AI-based analysis of social media language predicts addiction treatment dropout at 90 days

Adi Mukundan

Abstract

- Drug use in terms of addiction treatment and dropout

- Predicting these events using AI models

- Combining with traditional methods to build a digital phenotype of an individual prior to entering treatment

Introduction: ASI

- A broad spectrum of factors are considered before treatment intake to predict severity of an individual's addiction (ASI)
  - Information incomplete in this process
- These factors used to place individuals into ASI Indices

Introduction: AI

- Prevalence of social media lends itself to understanding drug use

- Inspired by previous work involving AI on social media not necessarily for drug use

- Interesting because first time this modern approach (BERT) being used in a clinical setting for understanding substance abuse

- Generally, use a modern deep learning approach to create a digital phenotype representation of users from social media posts
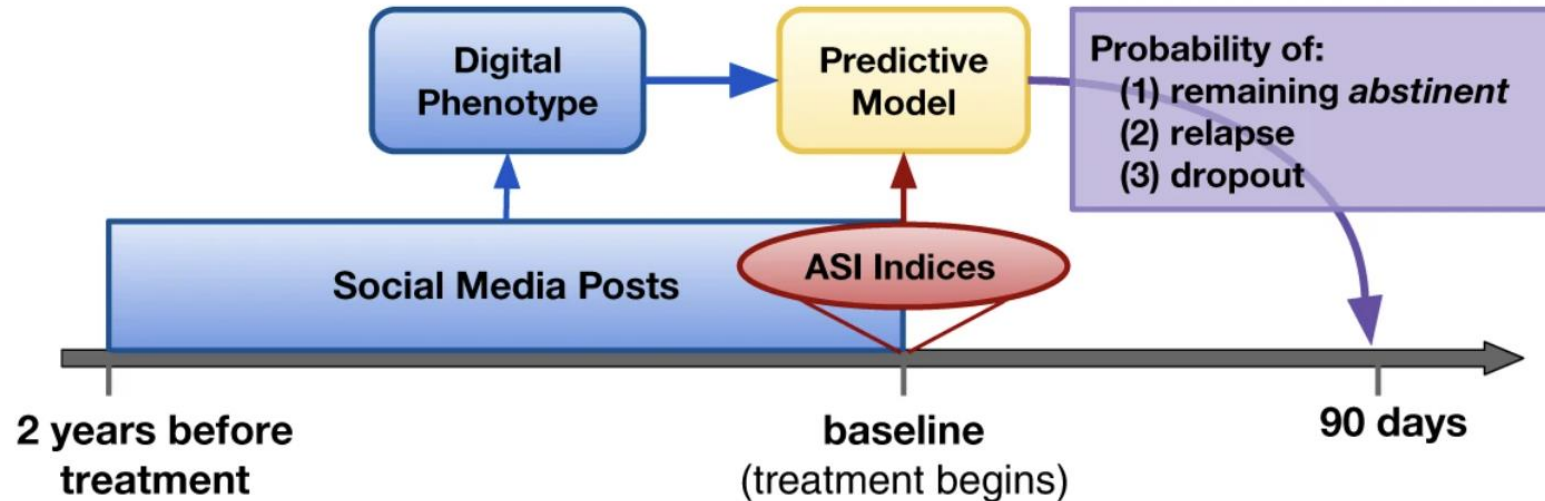
Relation to Class Project

- Both working with social media
- Showing how results could be obtained from social media posts
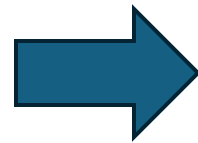
Study Design

