Feature Analysis and Hierarchical Classification of Anxiety Severity during early COVID-19

Binh Nguyen, Michael Nigro, Alice Rueda, Sharadha Kolappan, Venkat Bhat, and Sridhar Krishnan

Presentation for 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2021)

Ryerson University

St. Michael's
Inspired Care.
Inspiring Science.

Problem Statement

- Mental health is among the greatest cross-national inequalities with 80% of those affected living in low- and middle- income countries [1]
- Mental health can be affected by internal factors such as a person's physical health and genetic predisposition as well as external factors such as financial insecurity, food insecurity, and lifestyle changes [2], [3]
- Statistics Canada created the Canadian Perspectives Survey Series (CPSS) to understand social issues and effects of COVID-19 on the Canadian labour force (LF)

Objectives

- The motivation for this work is to identify characteristics from the CPSS [5] data that are indicative of anxiety for the Canadian LF population
- We hypothesize that successful identification of survey data characteristics will allow for indirect assessment of anxiety
- We propose to use the more quantified General Anxiety Disorder-7 labels to look at anxiety among the public during the pandemic

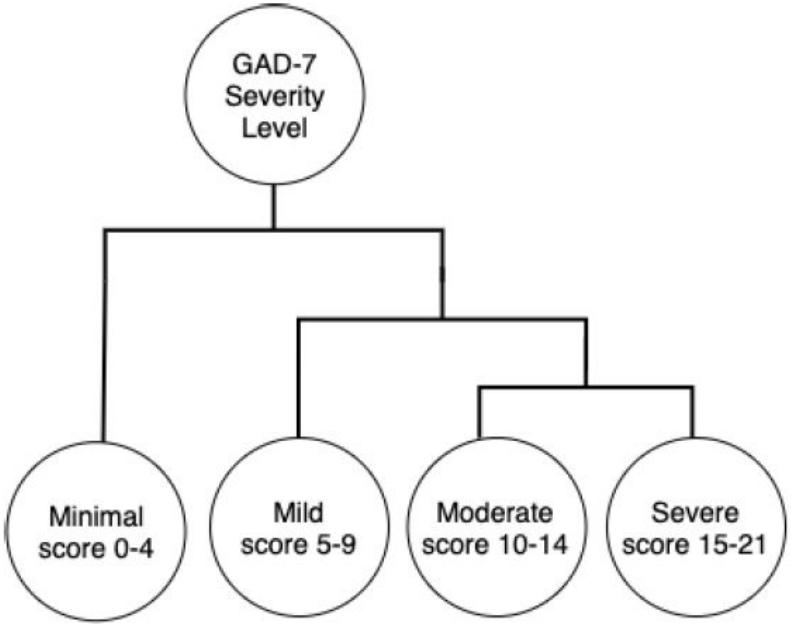
Related works

- A few studies have been conducted on CPSS using perceived mental health categories [4], [6]
  - Perceived mental health labels: Excellent, Very Good, Good, Fair, and Poor.
- Polsky and Gilmour compared food insecurity among Canadians during the COVID-19 pandemic [3]

Dataset

- CPSS2 [5]
  - Collected May 4 - May 10, 2020
  - The target population are Canadians aged 15 and older in the LF with the exception of full-time members of the Canadian Armed Forces
  - CPSS2 has 4600 sample points and 62 variables
  - Variables are categorized into
    - Behaviours and Health impacts (BH)
    - Demographics (DEM)
    - Derived Variables (DV)
    - Food security (FSC)
    - Labour market impacts (LM)
    - Mental health impacts (MH)
    - Survey related variables (SRV)

GAD-7


GAD-7
Severity
Level
Minimal
score 0-4
Mild
score 5-9
Moderate
score 10-14
Severe
score 15-21

Pre-processing

- Removal of data related to GAD-7
- Removal of Incomplete data

2	MH_15A	Feeling nervous, anxious or on edge - Freq last 2 weeks	discrete	numeric-1.0	4579	21	Over the last 2 weeks, how often have you been bothered by the following problems? - Feeling nervous, anxious or on edge
3	MH_15B	Not being able to stop or control worrying - Freq last 2 weeks	discrete	numeric-1.0	4563	37	Over the last 2 weeks, how often have you been bothered by the following problems? - Not being able to stop or control worrying
4	MH_15C	Worrying too much about different things - Freq last 2 weeks	discrete	numeric-1.0	4561	39	Over the last 2 weeks, how often have you been bothered by the following problems? - Worrying too much about different things
5	MH_15D	Trouble relaxing - Freq last 2 weeks	discrete	numeric-1.0	4566	34	Over the last 2 weeks, how often have you been bothered by the following problems? - Trouble relaxing
6	MH_15E	Being so restless that it is hard to sit still - Freq last 2 weeks	discrete	numeric-1.0	4558	42	Over the last 2 weeks, how often have you been bothered by the following problems? - Being so restless that it is hard to sit still
7	MH_15F	Becoming easily annoyed or irritable - Freq last 2 weeks	discrete	numeric-1.0	4568	32	Over the last 2 weeks, how often have you been bothered by the following problems? - Becoming easily annoyed or irritable
8	MH_15G	Feeling afraid as if something awful might happen - Freq last 2 weeks	discrete	numeric-1.0	4570	30	Over the last 2 weeks, how often have you been bothered by the following problems? - Feeling afraid as if something awful might happen
3	ANXDVGAC	Generalized Anxiety Disorder Cut-point	discrete	numeric-1.0	4512	88	-
1	ANXDVGAD	Generalized Anxiety Severity Score	continuous	numeric-2.0	4512	88	-

