

Comparing emotion feature extraction approaches for predicting depression and anxiety

Hannah A. Burkhardt¹, Michael D. Pullmann¹,
Thomas D. Hull², Patricia A. Areán¹, Trevor Cohen¹

¹University of Washington, Seattle ²Talkspace, New York, NY

Background

- Counseling is informed by patient symptoms
- Emotional state changes are symptoms; differ by emotion (e.g. pride), not just sentiment (pos./neg.)
- Patient-generated text from digital mental health services can be used to develop automatic assessments via measuring emotion
- There are well-established tools, but they have limitations

Can NN-based emotion extraction methods (GoEmo) outperform established word counting methods (LIWC)?

Linguistic Inquiry and Word Count (LIWC)

- Established word-counting tool used for mental-health prediction tasks (Pennebaker et al., 2007)
- Counts words belonging to pre-defined categories
- Categories with known relationship to anxiety/depression: first-person singular pronouns (“I”), first-person plural pronouns (“we”), bio, health, sadness, anxiety, anger, pos. and neg. emotion. (Tausczik and Pennebaker, 2010)

GoEmotions

- BERT-based classifier pipeline trained on GoEmotions (annotated Reddit posts) (Demszky et al., 2020)
- Granularity settings: **6 basic emotions** (Ekman, 1992) & **27 fine-grained emotions** (Cowen and Keltner, 2017). Positive and negative emotion features were calculated.

Positive		Negative		Ambiguous
admiration 😄	joy 😄	anger 😡	grief 😞	confusion 😟
amusement 😄	love ❤️	annoyance 😡	nervousness 😟	curiosity 😟
approval 👍	optimism 🙌	disappointment 😞	remorse 😞	realization 💡
caring 😊	pride 😊	disapproval 🗿	sadness 😞	surprise 😮
desire 😍	relief 😊	disgust 😞		
excitement 😄		embarrassment 😞		
gratitude 🙏		fear 😨		

GoEmotions taxonomy: Includes 28 emotion categories, including “neutral”.
<https://ai.googleblog.com/2021/10/goemotions-dataset-for-fine-grained.html>