- 269 patients had ASI conducted as well as social media posts analyzed
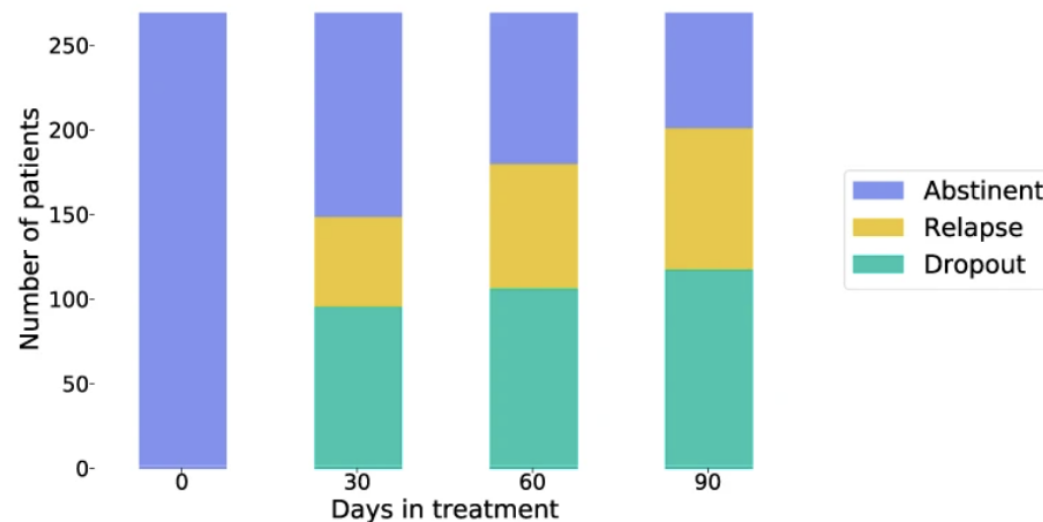- Previous 2 years of social media analyzed to make digital phenotype

Treatment outcomes

- Models used to predict outcomes at different time intervals
  - 30, 60, 90-day outcomes
  - Outcomes: remained abstinent, relapsed, dropped out
  - Alternative Outcomes:
    - 4 category: abstinent, relapse-in, relapse-out, dropped out
    - 2 category: stay in treatment, dropped out

(B)

Facebook Language

- Data from status posts(Status updates) and link posts(free text posted by users)

Table 1. Demographics in the study.

| Demographics	Total sample (N = 504)	Sample with 200-word restrictions			
		Total (N = 269)	Remained abstinent (N = 68)	Relapsed (N = 83)	Dropped out (N = 118)
Age (mean, SD)	33.1 (9.7)	33.2 (9.4)	37.1 (9.8)	32.1 (8.9)	31.7 (8.7)
Sex, % Male	69.3%	63.6%	60.3%	62.7%	66.1%
Race, % Black	59.7%	62.8%	66.2%	72.3%	54.2%
Ethnicity, % Hispanic	13.5%				
Never married	81.2%				
Social media language					
Number of words (mean, SD)	43,395 (7709)	4488 (8850)	4171 (7879)	2863 (4423)	5814 (11,178)
Drug use history					
Drug treatment attempts (mean, SD)	3.8 (4.1)				
Reason entering treatment					
Alcohol	9.9%				
Marijuana	28.0%				
Sedatives	2.2%				
Cocaine/Crack	18.8%				
Stimulants	1.0%				
Hallucinogens	7.7%				
Heroin	22.2%				
Other Opiates	7.3%				
Other Substances	2.8%				

