Explainable-Machine-Learning to discover drivers and to predict mental illness during COVID-19

Indra Prakash Jha[1,*], Raghav Awasthi[1,*], Ajit Kumar[2,3], Vibhor Kumar[1,#], Tavpritesh Sethi[1,#]

1 Indraprastha Institute of Information Technology, Delhi (IIIT-D)
2 XLRI - Xavier School of Management - Jamshedpur
3 Adobe Inc.
 # corresponding author
** authors contributed equally to this work*

Abstract

COVID-19 pandemic has deeply affected the global economy, education, and travel and stranded people in their respective zones. Most importantly, it has claimed more than half a million lives and is poised to create long and short term consequences through mental health impairment. Identification of important factors affecting mental health can help people manage emotional, psychological, and social well-being. Here, we focus on identifying factors that have a significant impact on mental health during COVID pandemics. In this study, We have used a survey of 17764 adults in the USA at different age groups, genders, and socioeconomic statuses. Through initial statistical analysis followed by Bayesian Network inference, we have identified key factors affecting Mental health during the COVID pandemic. Integrating Bayesian networks with classical machine learning approaches lead to effective modeling of the level of mental health. Overall, females are more stressed than males, and people of age-group 18-29 are more vulnerable to anxiety than other age groups. Using the Bayesian Network Model, we found that people with the chronic medical condition of mental illness are more prone to mental disorders during COVID age. The new realities of working from home, home-schooling, and lack of communication with family/friends/neighbors induces mental pressure. Financial assistance from social security helps in reducing mental stress during COVID generated economic crises. Finally, using supervised ML models, we predicted the most mentally vulnerable people with ~80% accuracy.

Keywords: Covid-19, Mental health , Bayesian Network , Machine learning

Introduction

As the coronavirus pandemic rapidly sweeps across the world, it is inducing a considerable degree of fear, worry, and concern in the population at large and among certain groups in particular, such as youth, frontline workers, caregivers, and people with chronic medical conditions. In public mental health terms, the main psychological impact to date is elevated rates of stress or anxiety. But as new measures and impacts are introduced – especially people working remotely, quarantine and its effects on many people's usual activities, routines or livelihoods – levels of loneliness, depression, harmful alcohol, and drug use, and self-harm or suicidal behavior are also expected to rise. Emergencies are always stressful, but specific stressors particular to the COVID-19 outbreak affect the population. Stressors include risk of being infected and infecting others, especially if the transmission mode of

COVID-19 is not 100% clear, Common symptoms of other health problems (e.g. a fever) can be mistaken for COVID-19 and lead to fear of being infected, Caregivers may feel increasingly worried about their children being at home alone (due to school closures) without appropriate care and support. School closures may have a differential effect on women, who provide most of the informal care within families, with the consequences of limiting their work and economic opportunities. Risk deterioration of physical and mental health of vulnerable individuals, for example, older adults and people with disabilities, if caregivers are placed in quarantine if other care and support are not in place. The Lancet Psychiatry[1] highlights the needs of vulnerable groups, including those with severe mental illness, learning difficulties, and neurodevelopmental disorders, as well as socially excluded groups such as prisoners, the homeless, and refugees. But further pressure from clinicians, researchers, service user advocacy organizations, and journalists is needed if the societal discourse is to be shifted. The mental health effects of COVID-19 on the general population might be profound and long-lasting and deserve serious attention. Early career psychiatrists are crucial in the medical response to COVID-19 [2,3].Telepsychiatry (ie, providing mental health care remotely, using telecommunications such as telephone or video conferencing tools) in several settings is suddenly being introduced or massively expanded to serve patients with pre-existing disorders, health professionals on the frontline, and the general population, during a time of uncertainty, misinformation, and physical distancing. Frontline medical workers are at risk of not just adverse physical outcomes from coronavirus disease 2019 (COVID-19) but psychological ones too[4]. Public health emergencies may affect the health, safety, and well-being of both individuals (causing, for example, insecurity, confusion, emotional isolation, and stigma) and communities (owing to economic loss, work and school closures, inadequate resources for medical response, and deficient distribution of necessities). These effects may translate into a range of emotional reactions (such as distress or psychiatric conditions), unhealthy behaviors (such as excessive substance use), and non-compliance with public health directives (such as home confinement and vaccination) in people who contract the disease and in the general population. Extensive research in disaster mental health has established that emotional distress is ubiquitous in affected populations — a finding certain to be echoed in populations affected by the Covid-19 pandemic[5] . Social media is also playing a crucial role Huge citizens are exposed to social media during a novel coronavirus disease (COVID-19) outbreak in Wuhan, China. We assess the prevalence of mental health problems and examine their association with social media exposure[6] .

In this study we used survey data conducted at different places in the USA, using a Probabilistic Graphical model we identified the most influential variable for Mental illness and quantified their Impact. and finally using these influential variables we built the supervised prediction model to predict the vulnerable people in advance.

Methods

Datasets

We extracted survey data of 17764 Adult people (https://www.covid-impact.org/)[7]. The survey is designed to provide weekly estimates of the U.S. adult household population nationwide and for 18 regional areas including 10 states (CA, CO, FL, LA, MN, MO, MT1, NY, OR, TX) and 8 Metropolitan Statistical Areas (Atlanta, Baltimore, Birmingham, Chicago, Cleveland, Columbus, Phoenix, Pittsburgh). Data collection is occurring over a week-long period. Interviews are conducted in English and Spanish. The dataset comprised variables on physical health, mental health, Insurance related Policy, economic security, and social dynamics. We performed the preprocessing over the data. [Table-1] is representing the socio-demographics of respondents participating in the Survey.

