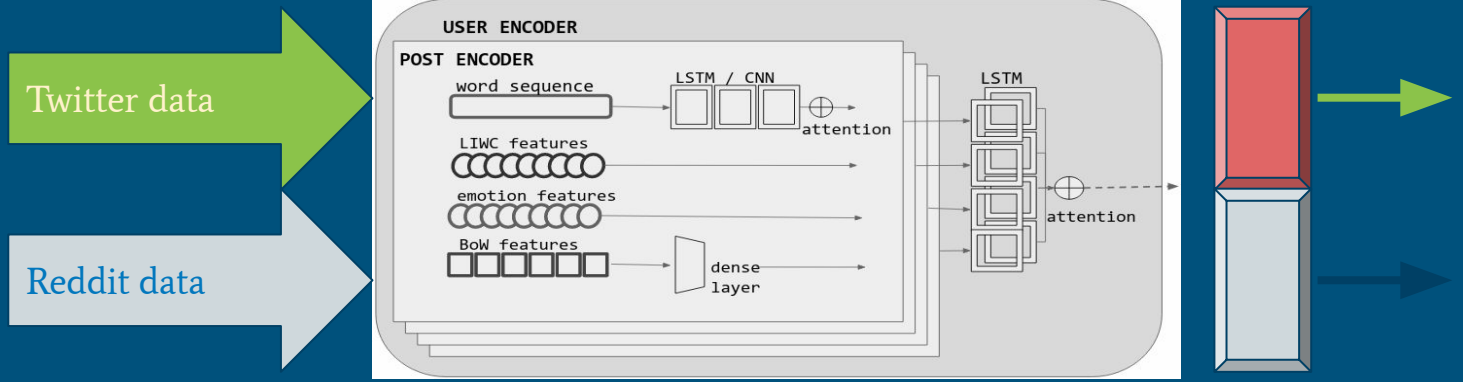
Example: cross-task (depression → anorexia)



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

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

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!

Transfer learning

Clinical evidence of comorbidity within mental disorders. (Exploring Comorbidity Within Mental Disorders Among a Danish National Population)

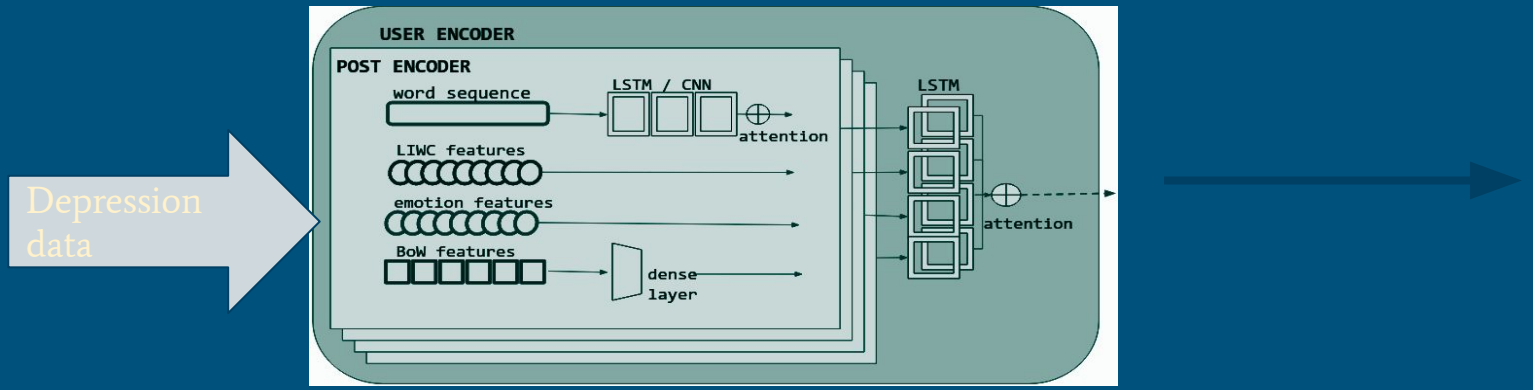
- Improve performance on tasks with less data (depression → other disorders)
- Understand connection/compatibility between disorders and expression media (genre/platform)

Cross-task - transfer knowledge between labels for different disorders

Cross-genre - transfer knowledge between different data platforms (reddit/Twitter)