Transformers

Transformers: Background
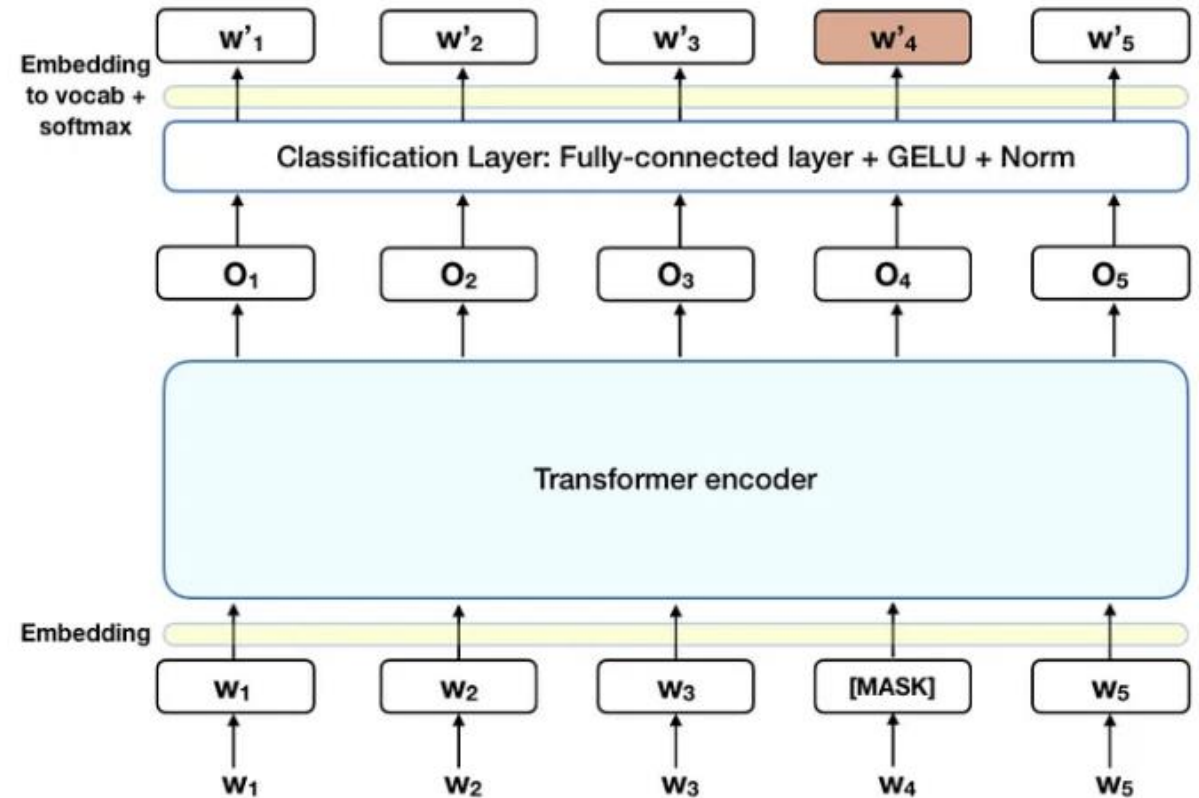
- General overview, Background, RNNS, etc
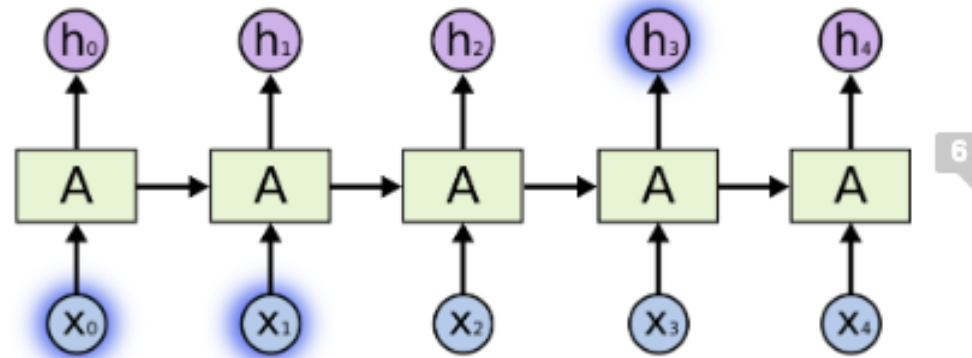
Methods: Transformers

