



WPI

EMU: Early Mental Health Uncovering Framework and Dataset

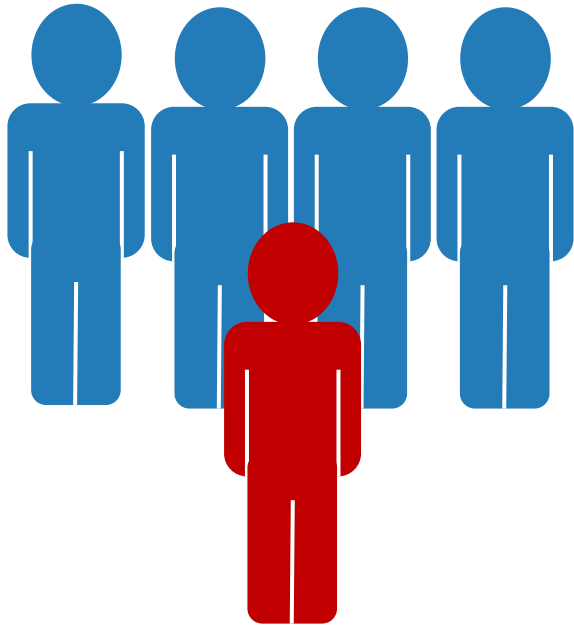
ML Tlachac, Ermal Toto, Joshua Lovering, Rimsha Kayastha, Nina Taurich, Elke Rundensteiner

Worcester Polytechnic Institute

ICMLA 2021 Special Session Machine Learning in Health

Mental Illnesses are Prevalent and Costly

Annually, more than
1 in 5
U.S. **adults** experience
mental illnesses



Depression and anxiety are the most prevalent mental illnesses

- Globally account for one trillion US dollars per year in lost productivity
- Treatment is effective
- There are long delays in diagnosis and treatment

[1] National Alliance on Mental Illness, "Mental health by the numbers," 2020. [Online]. Available: <https://www.nami.org/mhstats>

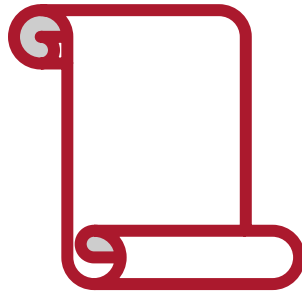
[2] K. Kroenke, R. L. Spitzer, and J. B. Williams, "The phq-9: validity of a brief depression severity measure," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.

[3] R. M. Epstein, P. R. Duberstein, M. D. Feldman et al., "I didn't know what was wrong: How people with undiagnosed depression recognize, name and explain their distress," *Journal of General Internal Medicine*, vol. 25, pp. 954–961, 2010.

Mental Illness Detection



Interview



Survey



App

People are More Willing to Share Active Modalities

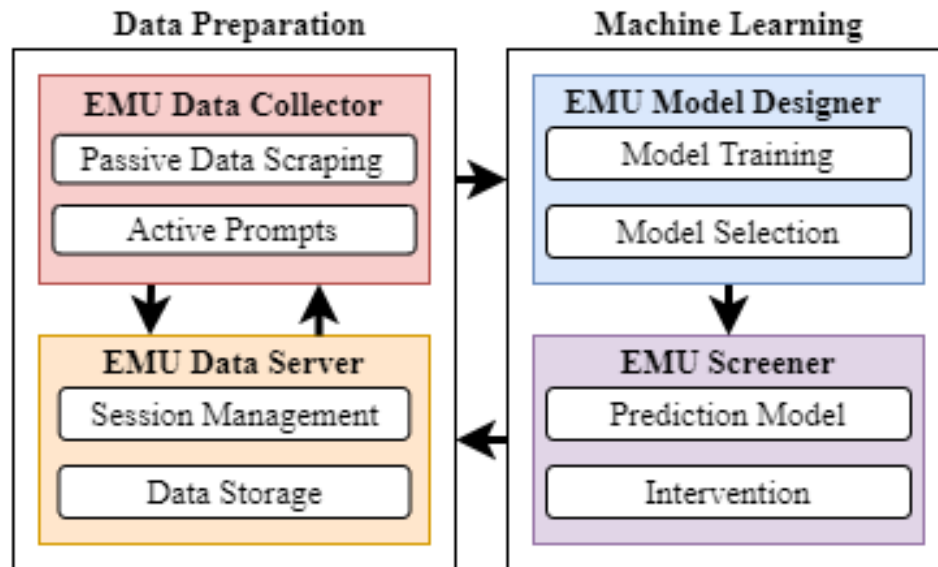
Modality	Willingness to Share
Voice Recordings	65%
Selfies	58%
GPS	48%
Twitter Tweets	44%
Call Logs	43%
Facebook Posts	42%
SMS Text Logs	41%
Browser History	36%

