


Predicting Psychological Distress During the COVID-19 Pandemic: Do Socioeconomic Factors Matter?

Social Science Computer Review
2022, Vol. 0(0) 1–25
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/08944393211069622
journals.sagepub.com/home/ssc


Nan Zou Bakkeli¹

Abstract

Background and purpose: The COVID-19 pandemic has posed considerable challenges to people's mental health, and the prevalence of anxiety and depression increased substantially during the pandemic. Early detection of potential depression is crucial for timely preventive interventions; therefore, there is a need for depression prediction.

Data and methods: This study was based on survey data collected from 5001 Norwegians (3001 in 2020 and 2000 in 2021). Machine learning models were used to predict depression risk and to select models with the best performance for each pandemic phase. Probability thresholds were chosen based on cost-sensitive analysis, and measures such as accuracy (ACC) and the area under the receiver operating curve (AUC) were used to evaluate the models' performance.

Results: The study found that decision tree models and regularised regressions had the best performance in both 2020 and 2021. For the 2020 predictions, the highest accuracies were obtained using gradient boosting machines (ACC = 0.72, AUC = 0.74) and random forest algorithm (ACC = 0.71, AUC = 0.75). For the 2021 predictions, the random forest (ACC = 0.76, AUC = 0.78) and elastic net regularisation (ACC = 0.76, AUC = 0.78) exhibited the best performances. Highly ranked predictors of depression that remained stable over time were self-perceived exposure risks, income, compliance with nonpharmaceutical interventions, frequency of being outdoors, contact with family and friends and work–life conflict. While epidemiological factors (having COVID symptoms or having close contact with the infected) influenced the level of psychological distress to a larger extent in the relatively early stage of pandemic, the importance of socioeconomic factors (gender, age, household type and employment status) increased substantially in the later stage.

Conclusion: Machine learning models consisting of demographic, socioeconomic, behavioural and epidemiological features can be used for fast 'first-hand' screening to diagnose mental health problems. The models may be helpful for stakeholders and healthcare providers to provide early diagnosis and intervention, as well as to provide insight into forecasting which social groups are more vulnerable to mental illness in which social settings.

¹Centre for Research on Pandemic & Society; Consumption Studies Norway, Oslo Metropolitan University, Oslo, Norway

Corresponding Author:

Nan Zou Bakkeli, Centre for Research on Pandemic & Society; Consumption Studies Norway, Oslo Metropolitan University, P.O. Box 4, St Olavs Plass, Oslo 0130, Norway.

Email: Nan.Bakkeli@OsloMet.no

Keywords

mental health, depression, COVID-19, social determinants of health, inequality, machine learning

Introduction

The COVID-19 pandemic has majorly impacted our lives. The pandemic's devastation has had a marked effect on people's mental health, leading to higher levels of anxiety, psychological stress and depression. A meta-analysis found that, during the early stages of the pandemic, the prevalence rates for stress, anxiety and depression were as high as 30% in the general population (Salari et al., 2020). Increases in psychological distress and mental illness have been documented after the pandemic struck. The prevalence of depression in China at the peak of the pandemic in 2020 was 20.4%, compared with a 4% prevalence in 2019 (J. Li et al., 2020b). In Norway, the prevalence of depression was three times higher than in pre-pandemic times (31% vs. 10%) (Ebrahimi et al., 2021). The mental health situation has not improved in the second year of the pandemic; according to the U.S. Centers of Disease Control and Prevention, more than 40% of adults reported symptoms of anxiety or depression between August 2020 and February 2021 (Vahratian et al., 2021). In the UK, four in 10 young women (43%) experienced depressive symptoms in 2021 (Williams et al., 2021).