Transfer learning

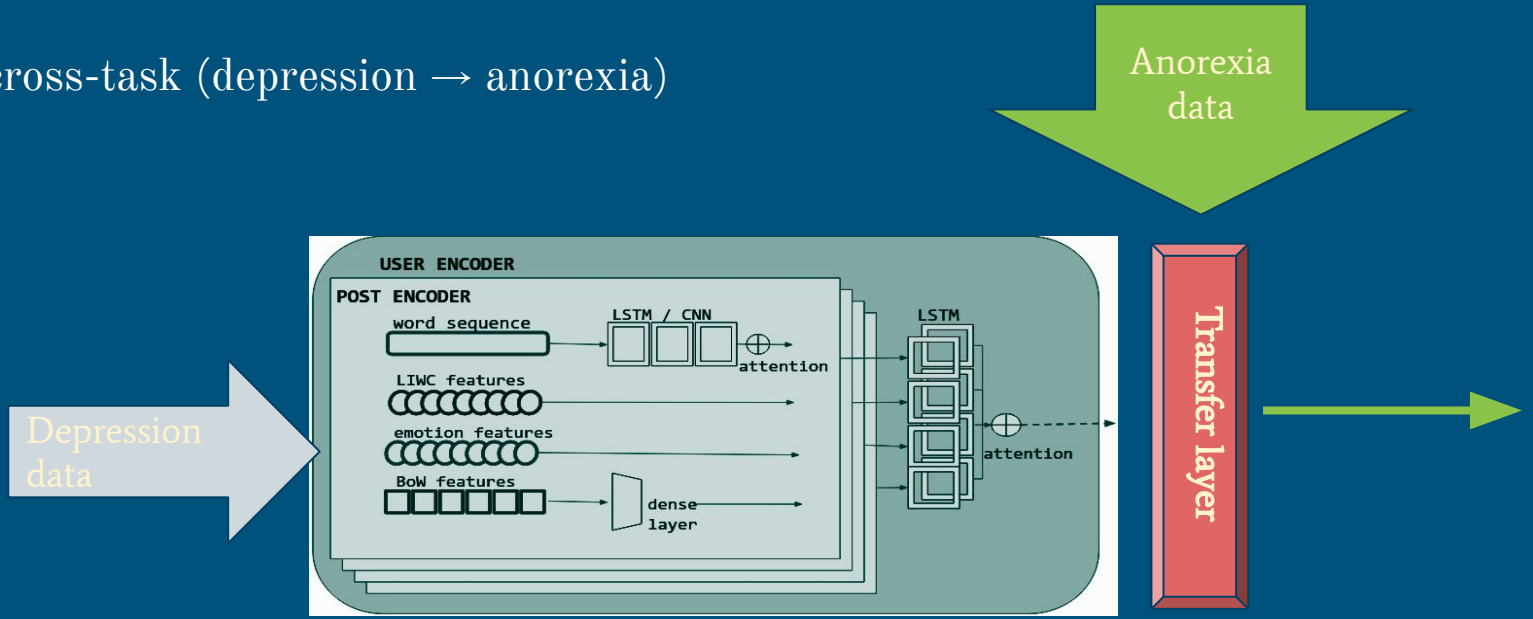
Strategy 0. No pre-training



Transfer learning

Strategy 1. Transfer layer

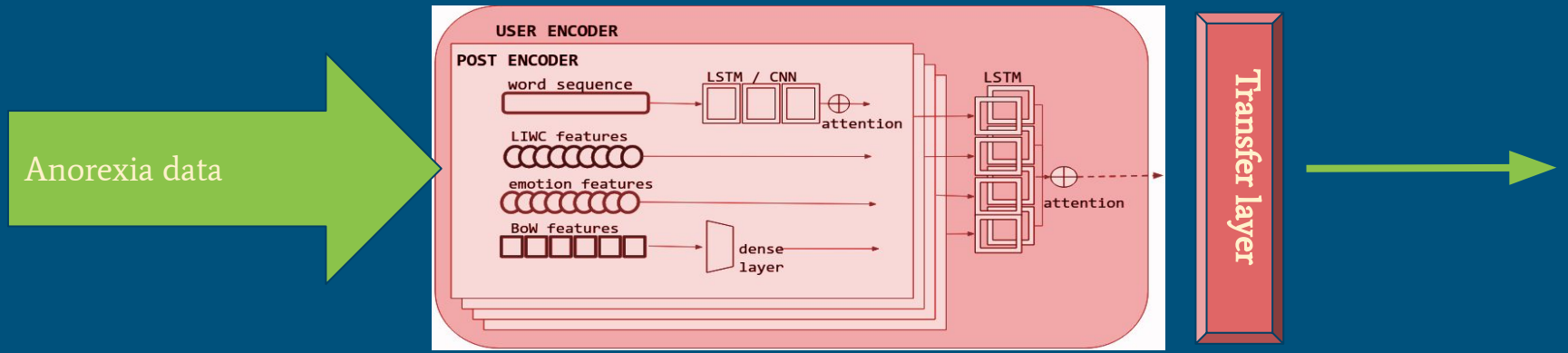
Example: cross-task (depression → anorexia)