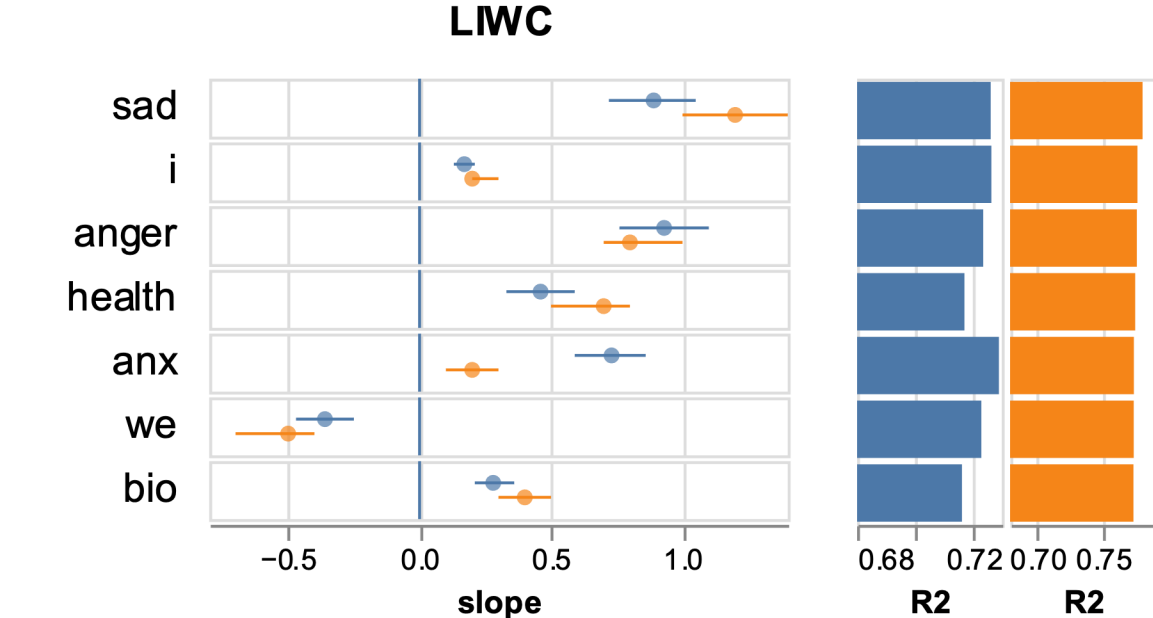
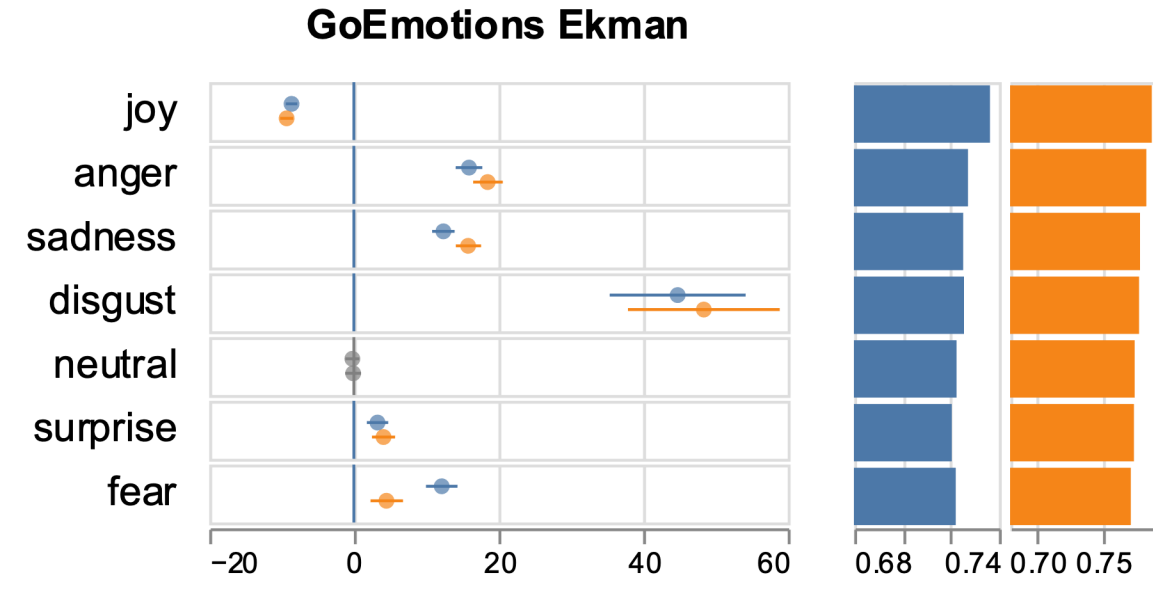