Depression was identified as the most preventable mental disorder (Muñoz et al., 1996), and its negative impact can be largely improved by early detection and treatment (Halfin, 2007). Therefore, early identification of depression is vital for timely intervention to decelerate the severity of the disease and lessen public health burdens. However, early detection and treatment of mental health disorders are rather limited. Most validated laboratory tests for diagnosing depression in psychiatric research are based on self-reports of one's mental status, and 'a common concern on the part of psychiatrists is the fear that numerical assessment is dehumanizing (...) the subject's individuality' (Hamilton, 1980). Moreover, some psychiatric information can be less accessible in cases where people are severely ill, and the depression rate has been found to be substantially underreported for certain social groups (Moussavi et al., 2007). Therefore, there is a need to predict depression based on rapid and easily accessible information as a first-hand screening, and social determinants of mental health problems can provide useful information for such predictions.

However, whilst most social determinant studies have examined causal relationships or contributing factors to mental health, relatively few have incorporated the identified social determinants into models that systematically predict potential mental health outcomes. Existing predictive studies on mental health have mostly been conducted in the fields of psychology and psychiatry, with a focus on psychopathological aspects (e.g. Elhai et al., 2021; Prout et al., 2020). Nevertheless, social factors and environmental aspects are key stressors for mental health risks (Allen et al., 2014; Silva et al., 2016). COVID-19 studies have identified several social determinants of mental health problems, including income, work, living situation, demographics, social contact, family support and the practice of prevention measures (see, for example, a review article from Hossain et al., 2020). Studies that integrate social and environmental features to predict mental health may provide valuable insight into forecasting which social groups are more vulnerable to mental illness in which social settings. Such insights can also help in the planning and provision of mental health care and in taking preventive measures.

The current study focuses on the social determinants of mental health during the COVID-19 pandemic and takes the further step of predicting depression in Norway using supervised machine learning. The input features consist of socioeconomic status, demographic background and pandemic-related social epidemiological and behavioural factors. Using datasets from earlier and

later phases of the COVID-19 pandemic, we will (1) explore important contributors to depression, (2) select models with the best predictive performance to forecast depression for each phase of the pandemic, (3) rank the importance of input variables in predicting depression and (4) map changes in the rankings between the two pandemic phases considered.

This study aims to contribute to the literature on the social determinants of health in several ways. First, while most previous studies have examined the *relationships* between certain factors and mental health, the current study contributes to *predicting* mental health. Using machine learning models, this study incorporates numerous input variables into its analysis and predicts depression based on a wide range of diverse and complex social realities. Second, the predictions of mental health problems in this study focus on social dynamics and people's behavioural patterns in response to pandemic interventions. Psychological distress can be closely related to social and structural possibilities and limitations of people's actions during the pandemic. This study considers social conditions, daily praxis, social contact and social support levels, as well as ongoing stressful circumstances and changes in daily life during the pandemic. It thereby provides a comprehensive view of the mental health crisis from a sociological perspective. Third, by classifying depression in different pandemic phases, this study provides insight into changes in variable importance; thus, it can be useful to adjust preventive measures targeting different social groups in different time periods. Finally, unlike traditional sociological studies that are hypothesis-driven, predictive models from machine learning are not solely based on our preconceived expectations. The machine learning models adopt a more inductive and explorative approach that facilitates the exploration of new relationships and the development of new theories.

Theories and Previous Studies

Social Determinants of Mental Health

Sociological studies in health and illness have assessed the role of social circumstances embedded within medical conditions and reviewed social determinants, such as socioeconomic status and environmental factors, that shape one's mental health outcomes (Allen et al., 2014; Silva et al., 2016). Such an approach emphasises that mental health depends not only on individuals' personalities or physiological functions but also on social conditions, such as demanding social roles, the level of social support, stressful circumstances and life events (Horwitz, 2009). Therefore, this perspective assumes that individuals with the same circumstances will have similar levels of mental health.

The current pandemic can therefore be viewed as a stressful event that triggers a series of mechanisms that frame diverse peoples' mental health differently based on the circumstances in which they are situated. Previous literature has identified four groups of societal features that played important roles regarding people's mental health status during the COVID-19 pandemic: (1) sociodemographic aspects, (2) socioeconomic status, (3) daily praxis and preventive measures and (4) social epidemiological factors. In the following subsections, we will explore how each of the four groups of social determinants contributes to higher levels of anxiety and distress during the pandemic.