Active Modalities

Passive Modalities

Dogrucu, et al. "Moodable", Smart Health, 2020

Early Mental Health Uncovering (EMU) Framework



EMU seamlessly combines the collection of both active and passive modalities to maximize the data collected for near instantaneous screening of a variety of mental illnesses

EMU Mobile Application

Upload Phone Data

Demographics | Earned: \$0.00

Demographic Information

1. What is your age?

☐ 18-22 ☐ 23-38 ☐ 39-54 ☐ 55-73 ☐ 73+

2. What is your gender?

☐ Male ☐ Female

3. What is your highest level of education?

☐ Less than high school ☐ High school degree or equivalent ☐ Some college, no degree ☐ College degree

4. Are you currently a student?

☐ Yes ☐ No

SUBMIT (\$0.10)

Depression Survey

PHQ9 | Earned: \$0.35

Over the last 2 weeks, how often have you been bothered by any of the following problems?

0 - Not At all
1 - Several Days
2 - More than Half the Days
3 - Nearly Every Day

3. Sleeping too much

☒ 0 ☐ 1 ☐ 2 ☐ 3

4. Feeling tired or having little energy

☐ 0 ☒ 1 ☐ 2 ☐ 3

5. Poor appetite or overeating

☐ 0 ☐ 1 ☒ 2 ☐ 3

6. Feeling bad about yourself - or that you're a failure or have let yourself or your family down

SUBMIT (\$0.10)

Anxiety Survey

GAD7 | Earned: \$0.45

Over the last 2 weeks, how often have you been bothered by the following problems?

0 - Not At all
1 - Several Days
2 - More than Half the Days
3 - Nearly Every Day

1. Feeling nervous, anxious or on edge

☒ 0 ☐ 1 ☐ 2 ☐ 3

2. Not being able to stop or control worrying

☐ 0 ☒ 1 ☐ 2 ☐ 3

3. Worrying too much about different things

☒ 0 ☐ 1 ☐ 2 ☐ 3

SUBMIT (\$0.10)

Scripted Audio

Voice Recording | Earned: \$0.55

Read Prompt

Please press record and read the following sentence:

That which we call a rose by any other word would smell as sweet.

RECORD

Press Record

SUBMIT (\$0.15)

Unscripted Audio

Voice Recording | Earned: \$0.70

Open Ended Prompt

Please press record, then answer the following prompt into the microphone until time runs out:

Describe a good friend.