Feature ranking

- Minimum redundancy maximum relevance (mRMR) was selected

$$\max \Phi(D, R), \Phi = D - R$$

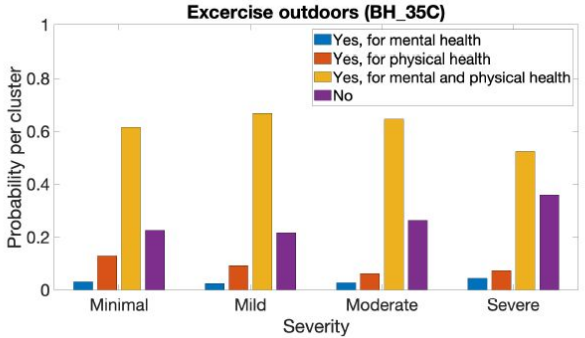
$$\max D(S, c), D = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c))$$

$$\min R(S), R = \frac{1}{|s|^2} \sum_{x_i, x_j \in S} I(x_i; x_j)$$

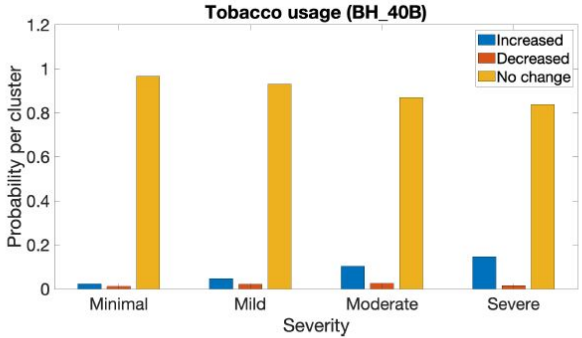
Results

- After pre-processing, 4512 samples remained
- mRMR reduced the feature set to 20

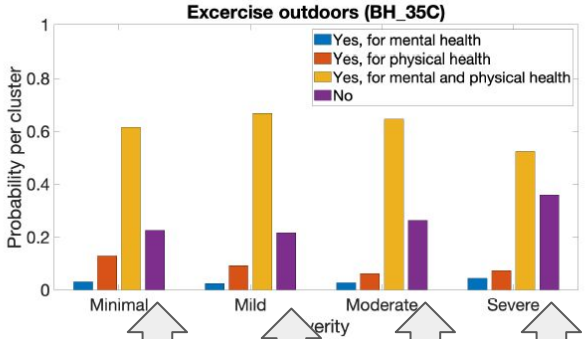
Feature	Description	Feature	Description
MH_05	Perceived mental health	LM_40	COVID impact ability meet financial obligations
BH_40D	Eating junk food or sweets	BH_20C	Made plan caring household member are ill
PFSCDV	Household food insecurity	BH_40F	Spending time on the internet
AGEGRP	Age group	SEX	Sex
MHDVMHI	Perceived mental health derived variable	BH_20M	Other precautions taken to reduce risk
BH_20A	Stocked up on essentials	BH_35C	Exercise outdoors
LM35BCDE	EI benefits (sickness/ caregiver/ worksharing/ other)	BH_40A	Consuming alcohol
RURURB	Rural or urban indicator	BH_110	Number of people in close contact
BH_40E	Watching TV	BH_20D	Made a plan for non-household member
BH_35E	Changing food choices	BH_40B	Using tobacco products