Analysis

We have performed a multi-step statistical analysis of the data [Figure 1(a)]. For this, firstly we clubbed questions that are similar in nature in survey data [Supplementary Table 1] and created the new indicators, such as *communication_indicators*, *behavioural_indicators*, *mental_health_Indicators,* etc. [Table-2]. In our analysis, We considered mental_health_indicators as dependent variables, and others as independent variables.

Item Reliability Analysis

Firstly, we performed Item Reliability Analysis [8] to check consistency in the survey questions' variables. For every mental health question (*soc5a, soc5b, soc5c, soc5d, soc5e*), we calculated the **Cronbach's alpha** value, which measures internal consistency and the quantification of the closeness of a set of items as a group. It is considered to be a measure of scale reliability. **Cronbach's alpha** value higher than 0.70 is regarded as a significant measure. The analysis was performed in the R package [9].

Test of independence among features

Next, to examine the dependency between mental_health_indiactors and other variables, we performed a pairwise chi-square test of independence. This chi-square test quantifies the correlation among variables. A p-value of less than 0.05 is regarded as a significant correlation.

Bayesian Network Analysis

Since Chi-square may be biased by spurious relationships and confounding, conditional associations were learned through a structure learning algorithm that learns a Bayesian probabilistic graphical model (PGM) from data. Hill climbing optimizer[10] with Bayesian Information Criterion(BIC)[11] based score was used to select the best PGM that explained the data. Robustness of associations was evaluated through bootstrapped learning and majority voting over 101 PGMs. Exact Inference using the belief propagation algorithm[12] was learned in order to quantify the strength of learned associations. The Analysis was performed in the R package [13] [14] [15]

Mental stress prediction using supervised Machine Learning

From Bootstrapped Network, we calculated the *Markov blanket* of *mental_health_indicators*, and then, using these Markov variables as a predictor, We built the supervised Machine learning model which can predict mental health in advance. For the same we divided the data into training and testing in an 80-20 ratio, and removed the class imbalance in the training set, we performed Oversampling using the Smote algorithm [16]in the training set. And further, we evaluated the performance of the model in the test dataset. We adopted SKlearn [17]Python Package for implementing ML models.

Results

We first performed item reliability analysis (IRA) on all mental-health indicators (*soc5a, soc5b, soc5c, soc5d, soc5e*). We discovered that the Cronbach's alpha value for 4 mental-health indicators is approximately 0.8 [Figure 1(b)], which confirms their consistency. However Cronbach's alpha value on the mental-health indicator, soc5e is 0.58.

We further analyzed the distribution of frequency of different categories for mental-health indicators. The trend for the primary dependent variable - soc5a, revealed two crucial observations. First, females are more stressed than males (Figure 2a,2b), and the youth (age-group 18-29) are more vulnerable to mental health than other age-groups. Second, the most vulnerable group (age-group 18-29 and mental stress >5 days in a week) is roughly the same for both genders- males and females (1.2% and 1.5%, respectively).

Using the Chi-Square test of independence, we attempted to show the significant correlation (dependencies) (Supplementary Figure 1) among different features and mental-health indicators. However, there could be indirect relationships and confounding factors, affecting the correlation (dependencies) among features. Hence we further performed Bayesian network inference to get a clear view of dependencies among different variables.

Predictors of anxiety identified by Bayesian Network analysis

Bayesian network analysis revealed several dependencies which could become topics of further investigation. However, we focused on mental health indicators. We found that in the learned Bayesian network, *SOC5A* (Felt nervous, anxious, or on edge in last seven days) is the driver variable for other mental health indicators (*soc5b:* Felt depressed; *soc5c*: Felt lonely; *soc5d*: Felt hopeless about the future; *soc5e*: sweating, trouble breathing, pounding heart, etc, in the last seven days) [Figure 3(a)]. This means that if we know the status of *soc5a*, then we can infer other mental health indicators with high probability. Hence, in this study, we regarded *soc5a* as the primary dependent variable.

The work_from_home indicators investigate the effects of remote working/schooling on people's mental health. The rise in the conditional probability of all work_from_home indicators (~4%, ~7%, ~5%, ~7% for *phys2_4*, *phy2_5*, *phys2_9*, *phys2_10* respectively) suggests this indicator as a significant stressor.

The deep fall of the inferred conditional probability of two features (>7%, >5% for soc2a, soc2b, respectively) and flat fall of the other two features (1.5%, 1% for *soc3a*, *soc3b*, respectively) suggest the fundamental effect of communication on mental stress. One-to-one or group communication seems to reduce mental stress; however, virtual communication does not have similar effects.

We also investigated the relationship between mental stress and covid symptoms indicators. World health organization (WHO) has recommended contacting the health-service-provider if individual encounters any [18]symptoms from *covid_symptoms indicators* (*from phys1a to phys1q*) in the last seven days, but interestingly, we discovered no significant association between this indicator and mental health (inferred probability difference ~0%, from 62.2% to 62.2%).

Similarly, WHO also recommends chronic medical conditions *(from phys3a to phys3m)* that increase the risk of serious illness from COVID-19. However, inferred conditional probabilities revealed that no disease is associated with mental health other than cancer (*phys3k*), High blood pressure or hypertension (*phys3b*), and mental health (*phys3h*) (inferred probability difference is -7.1, 33.4, -8.2 respectively). Most interestingly, the inferred conditional probabilities were such that cancer and High blood

pressure/hypertension appeared to be reducing mental stress during COVID-19. The possible reason for this is as follows- due to lock-down or remote distancing, people have to stay at home, which might cause for reducing mental stress for people suffering from a chronic medical condition like cancer, high-blood-pressure or hypertension.
Among all *issurance_assistance_indicators*, only *econ6d* and *econ6e* cause reduced mental stress (10.4% and 5.5%, respectively). Old age people found health insurance more relaxing than young age people. COVID has also severely affected the financial condition of individuals. That may also lead to mental stress. Our study found that assistance or income from social security (econ6d) or supplemental social security (econ6e) may reduce this stress significantly (>10% and >5%).
The behavioral changes are crucial to reducing the COVID-19 spread rate, and people are also affirming this. Overall, 83% of total people chose to wear the mask, 77% avoided restaurants, 83% avoided public and crowded places. Nevertheless, our analysis revealed no significant association (inferred probability difference is <5%) between these indicators and mental health.