- Encoder, decoder

- Attention
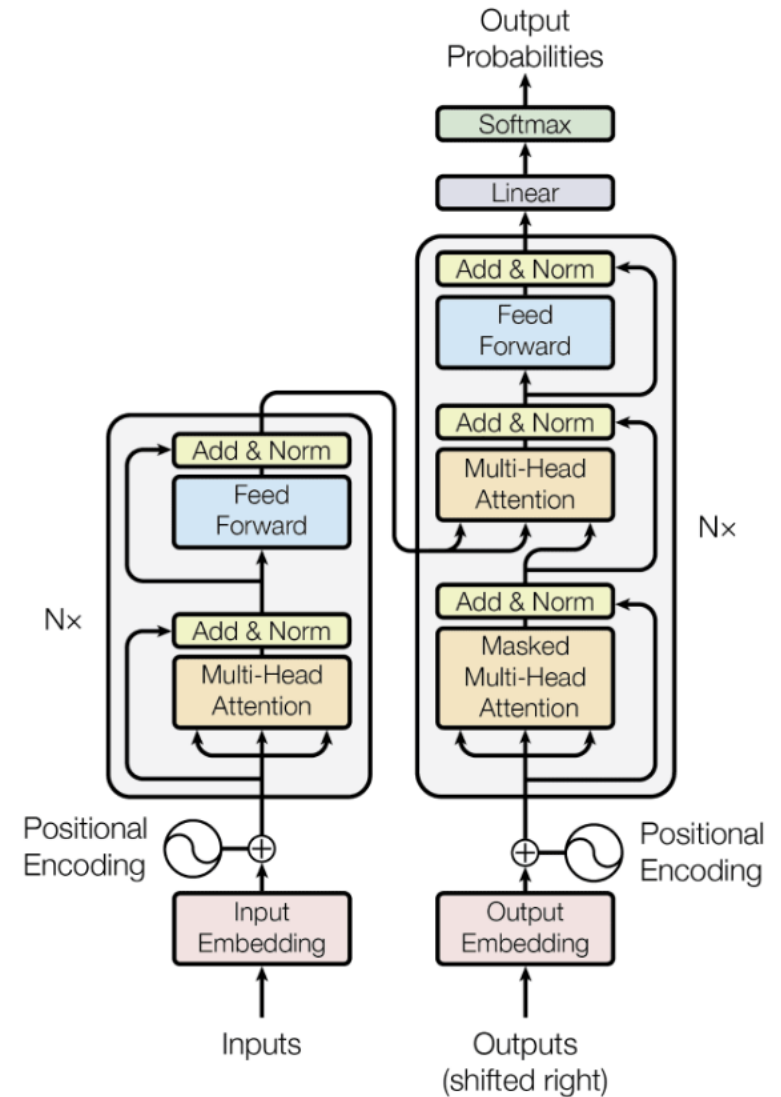
Transformers: BERT

- Pretrained Transformer model to understand

- Tokens from the attention step used bidirectionally

- Pretrained on general unlabeled data -> output layer

Digital Phenotype

- Most of the work already done in BERT pretraining

- 61 features from intake survey, demographic information and text analysis measures from the DLAKT package

Model Training

- 3 outcome model produced using ridge penalized logistic regression

- 4 and 2 outcome models which incorporated ASI and digital phenotype produced using random forest model with extremely randomized trees algorithm