Transfer learning

Strategy 2. Fine-tuning

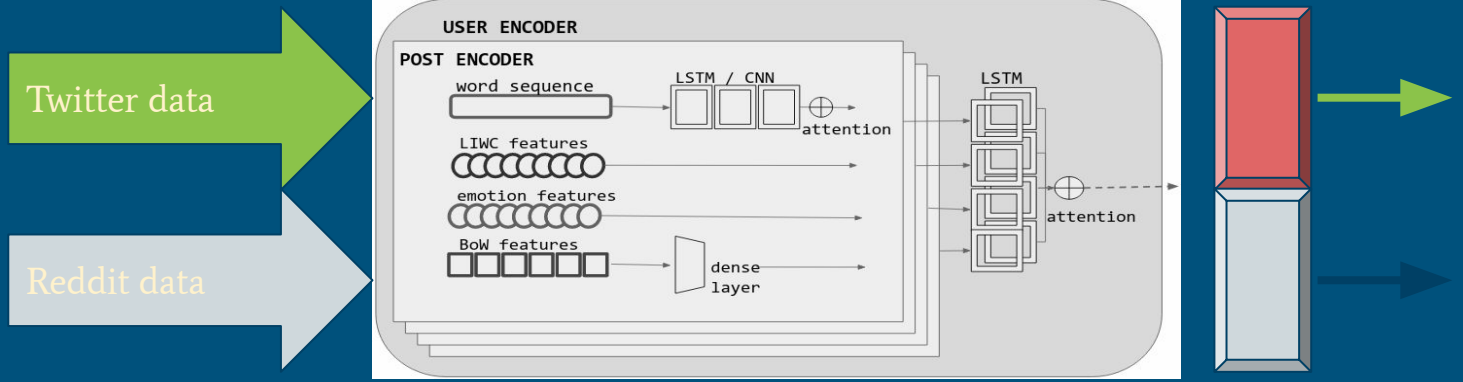
Example: cross-task (depression → anorexia)



Transfer learning

Strategy 3. Multi-task learning

Example: cross-genre (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

Table 3: 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	.40	.83	.75	.83	.56	.72

Depression and offensive language

Depression and aggression linked in psychology studies (especially self-aggression:)

Frontiers | Aggressive and Disruptive Behavior Among Psychiatric Patients With Major Depressive Disorder, Schizophrenia, or Alcohol Dependency and the Effect of Depression and Self-Esteem on Aggression | Psychiatry

Computational linguistics:

An Exploratory Analysis of the Relation between Offensive Language and Mental Health - ACL Anthology
(Ana-Maria Bucur, Marcos Zampieri and Liviu P. Dinu)

Depression and offensive language

RQ1: Are posts from individuals suffering from depression more likely to contain offensive language in existing datasets?

RQ2: Are there differences in the nature of offensive language used by individuals with depression compared to control groups?

Depression and offensive language

- Fine-tuned a BERT model on the OLID dataset (offensive language + subtypes according to target)
- Applied it for prediction on corpora of depressed users on social media
- Also selecting posts showing signs of depression based on emotions and polarity
- There are more posts with signs of depression labeled as offensive, the majority of them are untargeted (containing swears, profanity)
- Depressed individuals tend to use more self-deprecating content and less deprecation of others

Depression and offensive language

	Self-reported		Signs of depression	
Dataset	**Depression**	**Control**	**Showing**	**Not showing**
eRisk 2018	8.24%	5.91%	18.50%	7.40%
RSDD	11.31%	8.91%	24.33%	10.10%

Table 1: Percentage of posts labeled as offensive from total posts of self-reported individuals and of individuals showing/not-showing signs of depression measured with the H_s heuristic.

Depression and offensive language

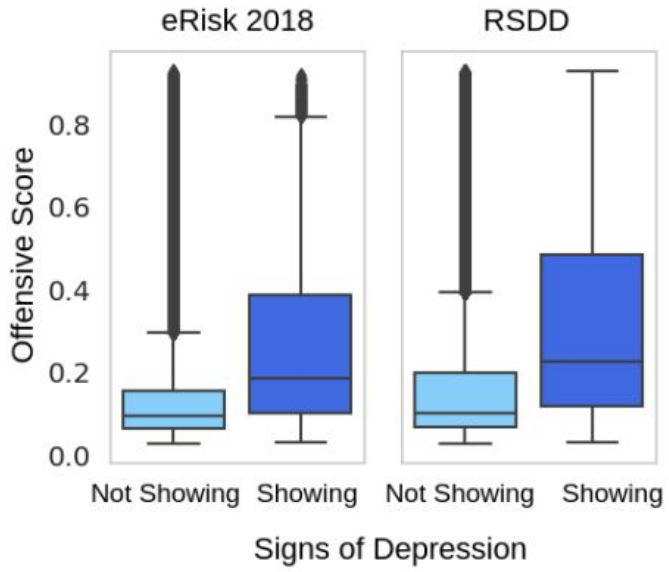


Figure 1: Distribution of the offensive language score for posts written by users with self-reported depression diagnosis and showing or not showing signs of depression measured with the H_s heuristic.