Predictability of anxiety

For further investigation, we used the supervised machine learning approach on the most predictive variable. For any node, we decided the most predictive variable as those which lie in its Markov blanket in the learned Bayesian network. Markov blanket for mental_health_indicator includes four variable *age4, phys7_4, phys2_18, phys3h*.

In order to study well, we considered three prediction scenarios -

1. Mental issues *less one day* in a week (class 1) vs. Mental issues *more than **one** day* in a week (class 0)
2. Mental issues *less one day* in a week (class 1) vs. Mental issues *more than **three** days* in a week (class 0)
3. Mental issues *less one day* in a week (class 1) vs. Mental issues *more than **five** days* in a week (class 0)

We observed linearly decay (Accuracy 0.80 to 0.64) in model predictability when we move from high risk of depression (case-3) to low risk of depression (case-1) [figure4].

Discussion

Mental Health is a serious public health concern, mood disorders, and suicide-related outcomes have increased significantly over the last decade among all age groups and genders[19,20]. The rapid spread of coronavirus infection forced the Government to close public gathering places, schools, colleges, Restaurants, and Industries. Every individual is isolated; students are taking their classes online; employees are working from Home. Industries are either closed or half-open. These situations are frightful and make people anxious, Stressed, and Depressed. In this study, we used a probabilistic graphical approach to discover interventions that could have a significant impact on keeping a healthy mind. Our study revealed that individuals suffering from the chronic medical condition of mental illness are the most vulnerable for mental illness during the COVID-19 phase, which recommends building national-level policies to regularly track their mental status and treat them accordingly. This study also recommends that people have regular one-to-one or group communication with family and neighbors, instead of communicating using phone, text, email, app, or using the Internet, during lockdown to reduce mental stress. Hence the distance socializing does not appear to be an alternative solution in the presence of social distancing. Thus our results clearly indicate a fundamental difference in the effect of two types of socializing. While the effect of online and telephonic socializing did not seem to have an effect on anxiety, but real one-to-one or group communications have.

The study recommends that the Government consider providing assistance or income in the form of social security or supplemental social security to reduce mental unrest among people. Since behavioral changes- wearing a mask and avoiding crowded places, sanitizing- are crucial to reducing the COVID-19 spread rate, people affirm this. However, these behavioral changes do not influence the mental condition.

Besides Discovering the possible intervention, we also build machine learning models to predict vulnerable people. This model achieved its best performance while predicting most vulnerable people (having mental stress more than five days in a week) vs. least vulnerable people (having no stress or less than one-day stress in a week). The proposed model finds its application with front-line healthcare workers, psychiatric professionals, and Government who can identify the individuals prone to mental illness. This can be implemented by running the individual's demonstrated factors through our Explainable ML model. Those individuals can then be catered to by specialized health professionals.

Author Contribution:

Study Design : VK,TP,IJ , Dataset: IJ , Data Analysis: IJ,RA , Paper writing: IJ,RA,AK, Paper Review: VK,TP

Conflict of Interest:

None

Funding Statement:

The authors did not have specific funding for this work.

References:

1. Psychiatry TL. Mental health and COVID-19: change the conversation. *Lancet Psychiatry*. 2020;7(6):463. doi:10.1016/S2215-0366(20)30194-2

2. Pereira-Sanchez V, Adiukwu F, Hayek SE, et al. COVID-19 effect on mental health: patients and workforce. *Lancet Psychiatry*. 2020;7(6):e29-e30. doi:10.1016/S2215-0366(20)30153-X

3. Chen Q, Liang M, Li Y, et al. Mental health care for medical staff in China during the COVID-19 outbreak. *Lancet Psychiatry*. 2020;7(4):e15-e16. doi:10.1016/S2215-0366(20)30078-X

4. Mahase E. Covid-19: Mental health consequences of pandemic need urgent research, paper advises. *BMJ*. 2020;369:m1515. doi:10.1136/bmj.m1515

5. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. *N Engl J Med*. 2020;0(0):null. doi:10.1056/NEJMp2008017

6. Gao J, Zheng P, Jia Y, et al. Mental health problems and social media exposure during COVID-19 outbreak. *PLOS ONE*. 2020;15(4):e0231924. doi:10.1371/journal.pone.0231924

7. Homepage | COVID Impact Survey. Untitled. Accessed June 26, 2020. https://www.covid-impact.org

8. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education | SpringerLink. Accessed June 26, 2020.

https://link.springer.com/article/10.1007%2Fs11165-016-9602-2

9. Revelle W. *Psych: Procedures for Psychological, Psychometric, and Personality Research*.; 2020. Accessed June 26, 2020. https://CRAN.R-project.org/package=psych

10. Gámez JA, Mateo JL, Puerta JM. Learning Bayesian networks by hill climbing: efficient methods based on progressive restriction of the neighborhood. *Data Min Knowl Discov*. 2011;22(1):106-148. doi:10.1007/s10618-010-0178-6

11. Neath AA, Cavanaugh JE. The Bayesian information criterion: background, derivation, and applications. *WIREs Comput Stat*. 2012;4(2):199-203. doi:10.1002/wics.199

12. Yedidia JS, Freeman WT, Weiss Y. Generalized Belief Propagation. In: Leen TK, Dietterich TG, Tresp V, eds. *Advances in Neural Information Processing Systems 13*. MIT Press; 2001:689–695. Accessed June 28, 2020. http://papers.nips.cc/paper/1832-generalized-belief-propagation.pdf