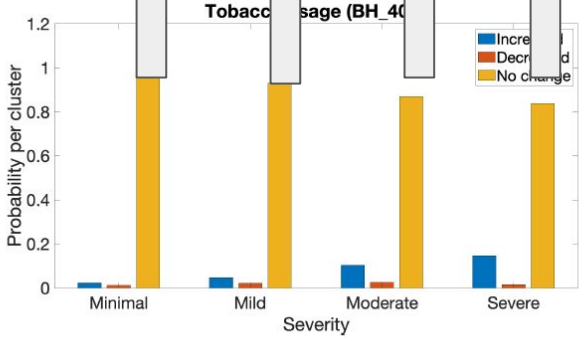
(a) Exercise outdoors



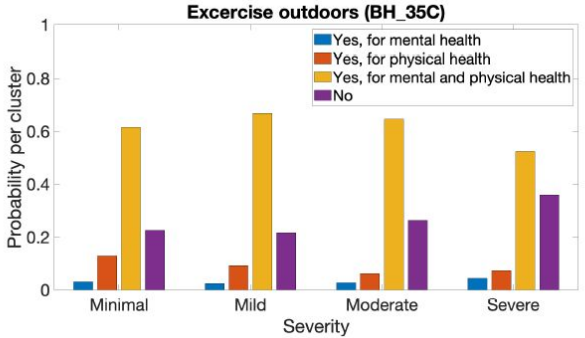
(b) Tobacco Usage



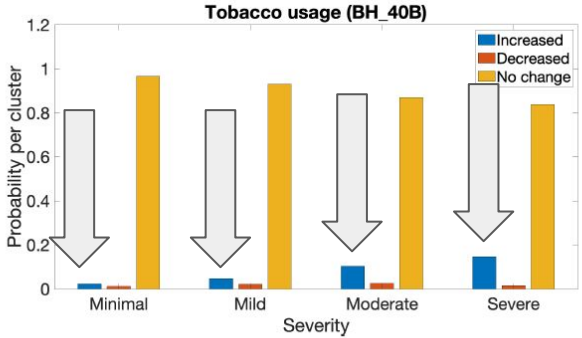
(a) Exer e outdoors



(b) Tobacco Usage



(a) Exercise outdoors

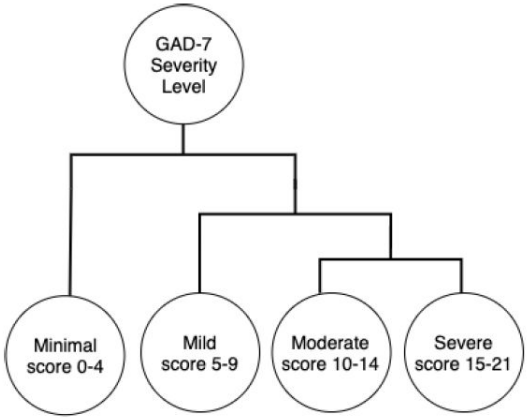


(b) Tobacco Usage

Results continue

- Minimal vs Severe GAD Severities
  - Decision tree: 92.03±0.30%
  - Support vector machine: 94.77±0.05%
    - Recall: 98.62±0.09%,
    - Precision: 95.72±0.05%
    - F1: 97.15±0.05%
- Hierarchical Classification

Hierarchical Classification



Classes	SVM(%)	DT(%)
Minimal vs Mild, Moderate, and Severe	76.99	68.79
Mild vs Moderate and Severe	71.05	62.64
Moderate vs Severe	63.94	57.52

Discussion and Conclusion

- The purpose of this work is to analyze the anxiety of the LF during the preliminary stages of COVID-19 using the CPSS2 dataset
- The application used the reduced 20 feature subset and 10-fold SVM to achieve an overall accuracy of 94.77±0.05% when comparing the classes Minimal and Severe
- Many of these features can be augmented as Pseudo Passive (PP) data. PP data is qualitative data that can be collected as passive data
  - RURURB (Rural or urban indicator) can use a GPS to determine a participant's location
  - BH 35C (Exercise outdoors) can use an accelerometer for activity recognition
  - BH 40E (Watching TV) can determine the audio environment to determine if the participant is watching TV
- Reduction of 62 survey questions to 20
  - Reduces survey fatigue, patients becoming apathetic or bored
- The study of continuous long-term data collection can further explore and understand how people cope during the COVID-19 pandemic [8,9]

References

[1] K. S. Jacob and V. Patel, “Classification of mental disorders: A global MH perspective,” The Lancet, vol. 383, pp. 1433–1435, 2014.
[2] L. Martinengo, L. Van Galen, E. Lum, M. Kowalski, M. Subramaniam, and J. Car, “Suicide prevention and depression apps’ suicide RA and management: a systematic assessment of adherence to clinical guidelines,” BMC Medicine, vol. 17, p. 231, dec 2019.
[3] J. Y. Polsky and H. Gilmour, “Food insecurity and MH during the COVID-19 pandemic.,” Health reports, vol. 31, no. 12, pp. 3–11, 2020.
[4] L. C. Findlay, R. Arim, and D. Kohen, “Understanding the Perceived Mental Health of Canadians During the COVID-19 Pandemic.,” Health reports, vol. 31, pp. 22–27, jun 2020.
[6] Statistics Canada, “Canadian Perspectives Survey Series 2: Monitoring the effects of COVID-19,” 2020.
[7] A. Zajacova, A. Jehn, M. Stackhouse, K. H. Choi, P. Denice, M. Haan, and H. Ramos, “MH and economic concerns from March to May during the COVID-19 pandemic in Canada: Insights from an analysis of repeated cross-sectional surveys,” SSM - Population Health, vol. 12, p. 100704, Dec 2020.
[8] B. Nguyen, Y. Coelho, T. Bastos, and S. Krishnan, “Trends in HAR with focus on machine learning and power requirements,” Machine Learning with Applications, vol. 5, p. 100072, sep 2021.
[9] S. Krishnan, “Biomedical signal analysis for connected healthcare,” Elsevier.