Sociodemographic aspects

Sociodemographic aspects, such as gender, age, ethnicity, place of residence and household type, play an important role in predicting mental health status. The link between these factors and mental health can be understood from a life course perspective. Young adults' and adolescents' mental health has been threatened by limited social interaction, reduced communication and peer

support and increased social isolation during the pandemic (Hawes et al., 2021). In addition to a 'rite de passage' with lockdowns and remote learning, the pandemic has also led to higher unemployment among younger people (Fana et al., 2020), resulting in higher levels of anxiety and depression (Varma et al., 2021).

Health consequences are also unequally distributed among men and women and among ethnic minorities and majorities. Compared to men, women have carried a heavier load in childcare provision during the pandemic, have been more likely to reduce working hours and, thus, have experienced a higher level of psychological distress (Zamarro & Prados, 2021). Ethnic minorities also have a substantially higher risk of poor health compared to the majority population (Proto & Quintana-Domeque, 2021). Additionally, minorities are more likely to have a job that is subject to higher infection through contact with others, and they often face higher risks of income loss, financial insecurity and unemployment during the pandemic crisis.

Finally, mental health status has been proven to be correlated with geographic residence and family structures during the pandemic (Grey et al., 2020; Stoye et al., 2020). People living in various residential areas can be differentiated by different interventions and indirectly measure the impact of interventions on anxiety levels. Considering household structure, the pandemic has increased the amount of time people spend at home. On the one hand, living with others can provide support and thus reduce loneliness (Grey et al., 2020). On the other hand, it may also increase the challenge of conducting professional activities from home (Wilder et al., 2020).

Socioeconomic Factors

The labour market structure has been significantly affected by the pandemic. Employment rates fell sharply, work hours were reduced and people with weak attachment to the labour market were most vulnerable to being laid off or losing their jobs (Eurostat, 2020; OECD, 2020). Income loss, as an immediate effect of job insecurity, leads to lower life satisfaction and higher mental distress (Hertz-Palmor et al., 2020). Thereby, constant worries about insecure work situations and low income had a significant impact on mental health during the COVID-19 pandemic (Bakkeli, 2021; Ganson et al., 2021).

People with different occupations were also faced with different levels of mental health challenges. Not all were able to work remotely from home, and many were subject to the risk of infection at work. Some researchers found that essential workers, such as medical staff and healthcare workers, were more likely to report depressive disorders and anxiety compared to other workers (Lu et al., 2020). Others have pointed out that an increased workload and the presence of job-related burnout might have a significant impact on anxiety and depression among healthcare workers (Kang et al., 2020). Among 26,174 surveyed public health workers in the U.S, 53% reported at least one mental health symptom, and the symptoms were more severe among those who had more working hours or less time off from work (Bryant-Genevier et al., 2021).

A growing body of studies has discussed how the work–life balance may impact one's mental health. As remote work has become more common due to COVID-19, new challenges have also emerged. Employees working remotely have reported longer working hours in addition to juggling work and family responsibilities (Wegner, 2021). Previous studies have reported that compared to people who have achieved work–life balance, those who have often experienced conflict between work and family life are more likely to suffer from clinical mental health problems and depression (Frone, 2000). This also applies to remote work during the pandemic; work–life conflict has been negatively associated with psychological wellbeing and mental health (Bakkeli, 2021).

Daily Praxis and Protective Measures

There is a link between clinical depression and people's daily experiences (Horwitz, 2009). Such experiences involve the perceptions of the risk situation, the willingness and ability to follow protective measures and the extent to which one sought support and social contact. Infection controls during the COVID-19 pandemic have mainly relied on nonpharmaceutical interventions (NPIs). These measures have substantially changed peoples' daily routines, normative expectations, and interactive and mobility patterns. People with different experiences and perceptions of situations may adopt different behavioural responses to the pandemic, and different responses may have different psychological impacts on their mental health. For example, perceived health threats moderate social distancing behaviours (Singh, 2020), while the perceived effectiveness of pandemic interventions and compliance with social distancing measures protects an individual from stress, anxiety and depression (Zhao et al., 2020).