13. Scutari M, Ness R. *Bnlearn: Bayesian Network Structure Learning, Parameter Learning and Inference*.; 2019. Accessed June 26, 2020. https://CRAN.R-project.org/package=bnlearn

14. Højsgaard S. *GRain: Graphical Independence Networks*.; 2020. Accessed June 26, 2020. https://CRAN.R-project.org/package=gRain

15. Sethi T, Maheshwari S. *WiseR: A Shiny Application for End-to-End Bayesian Decision Network Analysis and Web-Deployment*.; 2018. Accessed June 26, 2020. https://CRAN.R-project.org/package=wiseR

16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *J Artif Intell Res*. 2002;16:321-357. doi:10.1613/jair.953

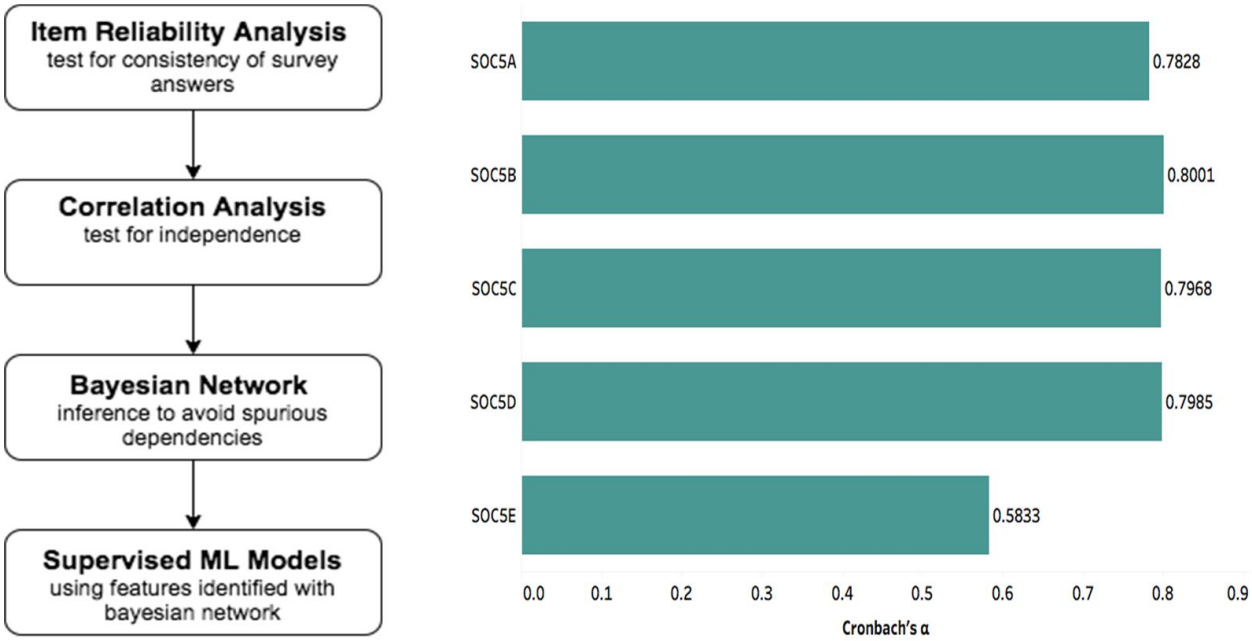
17. scikit-learn: machine learning in Python — scikit-learn 0.23.1 documentation. Accessed June 26, 2020. https://scikit-learn.org/stable/

18. WHO Declared Symptoms:. https://www.who.int/westernpacific/health-topics/coronavirus

19. Brådvik L. Suicide Risk and Mental Disorders. *Int J Environ Res Public Health*. 2018;15(9). doi:10.3390/ijerph15092028

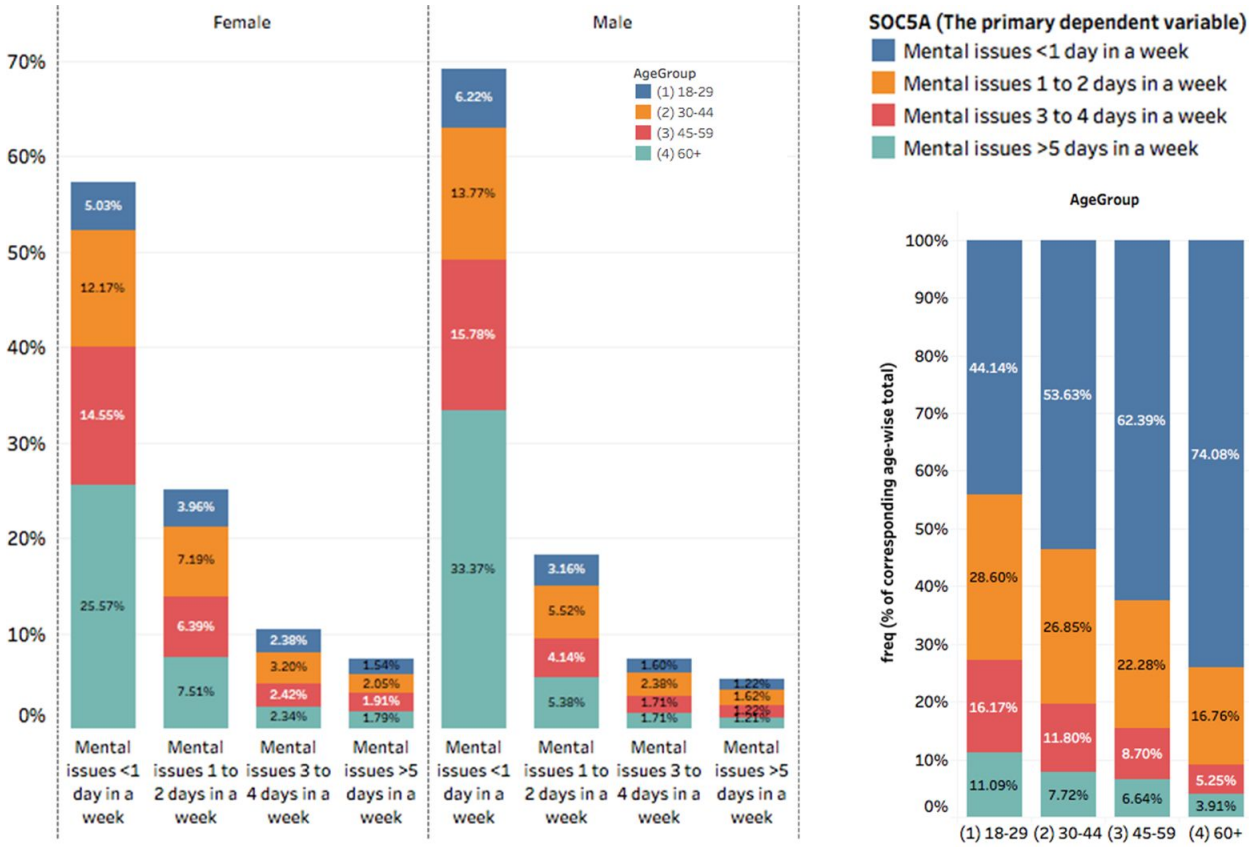
20. Bilsen J. Suicide and Youth: Risk Factors. *Front Psychiatry*. 2018;9. doi:10.3389/fpsyt.2018.00540