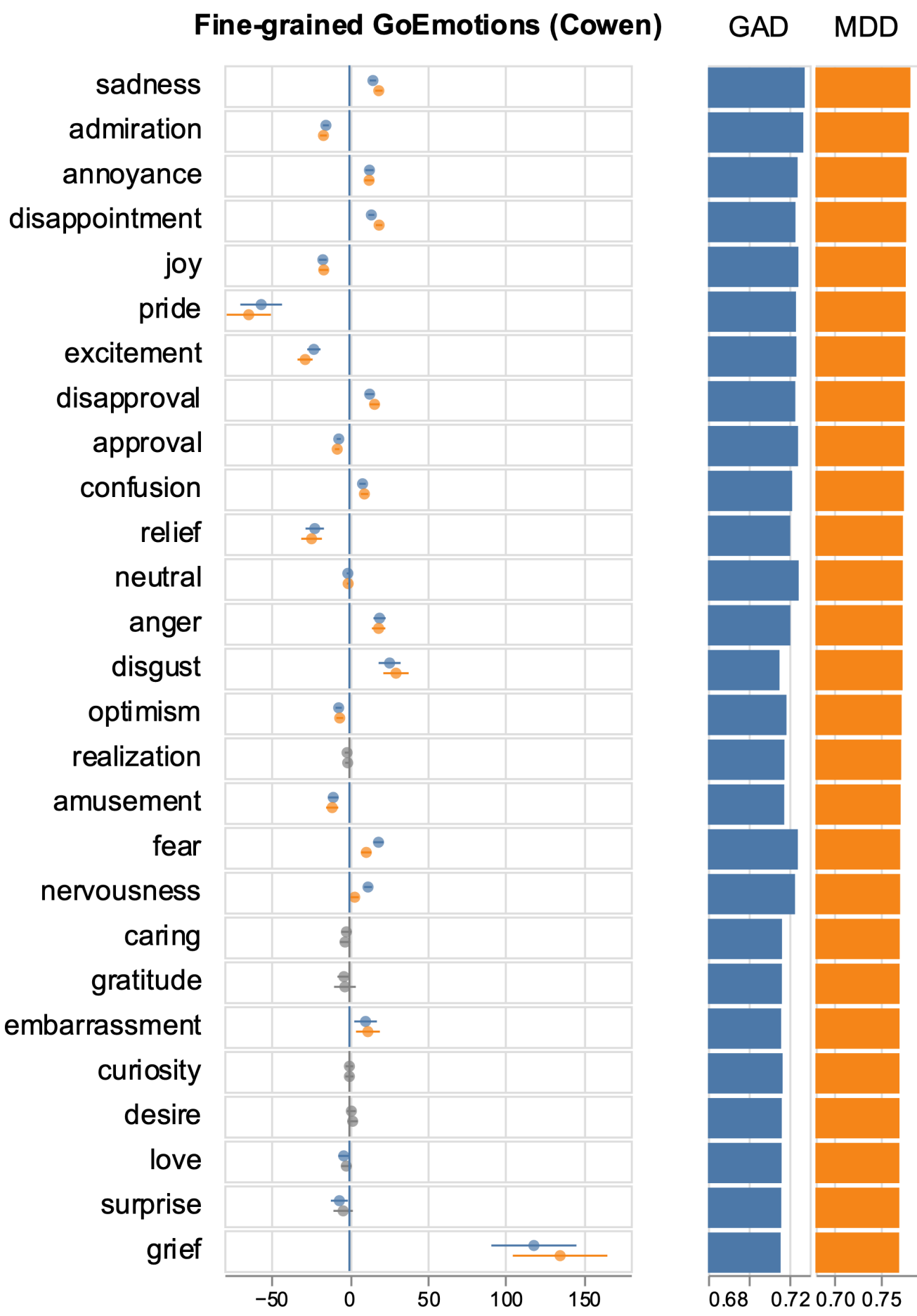
Data