Emotional responses to pandemics, such as the fear of being exposed to the coronavirus, are essential for the actions and praxis people adopted. For instance, Harper et al. (2021) found that virus-mitigating behaviour was not caused by political orientation or moral foundation, but by fear of COVID-19. This might explain why the inability to follow preventive measures was shown to increase stress and anxiety levels (Lewis, 2020), and why people who quarantined voluntarily were more like to have depressive symptoms due to fear of the disease (Harris & Sandal, 2021).

Contact-seeking behaviours can also have an impact on mental health status. A longer period of staying at home was correlated with a higher prevalence of depression (Zhao et al., 2020), and social support reduces this risk (Grey et al., 2020). With limited possibilities for mobility, contact through digital tools has become an important means of communication and contact. There might be a correlation between contact-seeking and loneliness – people who feel lonely might be more contact-seeking. However, digital contact facilitates better mental health and wellbeing (Sommerlad et al., 2021), as social contact, whether face-to-face or digital, contributes positively to social relationships and support.

Social Epidemiological Features

Mental health problems are related to a series of epidemiological predictors. Examples of such predictors include whether one was infected with the COVID-19 virus, had COVID-19-like symptoms, belonged to a high-risk group that can become seriously ill from COVID-19, had close contact with someone who contracted the disease, or had someone close who contracted the virus. The disease has been associated with stigmatisation and social disruptions, and having a high risk of being transmitted (or already transmitted) implies stress related to the physical consequences of being ill and fear of severe consequences from the disease (Bo et al., 2021). For hospitalised patients with COVID-19, studies have detected severe levels of stress and depression (Zandifar et al., 2020). Depressive symptoms were also high for people who had suspected or confirmed cases among family members or relatives (G. Li et al., 2020a; Yuan et al., 2020).

Predicting Mental Health Problems

Using social, demographic, behavioural and epidemiological factors previously recognised as having important impacts on mental health, our study aims to explore the use of machine learning for early detection of depression in different COVID-19 pandemic phases. Machine learning applications have gained popularity in health and medical sciences as they offer faster processing to diagnose disease symptoms, advance clinical decision-support tools and improve the accuracy (ACC) of treatment protocols and health outcomes (Chen et al., 2017). Machine learning has been

applied to the processing of disease biomarkers in applications such as detecting tumours (Bocchi et al., 2004), diagnosing neurodegenerative diseases (Salvatore et al., 2014), and predicting mental health disorders (Victor et al., 2019).

There have also been an increasing number of studies in the social sciences using machine learning methods with an aim to predict or forecast outcome variables. Unlike traditional sociological studies on health and illness that are often hypothesis-driven, or to investigate the mechanisms linking social factors to health, machine learning models are advantageous in studying social determinants of health and illness because they can effectively detect predictive patterns in the dataset and use large amounts of social variables combined in interactive ways. The use of predictive models makes it possible to explore generative models that reflect societal diversity and complexity, and help to understand the relationship between social circumstances and individual outcomes (Conte et al., 2012).

Theories can be used to guide machine learning model building, and the models can, in turn, contribute to the testing and further development of theories (Radford & Joseph, 2020). The computationally grounded theory approach combines previous knowledge with computational processing power and pattern recognition, efficiently produces theories grounded in producible results. For example, using social networks and feminist theory, Nelson (2020) constructed machine learning models to compare the dense connectivity between feminist communities and explained the rise of the second wave feminism in New York due to the city's tighter network connections. By testing parameters to validate theory-based hypotheses, Bamman et al. (2014) developed machine learning models to account for the influence of extralinguistic information when identifying literary characters. Using Bayesian networks, Alvarez-Galvez (2016) explained the link between socioeconomic status and self-reported health by exploring various contextual differences, such as welfare regimes in different countries. Using theory-grounded models, machine learning can shed light on both predictions and explanations of social phenomena.