Figure1: a) Analysis Flowchart b) Cronbach's alpha value (>0.70) determined by Item Reliability Analysis of Mental Health Indicators



(a) (b)

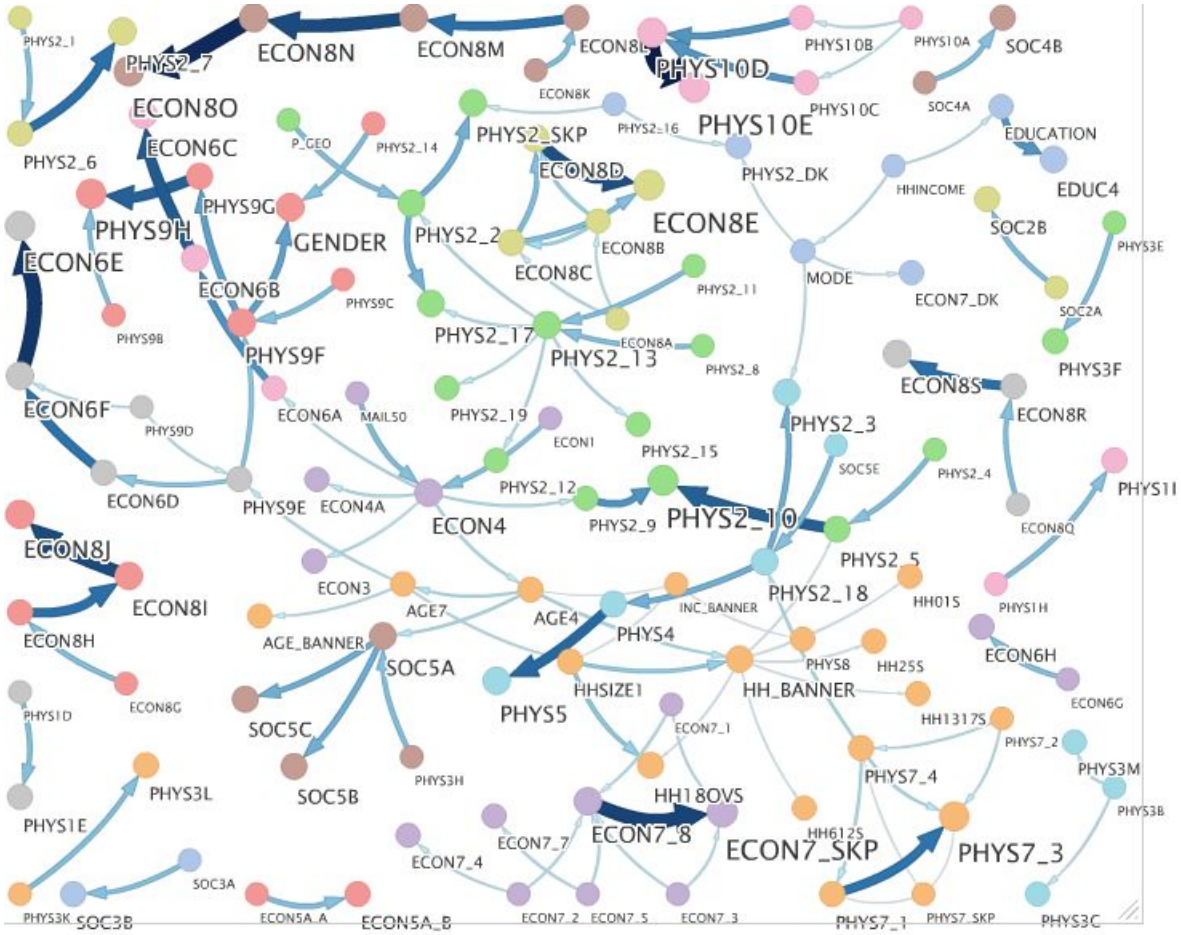
Figure2 : (a) Mental Illness (soc5a) wise distribution of Age. (b) Gender and Age-group wise distribution of Mental Illness (soc5a)



(a) (b)