Depression and offensive language

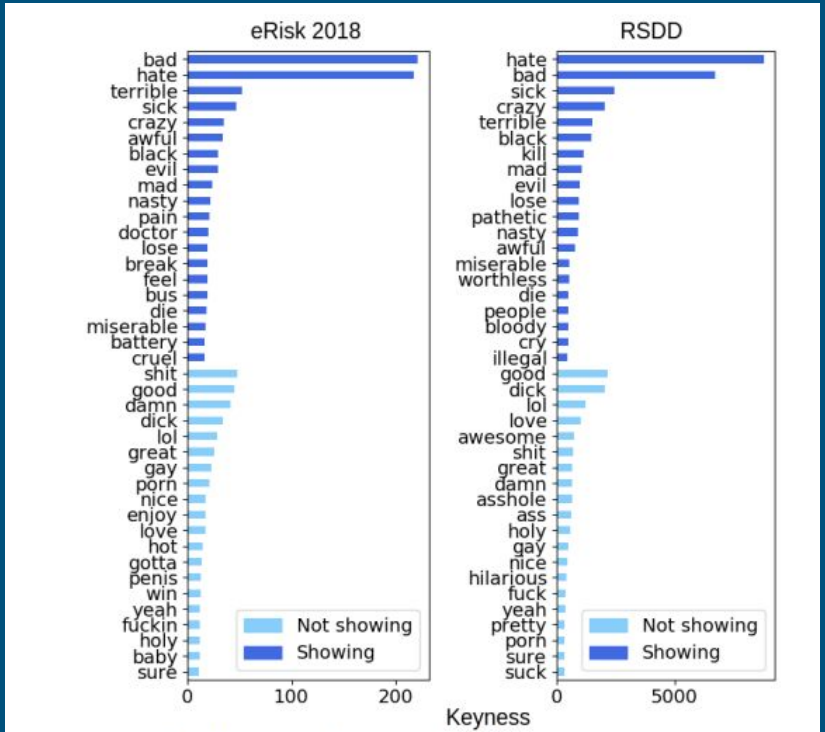


Figure 2: Keyness for words from posts showing/not showing signs of depression.

Mental disorders in conversational domain

Data:

Possible source: Therapy sessions transcripts (...)

Applications:

Chatbot “therapist”

Notable example: https://woebothealth.com/

Mental disorders in conversational domain

Woebot: https://woebothealth.com/

Clinically tested therapeutic approaches

Cognitive Behavioral Therapy (CBT), Interpersonal Psychotherapy (IPT), and Dialectical Behavioral Therapy (DBT) provide the foundation for Woebot's therapeutic support.

AI-powered delivery

Through the use of AI and NLP, Woebot forms trusted bonds and delivers clinically validated techniques in an approachable, conversational manner.

Mental disorders in conversational domain

Woebot: https://woebothealth.com/

Efficiency proved in studies:

Evidence of Human-level Bond Established with a Digital Conversational Agent: An Observational Study

Acceptability of Postnatal Mood Management Through a Smartphone-based Automated Conversational Agent

Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial

Mental disorders in conversational domain

Woebot: https://woebothealth.com/

Efficiency proved in studies:

Satisfaction with the chatbot was assessed by the Client Satisfaction Questionnaire (CSQ-8)

Acceptability of the chatbot as a mood management tool was assessed with the Working Alliance Inventory - Short Revised (WAI-SR) survey, a measure of therapeutic alliance that correlates with favorable psychotherapy outcomes.

Datasets on mental health in conversational domain + availability

- Not readily available
  - Crisis Counseling webpage
    - Real SMS Conversations
    - 5 millions conversation, several labels
    - Need to do a long process of starting cooperation with Crisis counseling (months)
  - Motivational Interviewing Dataset
    - Transcripts
    - 22,719 counselor utterances
    - Get after request? Various labels/disorders
  - Annotating Reflections for Health Behavior Change Therapy
    - Transcripts
    - 324 sessions

Datasets on mental health in conversational domain + availability

- Publicly available...
  - Counsel Chat
    - Single talk turn
    - <1k utterances
  - **Counseling and Psychotherapy Transcripts Series (CPTS)**
    - Transcripts
    - 4,000 transcripts of real therapy sessions (various labels/disorders)
    - (requires account)
  - **Distress Analysis Interview Corpus (MIT paper) (DAIC-WOZ)**
    - Transcripts
    - 142 interactions, 3K words

Mental disorders in conversational domain

Logistic regression model on conversational data - depressed vs non-depressed feature analysis:

environment (-7.5), open-minded (-6.3), or accomplish (-4.7) - associated to non-depressed patients.

insignificant (5.36), television (5.66), or pollution (+6.1) - associated to depressed patients

Mental disorders in conversational domain
Transfer learning from non-conversational domain (social media data)

Table 3. Results - biLSTM stands for bidirectional LSTM, use5 - Universal Sentence Encoder based on Transformers, "f" stands for additional features.

	DAIC-WOZ	eRisk/GPC without fine-tuning	eRisk/GPC with fine-tuning
biLSTM + use5	**0.660**	**0.651** / 0.440	**0.803** / 0.690
biLSTM + use5 + f	0.565	0.541 / 0.410	0.597 / 0.511

Mental disorders in conversational domain Transfer learning from non-conversational domain (social media data)

Is the size of the source domain dataset more critical than domain relatedness?

Using GPC as the source domain underperforms the setting in which eRisk data is used as the source domain, even though GPC is a more similar type of data to our target dataset =>

The size of the source domain dataset is more important than domain closeness.