- **13,000 documents labeled with PHQ-9 and GAD-7 scores** created from **>337,000 messages** from message-based therapy sessions from **>6,500 unique patients** collected via Talkspace (Hull et al., 2020)
- Patients and clinicians gave consent; IRB approved; data handled securely.

GoEmotions captures clinically relevant nuance

Mixed-effects linear regressions showed significant associations and high R2s. Findings reflect current understanding:

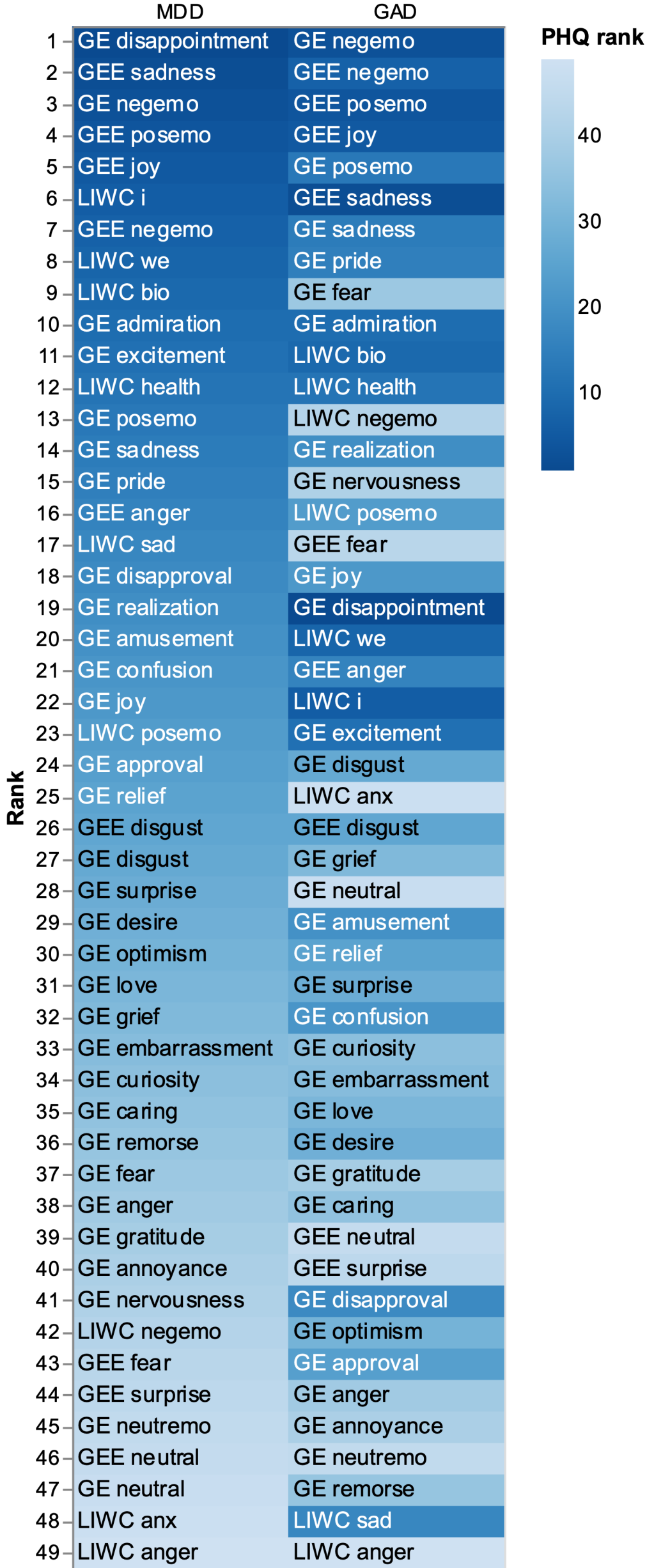
- lower reactivity in depression → less **excitement**
- **grief** often co-occurs with depression
- lower self-image in depression → less **pride**
- feeling of not fitting in, isolation in depression & anxiety → perception of violations of social norms aka self-**disgust**
- heightened self-criticism, interpersonal sensitivity in anxiety → increased **disapproval** / decreased **approval**



95% confidence intervals shown. Variables that were not significant (p>0.05) are shown in gray.

Feature importance differs between depression and anxiety

- Random Forest models trained w/ 80-20 data split (by patient), binary targets (PHQ-9≥10: major depressive disorder, MDD; GAD-7≥10: general anxiety disorder, GAD), using all features
- SHAP feature importance: fear ranked higher for predicting GAD than MDD. Sadness higher for predicting MDD than GAD.



GE = GoEmotions (fine-grained), GEE = GoEmotions Ekman

GoEmotions features are collectively more predictive

- Trained Random Forest models with combinations of feature sets as input.
- Fine-grained GoEmotions features combined with LIWC’s syntactic and topic (non-emotion) features were most predictive.

AUROC, F1 score (positive class), precision, and recall of models trained with combinations of non-emotion and emotion features for predicting MDD/GAD

	MDD				GAD			
	ROC	F1	Pr	Rc	ROC	F1	Pr	Rc
LIWC non-emo	0.577	0.413	0.525	0.341	0.549	0.290	0.478	0.209
LIWC emo	0.621	0.471	0.561	0.405	0.613	0.405	0.541	0.324
GoEmo Ekman	0.643	0.493	0.583	0.427	0.643	0.443	0.550	0.371
GoEmo Cowen	0.662	0.522	0.613	0.455	0.652	0.444	0.565	0.366
LIWC non-emo+								
LIWC emo	0.640	0.484	0.569	0.420	0.617	0.401	0.529	0.324
GoEmo Ekman	0.655	0.498	0.585	0.434	0.637	0.441	0.548	0.369
GoEmo Cowen	0.671	0.514	0.615	0.441	0.654	0.451	0.568	0.374
All three	0.671	0.520	0.612	0.453	0.657	0.456	0.567	0.382