Figure 3: (a)101 bootstrap structure-learning iterations Ensembled Network Learned upon Impact Survey Data. used hill-climbing search along with Bayesian Information Criterion as the scoring function. Here we are showing connections having edge strength and direction strength more than 90%; (b) SOC5A was found to be the parent Node of all other Mental Health Variables, so we assumed SOC5A as Primary dependent variable

(a)



(b)

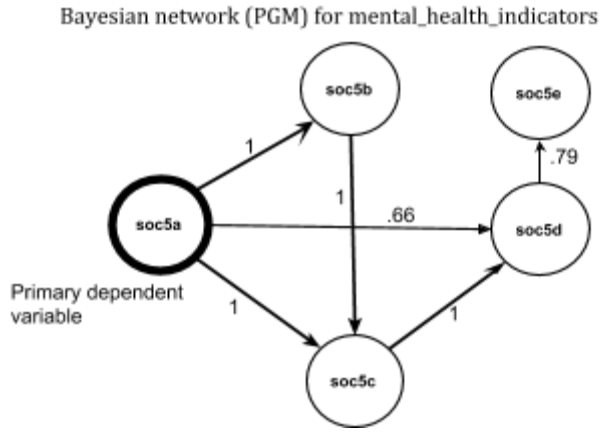


Figure 4: Performance of Stress prediction Model

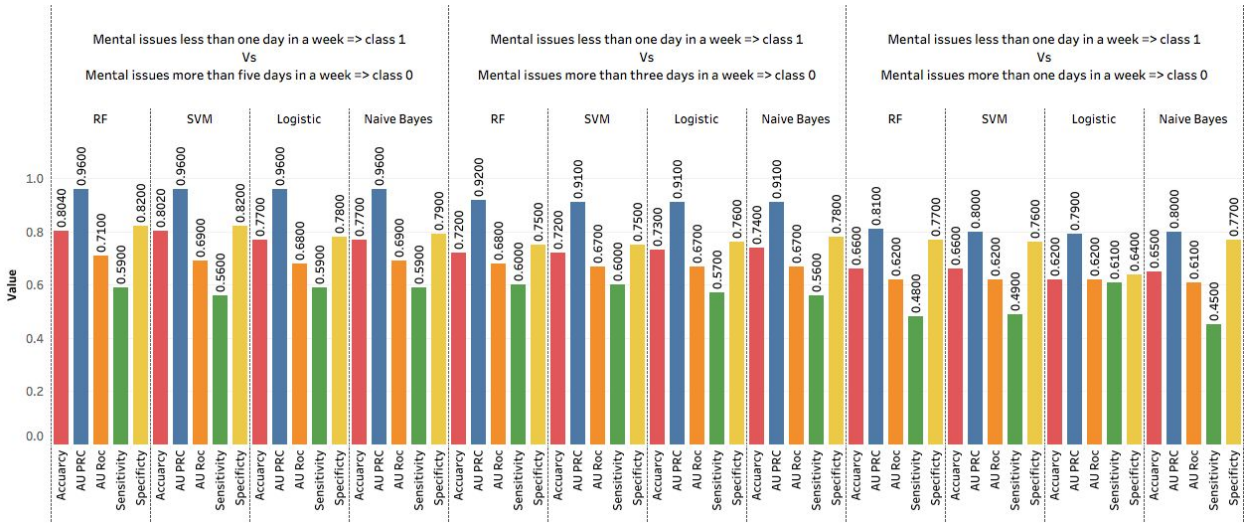


Fig5: Inferences from the Bayesian network: here we are showing the change in inferred probability difference for interventions. A positive association implies a mental-stress-inducer whereas a negative association implies mental-stress-reducer.