There are relatively few studies that have used only social determinants to predict health outcomes. Several have included psychological factors in their prediction model. For example, using psychopathology, sociodemographic, and exposure variables, Elhai et al. (2021) achieved from 65% to 75% root mean squared error (RMSE) when predicting anxiety and fear of COVID-19 using machine learning models. In another study, Prout et al. (2020) used demographics, emotion regulation and childhood trauma to predict anxiety, depression and post-traumatic stress. They found that the correlations between the predicted values and actual values were 0.71, 0.66 and 0.69, respectively, for the three mental health-related target variables. Therefore, it is interesting to investigate whether predictions regarding mental health problems using socioeconomic and social epidemiological predictors can achieve high performance and ACC, as well as to compare the results based on social factors and results predicted by biopsychosocial features.

This study will first briefly describe the mental health situations among different social groups in Norway. The main analysis will use machine learning models to predict depression in the earlier and later phases of the COVID-19 pandemic and select those models with the best performance for each pandemic phase. Furthermore, we will assess the importance of the contributing factors to depression and map the changes in importance between the two pandemic phases.

Methods

Data

The study used the 2020 and 2021 CorLife survey waves from Consumption Research Norway and the Work Research Institute. The data contained 3001 respondents in 2020 and 2000 respondents in 2021. The baseline data (t_1) was collected from 1000 respondents during the first

week of kindergarten and school shutdowns from 18 to 23 March 2020. Two months later, the second wave of data (t_2) was collected from 2000 respondents from 13 to 22 May 2020. A third wave of data (t_3) was collected from 21 April to 5 May 2021, during the second peak of the coronavirus pandemic in Norway.

Sentio Research Norway and Norstat collected the data. The response rates for the three waves were 41%, 28% and 30%, respectively. The discrepancies were mainly caused by the maintenance of the panel since the consulting firms frequently included new respondents and deleted people who had stopped responding. However, most of the dropouts in such surveys are random. Analysing Norwegian surveys that used the same sampling method as Norstat and Sentio, [Hellevik \(2016\)](#) argued that the sampling methods in collecting survey data were robust, that the nonresponses often happened by chance and that the results based on such data should not be biased.

Variables

The target variable in this study was *depression*, which was measured by using the Symptom Check-List (SCL-5). The SCL-5 is a five-item scale based on the Hopkins Symptom checklist (HSCL-25). Compared to other indicators of psychological distress and mental health problems, such as SCL-25, the Short Form 36 (SF-36) and the five-time mental health index (MHI-5), the SCL-5 has shown high levels of reliability (Cronbach's alpha = 0.87), specificity (82% with cut-off = 2.0) and sensitivity (96% with cut-off = 2.0) for capturing mental health problems, such as anxiety and depression ([Strand et al., 2003](#); [Tambs & Moum, 1993](#)).

A cut-off point at > 2.0 was used to determine if a person was mentally disturbed with symptoms of anxiety and depression ([Strand et al., 2003](#)). The SCL-5 measure was further dichotomised at a raw score of 2, where individuals who scored at least 2 were coded 1 for being depressed and coded 0 otherwise. In 2020 and 2021, respectively, 23.65% and 24.81% people had a score above the cut-off of 2. This means that, on average, one in four Norwegians suffered from mental distress and anxiety during COVID-19.

We used 55 feature variables to predict depression. *Demographics and socioeconomic predictors* included age, gender, education, income, size of geographic residence, household types, employment status, occupations and work–life conflict. *Behavioural and social epidemiological factors* were measured by social distancing behaviours, social contact, protective measures and nonpharmaceutical practices, self-reported disease exposure and corona disease-related risk factors and symptoms. Detailed information about the variables can be found in [Supplemental Appendix A](#).

Analytical Strategy

We applied classification supervised machine learning algorithms with the target and the predictor variables. The wave 1 and 2 surveys were merged to conduct analysis for 2020, and wave 3 was used for analysis in 2021. Analysis for 2020 and 2021 was done separately for each year. For each of the datasets in 2020 and 2021, the entire dataset was then divided into training and test sets. The datasets were then cleaned, merged, recoded and imputed (see [Supplemental Appendix B](#) for data pre-processing). For the training set, we addressed imbalanced datasets in the minority class (people who were psychologically distressed) by synthesising a new sample using the synthetic minority oversampling technique (SMOTE). The balanced samples were used to fit different machine learning models. After hyperparameter optimisation and model tuning, we predicted psychological distress in the test set and evaluated the fit and performance of the machine learning models. See [Figure 1](#) for the analytical framework of this study.