STOP

15

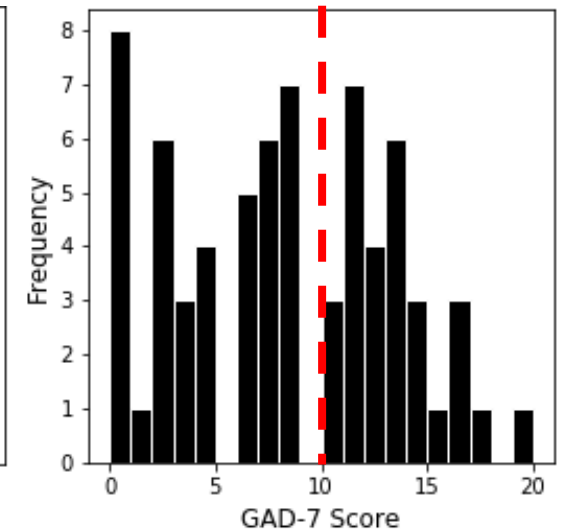
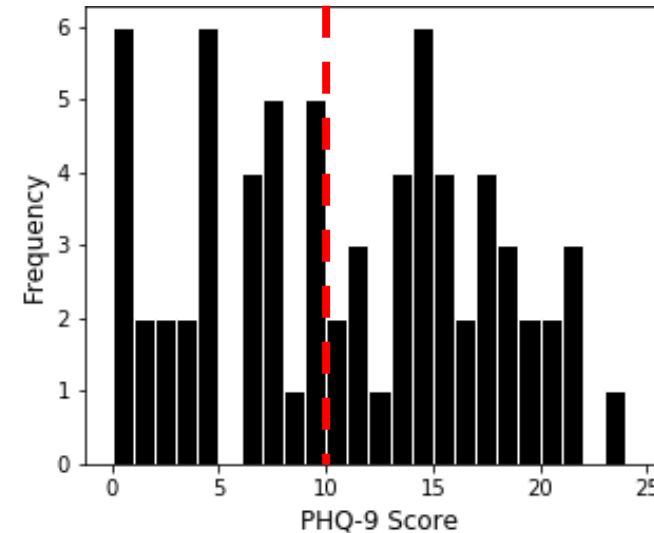
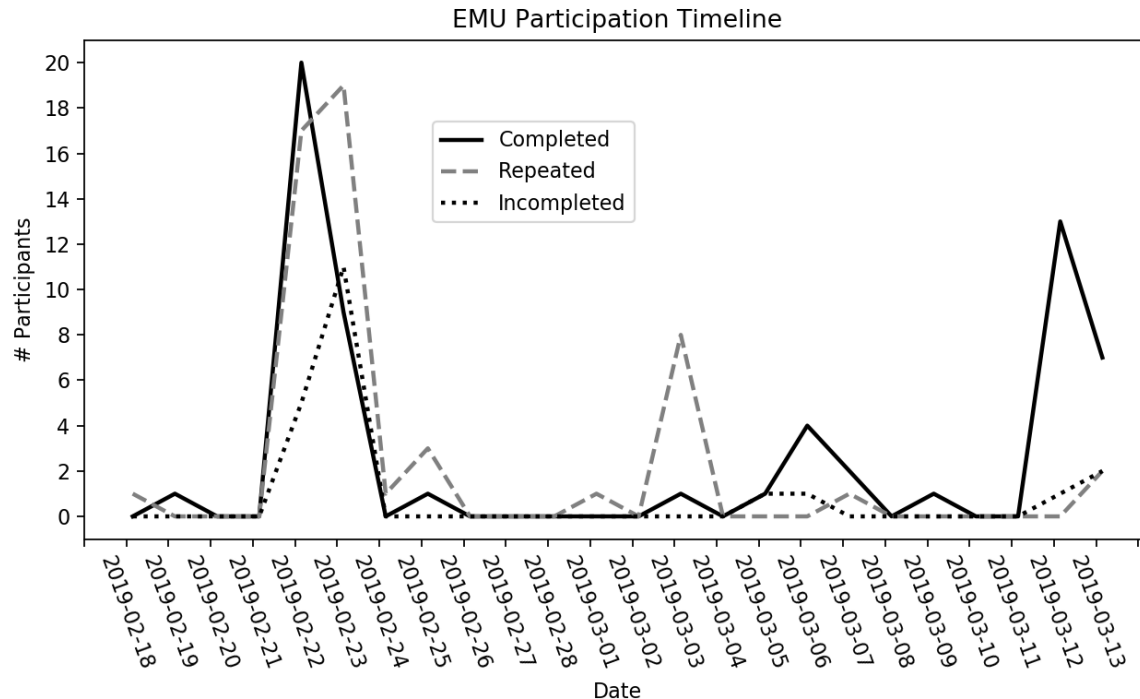
SUBMIT (\$0.20)