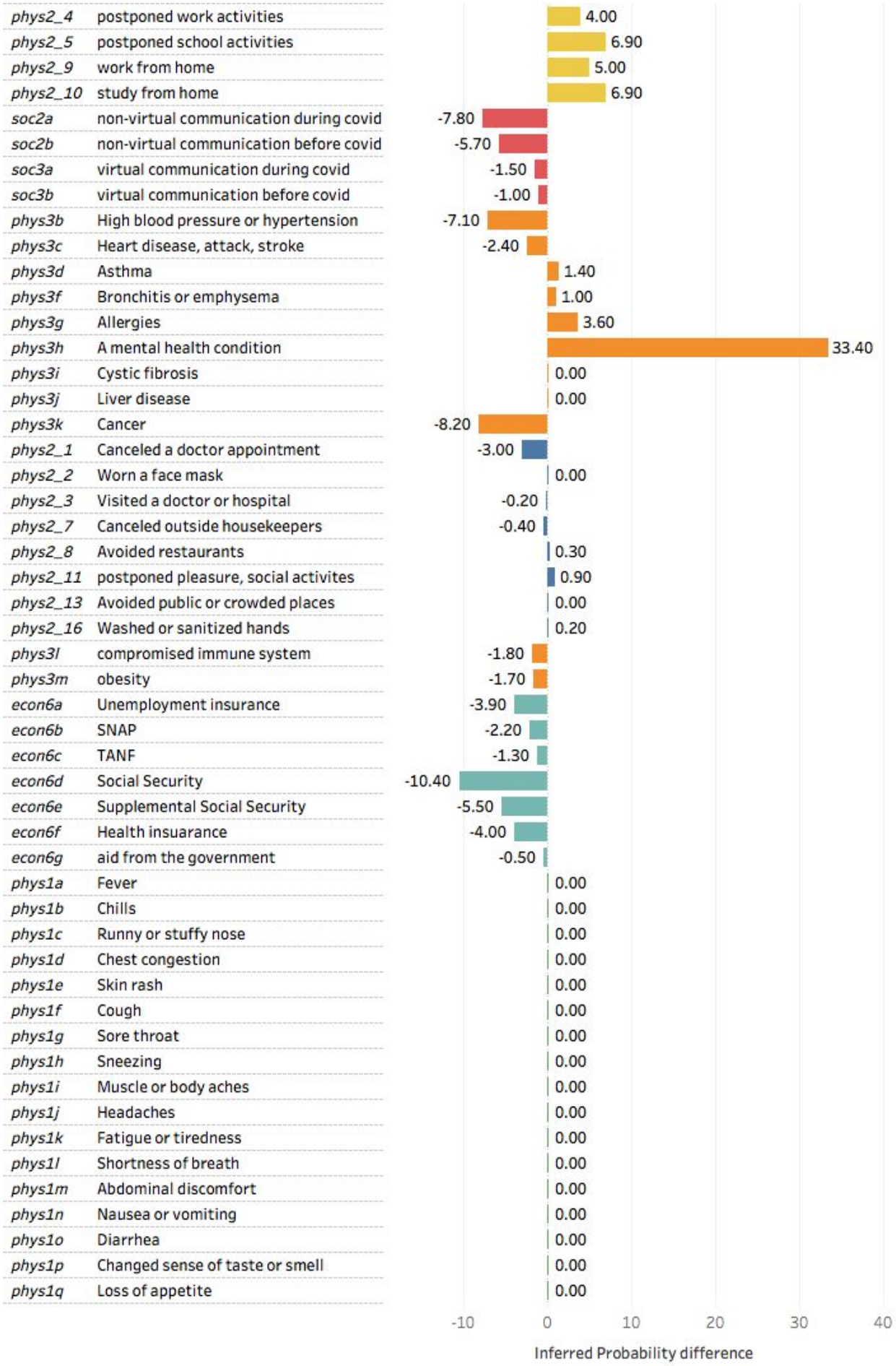

phys2_4 postponed work activities 4.00
phys2_5 postponed school activities 6.90
phys2_9 work from home 5.00
phys2_10 study from home 6.90
soc2a non-virtual communication during covid -7.80
soc2b non-virtual communication before covid -5.70
soc3a virtual communication during covid -1.50
soc3b virtual communication before covid -1.00
phys3b High blood pressure or hypertension -7.10
phys3c Heart disease, attack, stroke -2.40
phys3d Asthma 1.40
phys3f Bronchitis or emphysema 1.00
phys3g Allergies 3.60
phys3h A mental health condition 33.40
phys3i Cystic fibrosis 0.00
phys3j Liver disease 0.00
phys3k Cancer -8.20
phys2_1 Canceled a doctor appointment -3.00
phys2_2 Worn a face mask 0.00
phys2_3 Visited a doctor or hospital -0.20
phys2_7 Canceled outside housekeepers -0.40
phys2_8 Avoided restaurants 0.30
phys2_11 postponed pleasure, social activites 0.90
phys2_13 Avoided public or crowded places 0.00
phys2_16 Washed or sanitized hands 0.20
phys3l compromised immune system -1.80
phys3m obesity -1.70
econ6a Unemployment insurance -3.90
econ6b SNAP -2.20
econ6c TANF -1.30
econ6d Social Security -10.40
econ6e Supplemental Social Security -5.50
econ6f Health insuarance -4.00
econ6g aid from the government -0.50
phys1a Fever 0.00
phys1b Chills 0.00
phys1c Runny or stuffy nose 0.00
phys1d Chest congestion 0.00
phys1e Skin rash 0.00
phys1f Cough 0.00
phys1g Sore throat 0.00
phys1h Sneezing 0.00
phys1i Muscle or body aches 0.00
phys1j Headaches 0.00
phys1k Fatigue or tiredness 0.00
phys1l Shortness of breath 0.00
phys1m Abdominal discomfort 0.00
phys1n Nausea or vomiting 0.00
phys1o Diarrhea 0.00
phys1p Changed sense of taste or smell 0.00
phys1q Loss of appetite 0.00
-10
0
10
20
30
40
Inferred Probability difference

Table 1: Socio-Demographics of Respondents Participated in the Survey

Factors	GENDER	MALE	FEMALE	REFUSED	SKIPPED	TOTAL
AGE	All Age Group	7,886	9,868	1	9	17,764
	18-24	438	559	0	1	998
	25-34	1,224	1,662	0	4	2,890
	35-44	1,130	1,471	0	2	2,603
	45-54	1,096	1,467	0	0	2,563
	55-64	1,578	2,110	1	2	3,691
	65-74	1,654	1,885	0	0	3,539
	75+	765	714	0	0	1,479
	Not Provided	1	0	0	0	1
EDUACTION	All education Group	7,886	9,868	1	9	17,764
	BA or above	4,439	4,922	0	4	9,365
	Some college	2,317	3,192	1	1	5,511
	HS graduate or equivalent	913	1,375	0	0	2,288
	No HS diploma	210	362	0	0	572
	Not sure	1	2	0	0	3
	REFUSED	2	4	0	0	6
	SKIPPED on web	4	11	0	4	19
GEO-LOCATION	All Geo-Locations	7,066	9,134	1	9	16,210
	Atlanta-Sandy Springs-Al..	317	458	0	0	775
	Baltimore-Columbia-Tow..	288	439	0	0	727
	Birmingham-Hoover, Ala..	264	409	0	0	673
	California	608	681	0	0	1,289
	Chicago-perville-Elgin, Ill..	376	473	0	0	849
	Cleveland-Elyria, Ohio	350	475	1	1	827
	Colorado	384	458	0	1	843
	Columbus, Ohio	335	429	0	0	764
	Florida	461	514	0	0	975
	Louisia	249	423	0	0	672
	Minnesota	363	438	0	2	803
	Missouri	288	457	0	0	745
	Monta	358	453	0	1	812
	New York	425	484	0	1	910
	Oregon	355	495	0	1	851
	Phoenix-Mesa-Chandler, ..	267	317	0	0	584
	Pittsburgh, Pennsylvania	336	457	0	1	794
	Texas	326	467	0	1	794
	Not Mentioned/Skipped	716	807	0	0	1,523
	Missing	820	734	0	0	1,554
RACE	All RACE Group	7,886	9,868	1	9	17,764
	Hispanic	622	899	0	1	1,522
	Black, non-Hispanic	521	1,057	0	0	1,578
	White, non-Hispanic	5,240	6,204	0	3	11,447
	Other, non-Hispanic	620	589	0	1	1,210
	REFUSED	2	5	1	0	8
	SKIPPED on web	8	2	0	2	12
	Removed for disclosure ri..	871	1,112	0	2	1,985
	DON'T KNOW	2	0	0	0	2
	Missing	0	0	0	0	0
LANGUAGE	All language Group	7,886	9,868	1	9	17,764
	English	7,815	9,757	1	9	17,582
	Spanish	71	111	0	0	182
		0K 2K 4K 6K 8K	0K 5K 10K	-1 0 1 2	0 5 10	0K 10K