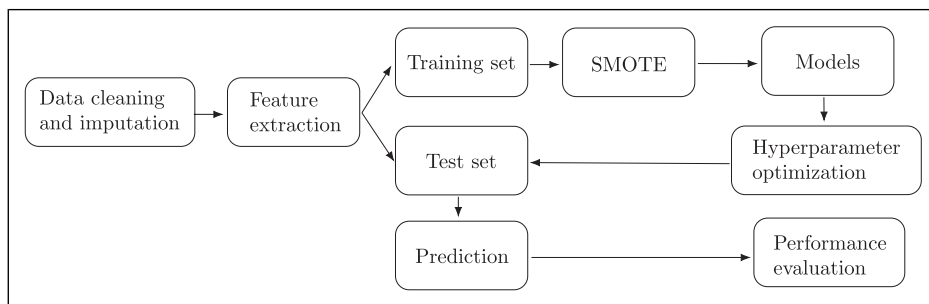


Figure 1. Machine learning framework.

Machine learning predictions are based on the assumption that the test set is large enough to yield statistically meaningful results and representative of the dataset as a whole (He & Ma, 2013). Furthermore, the test set must be as similar to the training set as possible. If the test set has significantly different characteristics than the training set, the classifier may not perform well on the test set. Due to the size of the two waves, we randomly selected 60% of the sample to train the parameters in the model, and we performed 10-fold cross-validation (CV) to evaluate the models. The remaining 40% of the sample served as the unseen test set, which was used to predict depression outcome.

The original dataset contained fewer respondents who suffered from mental distress for both 2020 and 2021 compared to those who did not experience the same level of anxiety and distress. When there are too few minority cases for a model to effectively learn the decision boundary, there might be important consequences for the learning process (Haibo He & Garcia, 2009). An imbalanced classification may produce classifiers with poor predictive ACC for the minority class, that is, the distressed, while classifying most new samples in the majority class for those who were not distressed.

To improve classification performance, the synthetic minority oversampling technique (SMOTE) was applied for data balancing. SMOTE is an oversampling approach that creates synthetic minority class samples, and it first selects cases at random that are close in the feature space, and then it creates synthetic observations based on the existing minority observations (Chawla et al., 2002). The synthetic sample S' is generated by adding the feature vector of the minority class sample X with the Euclidean distance between the minority class and its randomly chosen K -nearest neighbour X_k : $S' = X + |X - X_k| \times R$, where R is a random number and $R \in [0, 1]$.

The machine learning algorithms predict the proportion of people suffering from symptoms of mental distress and anxiety. The selected models represent a broad approach to machine learning. Gradient boosting machines (GBM), random forests (RF) and classification trees (CT) are ensemble decision tree algorithms; support vector machines (SVM) construct hyperplanes to classify data; SVM with sigmoid kernel is also a two-layer perceptron neural network (NN); and ridge (RG), lasso (LS) and elastic net (EN) are regularised methods that penalise the weights in logistic regressions. See [Supplemental Appendix C](#) for descriptions of each machine learning algorithm and hyperparameter optimisation.

Model Evaluation

Several measures were used to evaluate model performance. Prediction ACC and mean misclassification error (MMCE) were used to assess the overall model performance in predicting depression. ACC is the proportion of predictions that are correct among the total number of

predictions. The higher the ACC, the lower the error rate. However, ACC may overestimate model performance with class-imbalanced datasets, and not all types of misclassification errors are equal. The false negative error of wrongly indicating that a depressed person is healthy can be considered more serious than false positive errors, in which a healthy person is misclassified into the category of depression. To take different error types into account, we have adapted a cost-sensitive classification to adjust thresholds to map the depression and non-depression classes (see, e.g. Elkan, 2001).