Multimodality and mental disorder detection

Multimodal approach = approach involving multiple modes of communication

Multimodal data:

- Text +
- Images
- Video
- Speech

- Behavioral data
- Sensor data
- Physiologic testing

- Social media metadata
- Medical information
- etc

Features types for mental disorder detection

- **Text** posted by user on social media
- Audio/video data from interviews (therapy sessions) <- AVEC Workshop
- Sensor data from smartphones, eHealth: heart rate, eye sensor, accelerometer, calls, location, dermal activity, microphone, video cameras, gps, ...
- Behavioral data: social media metadata (login times, interactions, followers, ...)
- Demographic data (gender, age, ...)
- Medical statistics (psychometric and physiologic testing, indices of awareness, consciousness, insight and anosognosia, unawareness, and awareness of illness)
- Images posted by user on social media / profile pictures
- Disorders approached:
    - Images features: suicide, depression, anxiety
    - General: depression, types of PTSD (behavioral), bipolar, stress, others (audio/video)

Multimodal mental health: images

- Types of image features:
  - their five-color combinations, brightness, saturation, cool color ratio, and clear color ratio
  - Image embeddings, features extracted directly from images
  - Image captions/tags
- Platforms: Twitter, Instagram
- Types of images: avatar images/profile (Twitter), shared images on Instagram (selfies?), shared images on Twitter (generic)

Research questions

- ➔ Do users suffering from depression/a mental disorder post a different kind of images (various perspectives)?
- ➔ Is the content in the images (objects depicted) different than in healthy users?
- ➔ Can the combination of text and image features help with detection of mental disorders?

Colors and mental health: insights

The strong associations between color sensitivity and mood has been highlighted by several studies [...]. In an earlier research, a strong correlation between specific color selection such as yellow and depressive behavior has been reported by [...].

Reece and Danforth [...] analyzed uploaded images to Instagram and found that photos posted by depressed users were more likely to be bluer, grayer and darker.

In studies associating mood, color, and mental health, healthy individuals identified darker, grayer colors with negative mood, and generally preferred brighter, more vivid colors [...]. By contrast, depressed individuals were found to prefer darker, grayer colors

Mental health / personality multimodal studies

Study	# Users	Traits	Image Type	Image Features
Our Work	**887 + 4,132**	Depression & Anxiety	Twitter Posted & Profile Images	Color, Facial, Aesthetics, Content, VGG-Net
(Reece and Danforth 2017)	166	Depression	Instagram Photos	Colors
(Andalibi, Ozturk, and Forte 2015)	–	Depression	500 '#depression' Instagram Photos	Manual Annotation
(Ferwerda and Tkalcic 2018)	193	Personality (continuous)	Instagram Photos	Content
(Nie et al. 2018)	2238	Perceived Personality (continuous)	Web Portrait Images	Facial, Social information
(Samani et al. 2018)	300	Personality (continuous)	Twitter and Flickr Posts, Likes, & Profiles	Colors, Content, VGG-Net
(Farnadi et al. 2018)	5670	Personality (binary)	Facebook Profile Images	Facial, Text, Likes
(Guntuku et al. 2017a)	4132 + 161	Personality (continuous)	Posted, liked images and text on Twitter	Color, Bag of Imagga tags, VGG-Net
(Segalin et al. 2017)	11,736	Personality (continuous & binary)	Facebook Profile Images	Aesthetics, BOVW, VGG-Net, IATO
(Liu et al. 2016)	66,502	Personality (continuous)	Twitter Profile Images	Color, Facial
(Ferwerda, Schedl, and Tkalčič 2016)	113	Personality (binary)	Instagram Photos	Colors, #Faces, Filters
(Skowron et al. 2016)	62	Personality (binary)	Instagram Photos	Colors
(Guntuku et al. 2016)	300	Personality (continuous)	Liked ('Fave') images on Flickr	Colors, semantic features, aesthetics
(Guntuku et al. 2015)	123	Personality (continuous)	Selfies on Weibo	Color, Aesthetics, BOVW, Emotions
(Al Moubayed et al. 2014)	829	Personality (binary)	Face Images	Eigenfaces
(Celli, Bruni, and Lepri 2014)	112	Personality (binary)	Facebook Profile Images	Bag-of-Visual-Words (BOVW)

Table 1: Summary of data and methods used in previous work analyzing images of individual users.

Multimodal studies on mental health with social media images

Depression:

- Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution [1]
- What Twitter Profile and Posted Images Reveal about Depression and Anxiety [2]
- Multimodal mental health analysis in social media [6]
- Depression-related Imagery on Instagram [8]

Suicide:

- Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis [3]

Image-only/depression:

- Instagram photos reveal predictive markers of depression [4]

Emotion:

- Multimodal Classification for Analysing Social Media [5]

Multimodal studies on mental health with social media images (References)

[1] Shen, Guangyao, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, Tat-Seng Chua, and Wenwu Zhu. "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution." In IJCAI, pp. 3838-3844. 2017.

[2] Guntuku, Sharath Chandra, Daniel Preotiuc-Pietro, Johannes C. Eichstaedt, and Lyle H. Ungar. "What twitter profile and posted images reveal about depression and anxiety." In Proceedings of the international AAAI conference on web and social media, vol. 13, pp. 236-246. 2019.