Mental illness labels for the data

EMU Dataset Participants

Collected data from Mturk workers for three weeks in early 2019



Most of the **70** participants were 23 to 38-year-old men with college degrees

Modalities Shared by EMU Participants

Modality	Participants	mean PHQ-9	mean GAD-7
PHQ-9	70	10.37	7.77
GAD-7	69	10.43	7.77
Demographic	70	10.37	7.77
Scripted Audio	63	10.08	7.63
Unscripted Audio	55	10.04	7.71
Text Messages	31	10.29	8.20
Calendar	28	10.89	8.18
Call Log	25	10.88	8.04
GPS	14	9.29	7.36
Twitter Username	14	8.50	7.21
Twitter Posts	11	6.91	6.00
Instagram	0	0	0

Screening Results with Audio Recordings

		Audio Recordings			Voice Recordings		
Mental Illness	Audio	<i>F1</i>	<i>AUC</i>	<i>Acc</i>	<i>F1</i>	<i>AUC</i>	<i>Acc</i>
Depression	Scripted	0.746	0.729	0.730	0.746	0.735	0.732
Depression	Unscripted	0.691	0.691	0.691	0.667	0.675	0.673
Anxiety	Scripted	0.655	0.696	0.683	0.667	0.733	0.750
Anxiety	Unscripted	0.625	0.681	0.673	0.636	0.695	0.673
Suicidal Ideation	Scripted	0.706	0.752	0.762	0.648	0.729	0.768
Suicidal Ideation	Unscripted	0.627	0.667	0.655	0.629	0.714	0.735

Leave-one-out evaluation strategy to maximize training data

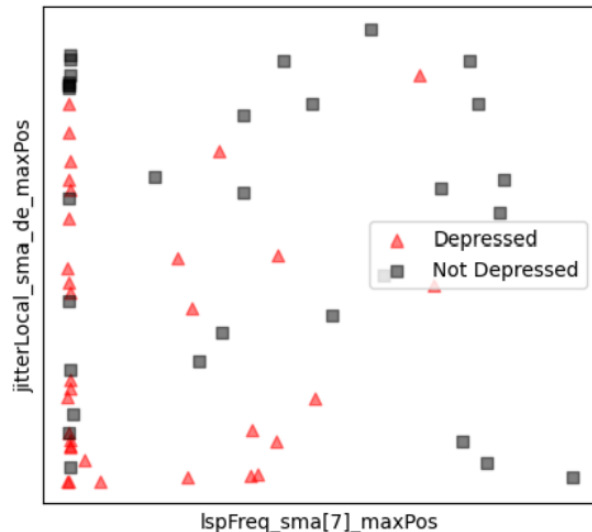
The most effective models were **Support Vector Classifiers**