The cost-sensitive classification considers two measures: *sensitivity*, or ‘true positive rate’, which shows the proportion of depressed people who were correctly identified (the fraction of true positives dividing the sum of true positives and false negatives), and *specificity*, or ‘true negative rate’, which refers to the proportion of correctly predicted non-depressed people (the fraction of true negatives dividing the sum of true positives and false negatives). We tuned the proportion of negative examples (the threshold) to make optimal cost-sensitive classification decisions. The optimal thresholds should achieve a balance between the true positive rate and true negative rate for each model, while at the same time having the lowest possible misclassification errors.

Two other relevant prediction results are *precision* and *negative predictive value* (NPV). Precision represents the proportion of people predicted to be depressed who are truly depressed (the fraction of true positives dividing the sum of true positives and false positives), and NPV indicates the probability that a person predicted to be healthy is indeed healthy.

The receiver operating curve (ROC) was plotted to illustrate the true positive rate (sensitivity) in the function of the false positive rate (1-specificity) for different cut-off points. The probability ROC curve and the area under the curve (AUC) indicate the models’ capability of distinguishing between classes. AUC ranges from 0 to 1, where a model that predicts 100% correctly has an AUC of one.

Finally, we ranked the importance of predictors by using a random forest algorithm. A random forest algorithm uses permutation feature importance to calculate how much each variable decreases the weighted impurity in a tree.

Results

Descriptive Statistics and Odds Ratio of Depression

Table 1 shows the demographic and social epidemiological attributes for the predictor variables and odds ratio of people who were depressed in 2020 and 2021. The descriptive statistics was based on the original datasets before data-partitioning. The columns ‘Yes’ and ‘No’ show the number and percentage of depressed and non-depressed people, while the columns with stars show the odds ratio and standard error of the coefficients in bivariate logistic regressions.

The trends in 2020 and 2021 were very similar. The most distressed were those who were younger, female, had a lower level of education, poorer, lived in Oslo, and those who were not currently employed. Leaders were less distressed than almost all other occupational groups, while a larger proportion of people working in service, sale and elementary occupations suffered from depression. Jobseekers and sickness and disability benefit recipients were also more likely to be depressed. Moreover, those who were outdoors more often were more distressed than those who were less outdoors, and a higher degree of nonpharmaceutical practices correlated with increased anxiety. In addition, those who engaged in digitally contacting family and friends at a higher frequency had higher levels of depression. Anxiety also increased with self-rated disease exposure. Finally, several risk factors for disease transmission and COVID-19-related symptoms increased the proportion of depressed people.

When comparing the odds ratio of depression in 2020 and 2021, the descriptive statistics show that in 2020, only single parents and people living in other types of households were significantly more stressed compared to couples with children. In 2021, the proportion of depressed people

Table 1. Distribution and Odds Ratio of the Depressed.