[3] Ramírez-Cifuentes, Diana, Ana Freire, Ricardo Baeza-Yates, Joaquim Puntí, Pilar Medina-Bravo, Diego Alejandro Velazquez, Josep Maria Gonfaus, and Jordi Gonzàlez. "Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis." Journal of medical internet research 22, no. 7 (2020): e17758.

[4] Reece, Andrew G., and Christopher M. Danforth. "Instagram photos reveal predictive markers of depression." EPJ Data Science 6 (2017): 1-12.

[5] Duong, Chi Thang, Remi Lebret, and Karl Aberer. "Multimodal classification for analysing social media." arXiv preprint arXiv:1708.02099 (2017).

Multimodal studies on mental health with social media images (References)

[6] Yazdavar, Amir Hossein, Mohammad Saeid Mahdavinejad, Goonmeet Bajaj, William Romine, Amit Sheth, Amir Hassan Monadjemi, Krishnaprasad Thirunarayan et al. "Multimodal mental health analysis in social media." Plos one 15, no. 4 (2020): e0226248.
[7] Garcia-Ceja, Enrique, Michael Riegler, Tine Nordgreen, Petter Jakobsen, Ketil J. Oedegaard, and Jim Tørresen. "Mental health monitoring with multimodal sensing and machine learning: A survey." Pervasive and Mobile Computing 51 (2018): 1-26.
[8] Nazanin Andalibi, Pinar Ozturk, and Andrea Forte. 2015. Depression-related Imagery on Instagram. In Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing (CSCW'15 Companion). Association for Computing Machinery, New York, NY, USA, 231–234. DOI:https://doi.org/10.1145/2685553.2699014

Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution [1]

- Disorder: Depression
- Platform: Twitter
- Features - 6 modalities: social network features, user profile features, visual features, emotional features, topic-level features, and domain-specific features.
- Image types: user avatars
- Image features: their five-color combinations, brightness, saturation, cool color ratio, and clear color ratio
- Model: multimodal depressive dictionary learning model (MDL) to learn the sparse user representations + relations between the modalities

What Twitter Profile and Posted Images Reveal about Depression and Anxiety [2]

- Platform: Twitter, + Facebook (text only)
- Disorders: depression, anxiety
- Labels: survey-based ground-truth labels (regression)
- Features: text, images (avatars of users + posted images), demographic information
- Image features: Color, Facial (emotion, posture, ...), Aesthetics (symmetry, lightning, harmony,...), Content, VGG-Net (bag-of-tags + dimensionality reduction; image vectors)
- Results indicate that multi-task learning gives significant improvements in performance for modeling mental health conditions jointly with demographics (here age and gender), factors which clinicians usually consider while diagnosing patients. Further, models trained on larger data sets using text-predicted outcomes show reliable performance when predicting more reliable survey based mental health outcomes.

What Twitter Profile and Posted Images Reveal about Depression and Anxiety [2]

- *For depression, we find that profile pictures suppress positive emotions rather than display more negative emotions, likely because of social media self-presentation biases. They also tend to show the single face of the user (rather than show her in groups of friends), marking increased focus on the self, emblematic for depression.*
- *Posted images are dominated by grayscale and low aesthetic cohesion across a variety of image features.*
- *Anxious users additionally seem to post marginally more content related to family and work*
- *Users high in depression post images which emphasize foreground objects and are low in brightness, (...) images containing text, animals, while users low in depression and anxiety post images of sports, nature, every day things*

Detection of Suicidal Ideation on Social Media: Multimodal, Relational, and Behavioral Analysis [3]

- Disorder: Suicide, platform: Twitter
- Collected dataset: 252 users annotated by clinicians
- Features: images, behavioral, texts
- Image features: CNN (ResneXt) pre-trained on Instagram suicide-related images => user score
- *The combination of textual, visual, relational, and behavioral data outperforms the accuracy of using each modality separately. We defined text-based baseline models based on bag of words and word embeddings, which were outperformed by our models, obtaining an increase in accuracy of up to 8% when distinguishing users at risk from both types of control users.*

Instagram photos reveal predictive markers of depression [4]

- Disorder: depression
- Only images; various hypotheses tested & statistical analyses
- Statistical features were computationally extracted from 43,950 participant Instagram photos
- Image features: color analysis, metadata components, and algorithmic face detection.
- *Human ratings of photo attributes (happy, sad, etc.) were weaker predictors of depression*
- *Our results supported Hypothesis 1, that markers of depression are observable in Instagram user behavior, and Hypothesis 2, that these depressive signals are detectable in posts made even before the date of first diagnosis.*