Supplementary Table 1: variable groups as Indicators

Code	Attributes Description
mental_health_indicators *(questions related to mental stress in the last seven days)*	
soc5a	Felt nervous, anxious, or on edge
soc5b	Felt depressed
soc5c	Felt lonely
soc5d	Felt hopeless about the future
soc5e	sweating, trouble breathing, pounding heart, etc
work_from_home_indicators *(questions related to work and school from home)*	
phys2_9	work from home
phys2_10	study from home
phys2_4	postponed work activities
phys2_5	postponed school activities
communication_indicators *(questions related to communication with friends/neighbor/family)*	
soc2a	verbal communication during COVID
soc2b	verbal communication before COVID
soc3a	virtual communication during COVID
soc3b	virtual communication before COVID
covid_symptoms_indicators *(questions on covid symptoms Prescribed by WHO)*	
phys1a	Fever
phys1b	Chills
phys1c	Runny or stuffy nose
phys1d	Chest congestion
phys1e	Skin rash
phys1f	Cough

phys1g	Sore throat
phys1h	Sneezing
phys1i	Muscle or body aches
phys1j	Headaches
phys1k	Fatigue or tiredness
phys1l	Shortness of breath
phys1m	Abdominal discomfort
phys1n	Nausea or vomiting
phys1o	Diarrhea
phys1p	Changed or lost sense of taste or smell
phys1q	Loss of appetite

chronic_medical_conditions_indicator
(questions on the chronic medical condition or prior clinical history of diseases)

phys3a	Diabetes
phys3b	High blood pressure or hypertension
phys3c	Heart disease, attack, stroke
phys3d	Asthma
phys3e	Chronic lung disease or COPD
phys3f	Bronchitis or emphysema
phys3g	Allergies
phys3h	A mental health condition
phys3i	Cystic fibrosis
phys3j	Liver disease
phys3k	Cancer
phys3l	compromised immune system
phys3m	obesity

behavioural_indicators
(questions on behavioral changes proposed/enforced by the government)

phys2_1	Canceled a doctor appointment
phys2_2	Worn a face mask
phys2_3	Visited a doctor or hospital
phys2_7	Canceled outside housekeepers
phys2_8	Avoided restaurants

phys2_11	postponed pleasure, social activities
phys2_12	Stockpiled food or water
phys2_13	Avoided public or crowded places
phys2_16	Washed or sanitized hands
insurance_assistance_indicators *(questions on insurance and different assistance)*	
econ6a	Unemployment insurance
econ6b	SNAP
econ6c	TANF
econ6d	Social Security
econ6e	Supplemental Social Security
econ6f	Health insurance
econ6g	aid from the government
other_useful_Indicators	
phys7_4	did not feel hot, chilly, sweating
phys2_18	Stayed home because I felt unwell

Supplementary Figure 1: Heatmap for negative log p values for the significant association.

	SOC5A	SOC5B	SOC5C	SOC5D	SOC5E
PHYS2_1	3.929	5.110	4.370	3.732	5.999
PHYS2_2	5.834	3.518	5.701	5.602	5.999
PHYS2_3	4.527	4.797	3.146	5.810	6.000
PHYS2_6	5.999	6.000	6.000	6.000	5.978
PHYS2_7	0.986	3.797	2.125	2.636	1.698
PHYS2_9	6.000	6.000	6.000	6.000	5.420
PHYS2_14	1.086	1.349	3.591	1.484	0.768
PHYS2_15	6.000	6.000	6.000	6.000	6.000
PHYS2_16	2.699	3.970	2.889	3.165	2.324
PHYS2_17	5.772	1.669	2.388	2.563	0.668
PHYS2_18	6.000	6.000	6.000	6.000	6.000
PHYS2_19	4.722	3.945	5.556	3.338	6.000
ECON6C	5.973	6.000	5.999	6.000	6.000
PHYS3A	6.000	6.000	6.000	5.951	6.000
PHYS3B	6.000	6.000	6.000	6.000	5.993
PHYS3C	6.000	6.000	6.000	5.979	5.999
PHYS3E	6.000	6.000	6.000	4.895	6.000
PHYS3I	6.000	6.000	4.516	2.490	6.000
PHYS3J	6.000	6.000	5.777	5.756	6.000
PHYS3K	6.000	6.000	6.000	5.999	6.000
SOC2A	6.000	6.000	6.000	6.000	5.671
SOC2B	6.000	6.000	6.000	6.000	5.972
SOC3A	5.847	6.000	6.000	3.953	2.732
SOC3B	5.955	5.835	6.000	4.156	1.039

F3

0.668 6.000