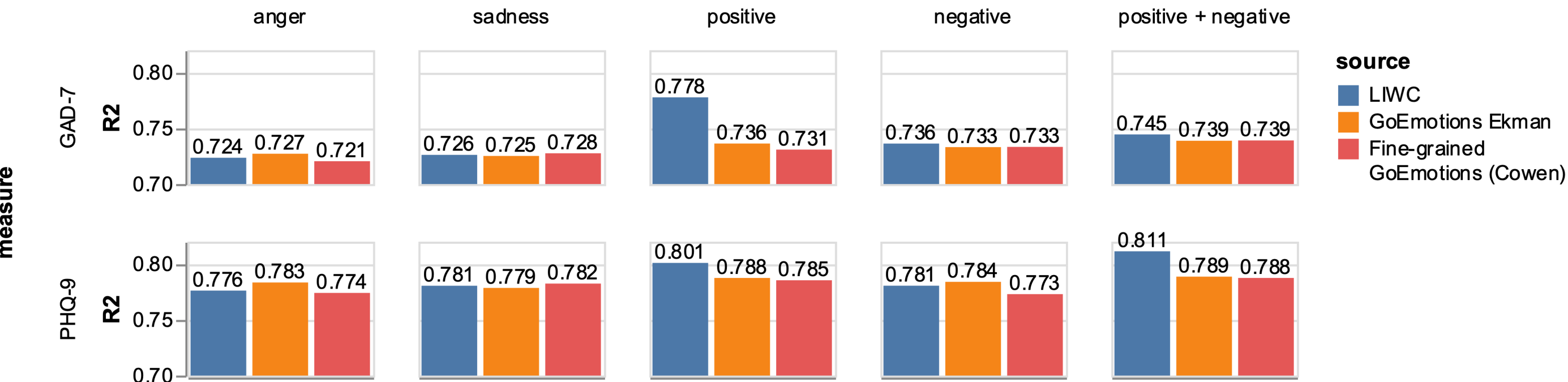
Summary

1. LIWC’s emotion features are as predictive as GoEmotions features → still a good choice.
2. GoEmotions features capture emotional state comprehensively, yielding additional clinically relevant nuance and benefitting predictive performance.

- Limitations: Non-diverse patient sample (79% ≤ 35 y.o., 79% female, 75% BS or higher)
- Future work: Clinical decision support tools. **Interpretability is key:** models based on interpretable emotion features are preferred over black-box models
- Ethics: Monitoring may be considered invasive - informed consent is paramount. Further research & applications must take ethical considerations into account.

Comparable features similarly associated

- Obtainable from all 3 sources: anger, sadness, positive & negative emotion
- Significant associations between these features and the PHQ-9/GAD-7 scores, and comparable predictive power (measured in R2).
- LIWC emotion features performed well, indicating that these features remain a good choice, e.g. if computational constraints preclude NN models.



Acknowledgements & References

This work was supported by the National Library of Medicine (grant number 67-3780) and by Innovation Grant “Informatics-Supported Authorship for Caring Contacts (ISACC)” from the Garvey Institute for Brain Health Solutions. Conflicts of interest: TDH is an employee of the platform that provided the data.

Alan S. Cowen and Dacher Keltner. 2017. [Self-report captures 27 distinct categories of emotion bridged by continuous gradients](#). *Proceedings of the National Academy of Sciences of the United States of America*, 114(38):E7900–E7909, September.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. [GoEmotions: A Dataset of Fine-Grained Emotions](#), May.

Paul Ekman. 1992. [Are There Basic Emotions?](#) *Psychological Review*, 99(3):550–553.

Thomas D. Hull, Matteo Malgaroli, Philippa S. Connolly, Seth Feuerstein, and Naomi M. Simon. 2020. [Two-way messaging therapy for depression and anxiety: longitudinal response trajectories](#). *BMC Psychiatry*, 20(1):297, December.

James W. Pennebaker, R. J. Booth, and M. E. Francis. 2007. Linguistic Inquiry and Word Count: LIWC.

Yla R. Tausczik and James W. Pennebaker. 2010. [The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods](#). *Journal of Language and Social Psychology*, 29(1):24–54, March.