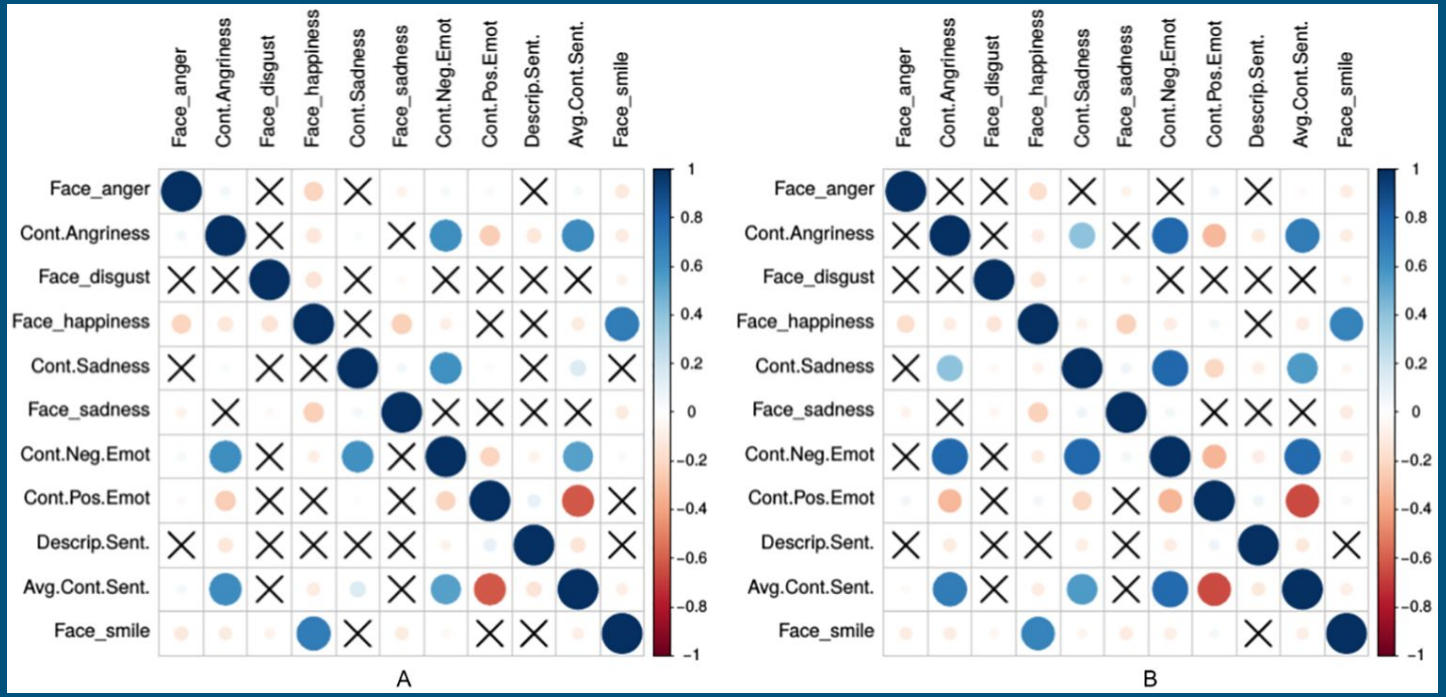
Depression-related Imagery on Instagram [8]

- Disorder: depression
- Platform: Instagram
- Dataset: images posted by users, #depression
- Quantitative analysis (no classification): Examined the distribution of themes in 500 images, uncovering themes posted by depressed users + relationship between image and caption
- Image features: manually coded image descriptions
- *Images alone were used to communicate about most topics; however, seeking or providing help/support/engagement, and positive emotions were shared more often through textual captions.*
- *In posts that depicted individuals and were not celebrities or screen shots, we identified 72% females and 17% males and around 4% minors*
- *13% of captions were descriptive of the image, 35% provided contextual and different information, 32% provided additional but similar information, and 20% were unrelated.*

Multimodal mental health analysis in social media [6]