Screening Results with Transcripts

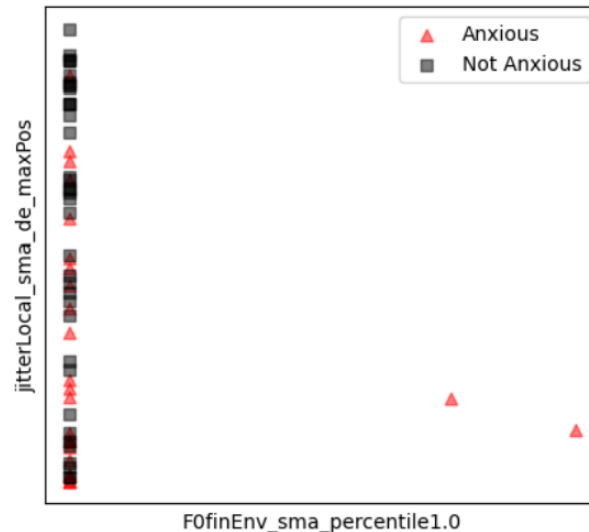
Mental Illness	Data	$F1$	AUC	Acc
Depression	Audio	0.600	0.633	0.628
Depression	Transcript	0.538	0.489	0.442
Depression	Both	0.600	0.633	0.628
Anxiety	Audio	0.600	0.693	0.721
Anxiety	Transcript	0.629	0.706	0.698
Anxiety	Both	0.545	0.639	0.651
Suicidal Ideation	Audio	0.667	0.754	0.767
Suicidal Ideation	Transcript	0.545	0.649	0.651
Suicidal Ideation	Both	0.621	0.718	0.744

Important Screening Features

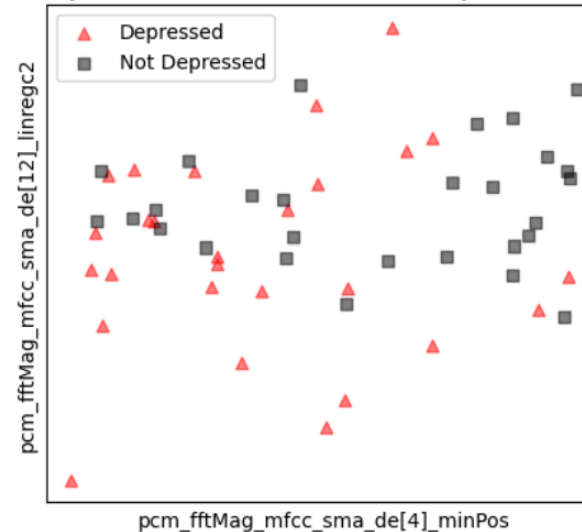
Depression Features for Scripted Audio



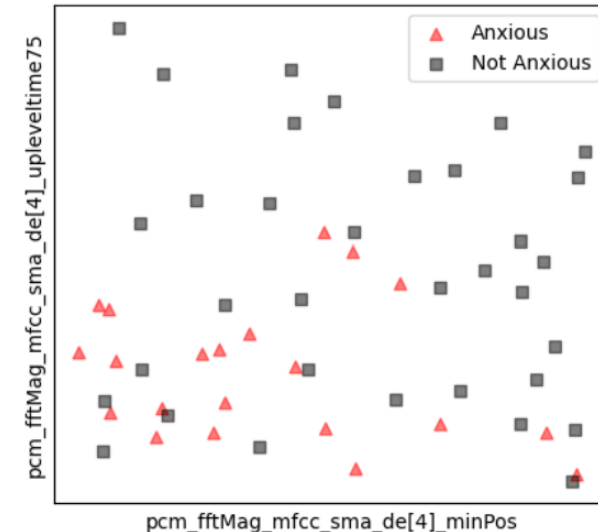
Anxiety Features for Scripted Audio



Depression Features for Unscripted Audio



Anxiety Features for Unscripted Audio



Jitter
is important for
scripted audio

Mel-frequency Cepstral Coefficient
is important for
unscripted audio

Takeaways

- The **EMU framework** for mental illness screening integrates both passive and active modalities
- The **EMU dataset** is a valuable community data resource composed of digital phenotype and audio features labeled with mental illness screening scores: <https://github.com/mltlachac/EMU>
- The **EMU participants** were more willing to share active than passive modalities
- Future work involves combining existing datasets in this domain for multimodal models that can use both active and passive modalities

Thank You. Questions?

This work was supported by:

- US Depart. of Ed. P200A150306 & P200A180088: GAANN grants
- NSF III: Small #1910880
- Data Science department at WPI