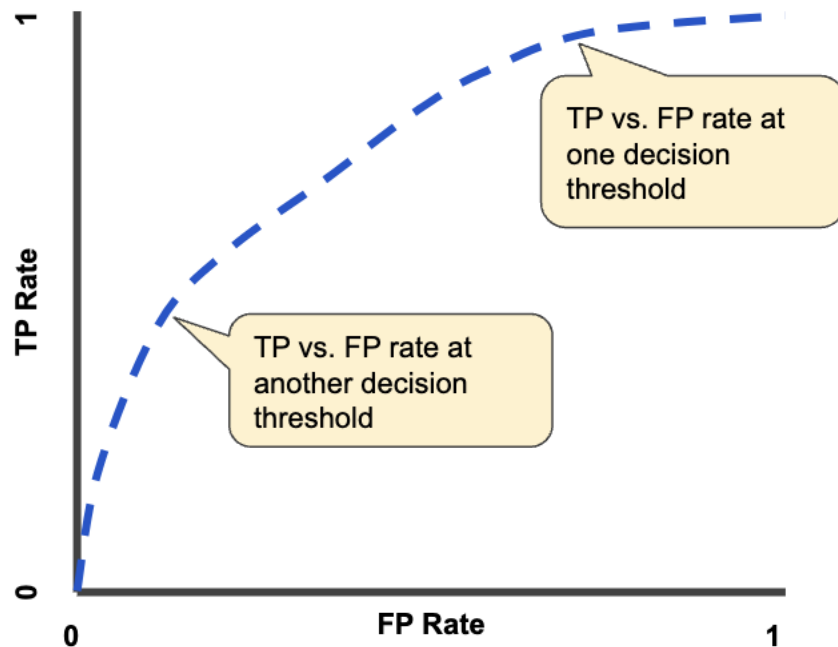- 10-fold cross-validation used to avoid overfitting of models

Classifying model accuracy

$$\text{Precision} = \frac{TP}{TP + FP}$$

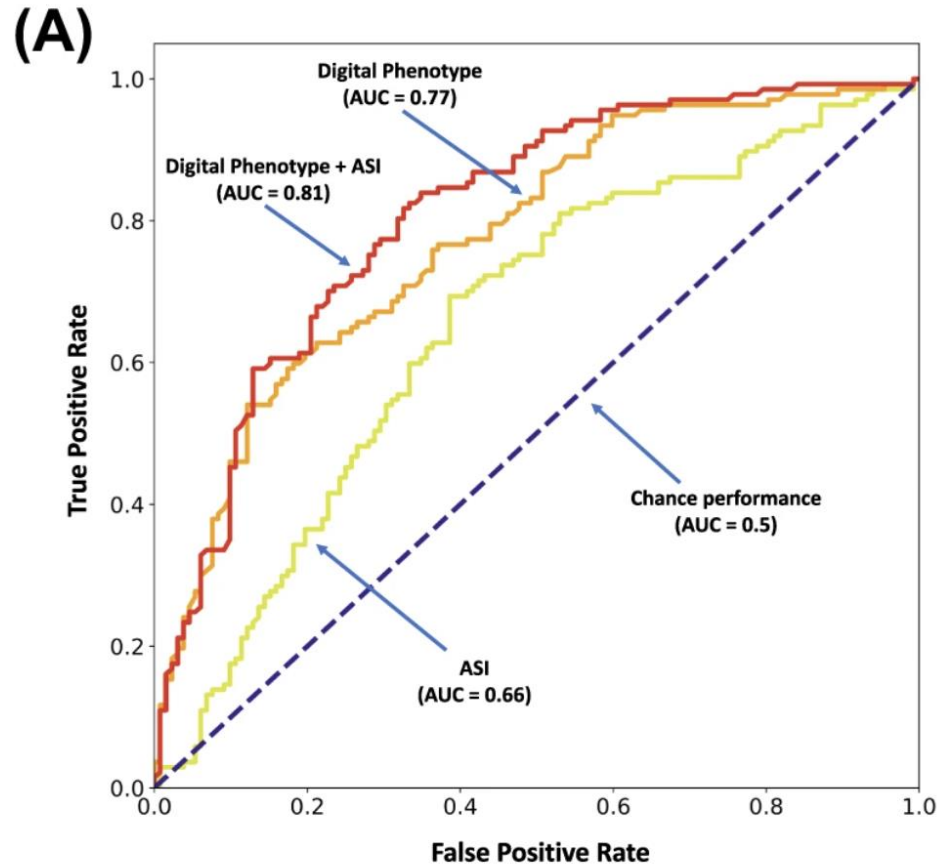$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Classifying model accuracy: AUC ROC