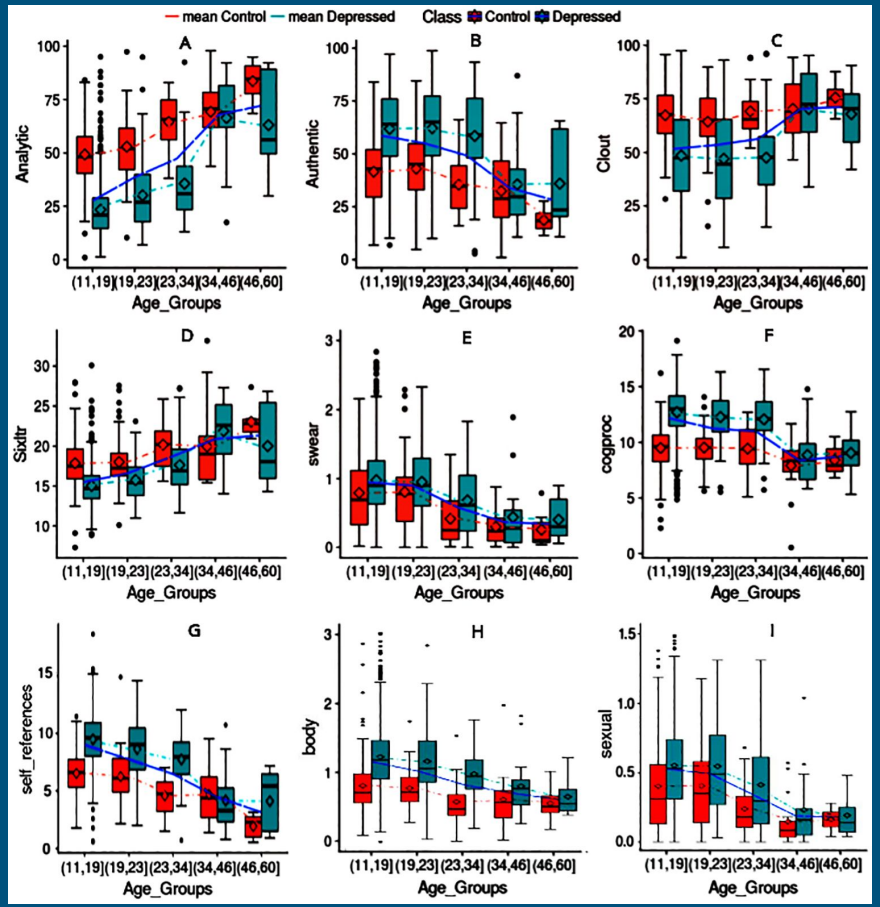
- Disorder: depression
- Data: Twitter (collect the Twitter profiles of individuals with self-declared depressive symptoms + age, gender)
- Features: visual, textual, and user interaction data
- Analysis of the content of posted images in terms of colors, aesthetic, facial presentation, and their associations with depressive symptoms;
- Uncovering the underlying relationships between visual and contextual content
- Improve the state-of-the-art for depression detection on Twitter: average F1-Score increased by 5 percent
- Feature importance analysis
- Demographic analysis: predicting age and gender on basis of text + images
- Thinking styles features & analysis: authenticity, self-references, informal, certainty (LIWC-based)
- *(...) a strong association between female gender, and expression of depressive symptoms on Twitter.*

Multimodal mental health analysis in social media [6]



The Pearson correlation between the average emotions derived from facial expressions through the shared images and emotions from textual content for depressed-(a) and control users-(b).

Multimodal mental health analysis in social media [6]



Characterizing linguistic patterns in two aspects: Depressive behavior and age distribution.

Multimodal Classification for Analysing Social Media [5]

- Emotion classification task
- Datasets: reddit, flickr
- Approach: neural networks (CNN for images) with fusion for text + image classification, + emotions
- *For the reddit dataset, the difference is only 0.8% for common space fusion while it is 34% for joint fusion. The discrepancy for common space fusion is lower as it considers image and text vector of a post as equally important. As a result, using either image or text to classify, we get similar accuracy. On the other hand, joint fusion considers textual information as more important. This makes the classification using text significantly better than using only image*

multiRedditDep - a multimodal depression dataset

Collected images posted on Reddit + titles (text)
Positive users labeled based on self-stated diagnoses: ”I was diagnosed with depression”...
Control users are random users who have posted at least once on the */r/depression* forum, but that have not mentioned any diagnosis of depression in the recorded history of their activity

	Images	Unique users
Depressed	48,121	1,419
Control	72,855	2,344

Classification experiments - multimodal

Models for classifying posts automatically based on text and visual features: feed-forward neural network

Text features (on image titles):

- word embeddings (word2vec)
- sentence embeddings (BERT)

Image features:

- image tags extracted with pretrained VGG model
- Text extracted from images which constitute screenshots
- + word2vec embeddings of tags

Classification experiments - multimodal

Future:

+ Visual features:
- Image embeddings
- Color information (color histogram, cool color ratio, brightness, saturation)
- Additional features from image (facial features, emotion/sentiment features,...)
- Fusion techniques

- Other disorders (anorexia!)

Classification experiments - multimodal

Features	Accuracy	AUC
Text only: word2vec embeddings	62.73	65.29
Images only: word2vec embeddings on image tags	61.88	61.63
Text and images: word2vec embeddings of text, image tags	**65.40**	**68.69**
Text only: BERT + word2vec embeddings	66.24	69.20
Text and images: BERT + word2vec embeddings	**66.39**	**69.37**

Multimodal dataset - image analysis

We build a taxonomy of objects depicted in the images in the dataset, based on clustering the image tags semi-automatically: word embeddings + manual curation => **6 categories** + sub-categories

	Depressed	Control	Significant sub-categories
Animal	16.92	13.51	Dogs, cats, wild mammals, wild aquatic, insects&reptiles, wild birds, rodents
Vegetal	0.07	0.06	-
People	14.95	15.71	Body parts
Actions & movement	11.97	14.64	shopping
Things	48.45	48.53	Food, health objects, technology, musical instruments
Scenery	7.61	7.52	Landscapes and buildings

More multimodal classification on multiRedditDep

See Ana Bucur's research in the Valencia team: https://scholar.google.com/citations?user=TQuQ5IAAAAAJ&hl=en

Presentation tomorrow: https://isds.unibuc.ro/interdisciplinary-research-training-groups/isds-conferences/