- Look at integral from 0 to 1 of this curve

- Gives performance of model across all possible classification thresholds

Results: Predicting dropout risk

- Both models which include the AI based digital phenotype performed better than the traditional ASI

- Relatively large improvement over solely ASI suggests the value of the digital phenotype in predicting treatment dropout at 90 days
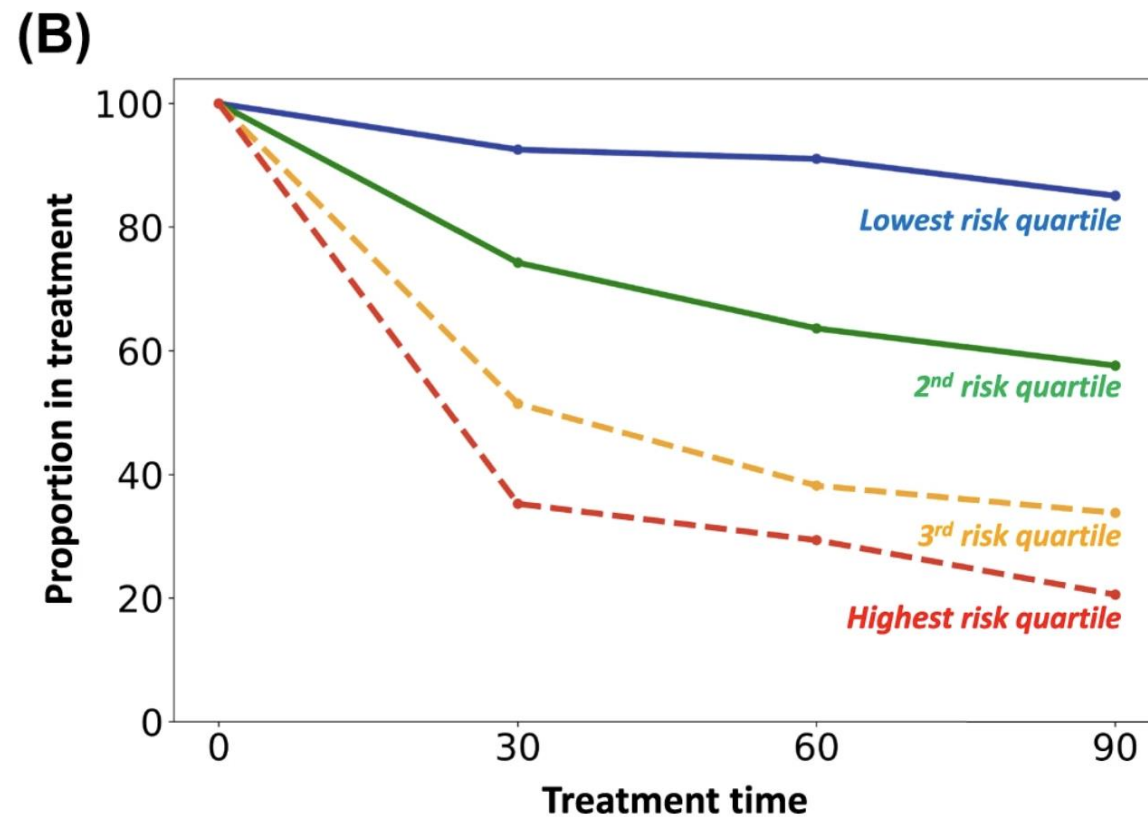
Results: model evaluation

- 3 outcome, 4 and 2 outcome results
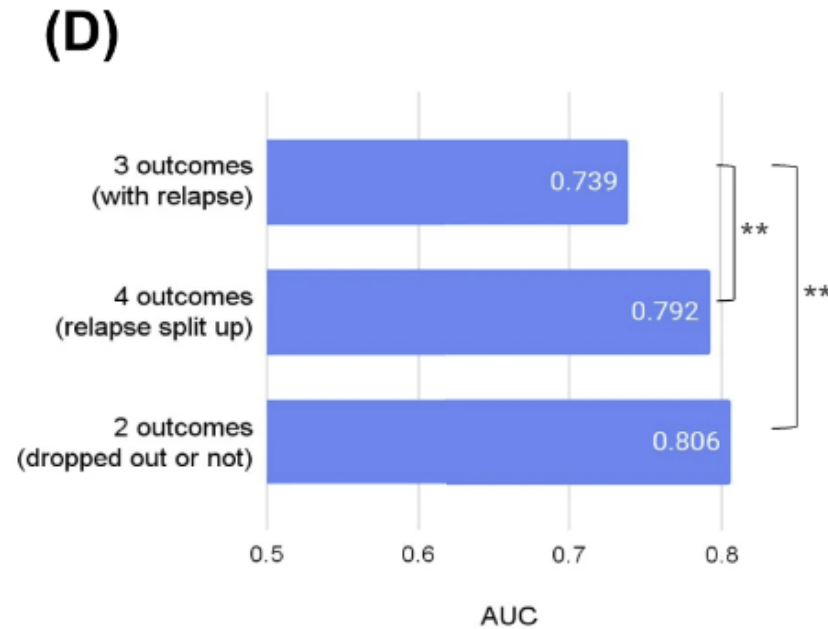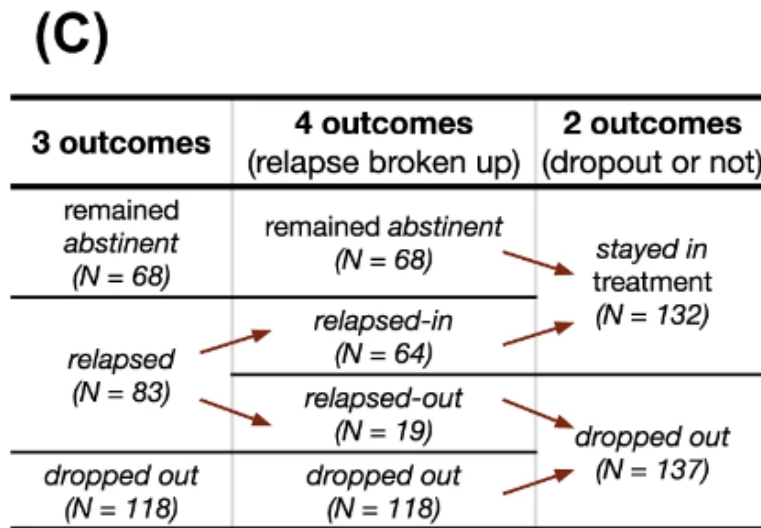
Results: Dropout risk evaluation

- Placing individuals into pretreatment risk quartiles aligns with their eventual categorization into the 4 outcome categories

Discussion: ASI and DP

- Combining ASI and DP provides valuable insights and predictive power for outcomes of individuals in substance use treatment

- Dropout risk and treatment outcomes can be predicted more effectively using the ASI and DP models

Discussion: Limitations

- These models have a very small sample size

- Abstinent and relapse classification requires a patient to self report

- BERT too general

Discussion: Feasibility

- Models may not generalize across wider population

- 70% of adults use social media; still a large source of data

- Reddit now restricts downloads of user and subreddit data, unsure availability of such data

Conclusions

- Predicting treatment outcomes

- Continuous monitoring

- AI identifies new variables of importance

- All this can improve efficiency and accessibility of substance use treatment

Credits, Questions?

Curtis, B., Giorgi, S., Ungar, L. et al. AI-based analysis of social media language predicts addiction treatment dropout at 90 days. Neuropsychopharmacol. 48, 1579–1585 (2023). https://doi.org/10.1038/s41386-023-01585-5

https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

https://towardsdatascience.com/transformers-141e32e69591

https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270

https://arxiv.org/pdf/1810.04805.pdf