	Distressed: (2020)			Distressed: (2021)		
	Yes (%)	No (%)	OR [95% CI]★	Yes (%)	No (%)	OR [95% CI]★
Demographic and socioeconomic factors						
Age						
18–22	116 (39.7)	176 (60.3)	Ref	105 (45.9)	124 (54.1)	Ref
23–35	250 (34.1)	484 (65.9)	0.784 [0.59, 1.04] [†]	180 (38.1)	293 (61.9)	0.725 [0.53, 1.00]*
36–55	237 (24.2)	743 (75.8)	0.484 [0.37, 0.64] ^{***}	135 (20.8)	514 (79.2)	0.310 [0.22, 0.43] ^{***}
56–80	113 (11.7)	856 (88.3)	0.200 [0.15, 0.27] ^{***}	76 (11.9)	564 (88.1)	0.159 [0.11, 0.23] ^{***}
Gender						
Male	261 (19.1)	1109 (80.9)	Ref	170 (18.4)	753 (81.6)	Ref
Female	455 (28.4)	1150 (71.6)	1.681 [1.41, 2.00] ^{***}	326 (30.5)	742 (69.5)	1.946 [1.58, 2.40] ^{***}
Education						
High	357 (20.1)	1416 (79.9)	Ref	276 (21.4)	1016 (78.6)	Ref
Low	359 (29.9)	843 (70.1)	1.689 [1.43, 2.00] ^{***}	220 (31.5)	479 (68.5)	1.691 [1.37, 2.08] ^{***}
Income	2.96±0.10	3.43±0.06	0.768 [0.72, 0.82] ^{***}	2.97±0.12	3.62±0.07	0.703 [0.65, 0.76] ^{***}
Geography						
Oslo	142 (31.0)	316 (69.0)	Ref	102 (34)	198 (66)	Ref
>50,000	207 (23.9)	661 (76.1)	0.697 [0.54, 0.90] ^{**}	149 (24.6)	456 (75.4)	0.634 [0.47, 0.86] ^{**}
5,000–50,000	193 (22.6)	660 (77.4)	0.651 [0.50, 0.84] ^{**}	153 (25.1)	457 (74.9)	0.650 [0.48, 0.88] ^{**}
2000–4999	95 (24.8)	288 (75.2)	0.734 [0.54, 1.00] [*]	46 (20.5)	178 (79.5)	0.502 [0.34, 0.75] ^{**}
<2000	79 (19.1)	334 (80.9)	0.526 [0.38, 0.72] ^{***}	46 (18.3)	206 (81.8)	0.433 [0.29, 0.65] ^{***}
Household type						
Couple and child	163 (21.8)	585 (78.2)	Ref	72 (15.6)	389 (84.4)	Ref
Couple no child	225 (21.6)	817 (78.4)	0.988 [0.79, 1.24]	142 (20.6)	549 (79.4)	1.397 [1.02, 1.91] [*]
Living alone	44 (24.3)	137 (75.7)	1.153 [0.79, 1.69]	42 (35.0)	78 (65.0)	2.909 [1.85, 4.57] ^{***}
Single parent	186 (25.7)	539 (74.3)	1.238 [0.97, 1.58] [†]	142 (29.8)	335 (70.2)	2.290 [1.66, 3.15] ^{***}
Other types	98 (35.1)	181 (64.9)	1.943 [1.44, 2.63] ^{***}	98 (40.5)	144 (59.5)	3.677 [2.57, 5.27] ^{***}
Work and occupation						
Not currently employed						
No	334 (22.1)	1175 (77.9)	Ref	232 (21.5)	847 (78.5)	Ref
Yes	382 (26.1)	1084 (73.9)	1.240 [1.05, 1.47] [*]	264 (29)	648 (71)	1.488 [1.21, 1.82] ^{***}
Occupation						
Leaders	14 (13.2)	92 (86.8)	Ref	5 (8.8)	52 (91.2)	Ref
Professionals	60 (21.2)	223 (78.8)	1.768 [0.94, 3.32] [†]	32 (15)	182 (85)	1.829 [0.68, 4.93]
Technicians	58 (20.3)	228 (79.7)	1.672 [0.89, 3.14]	29 (20.7)	111 (79.3)	2.717 [0.77, 7.42] [†]
Clerical support	59 (24.3)	184 (75.7)	2.107 [1.12, 3.97] [*]	44 (22.3)	153 (77.7)	2.991 [1.13, 7.95] [*]
Service and sale	118 (31.0)	263 (69.0)	2.948 [1.61, 5.39] ^{***}	63 (30.7)	142 (69.3)	4.614 [1.76, 12.1] ^{**}
Agricultural	4 (22.2)	14 (77.8)	1.878 [0.54, 6.52]	-	11 (100)	-
Craft and trades	22 (24.4)	68 (75.6)	2.126 [1.01, 4.46] [*]	9 (16.4)	46 (83.6)	2.035 [0.64, 6.51]
Plant and machine	12 (19.4)	50 (80.6)	1.577 [0.68, 3.67]	7 (20.0)	28 (80.0)	2.600 [0.76, 8.95]
Elementary	21 (42.9)	28 (57.1)	4.929 [2.22, 10.9] ^{***}	17 (33.3)	34 (66.7)	5.200 [1.75, 15.4] ^{**}
Job-seekers	141 (37.6)	234 (62.4)	3.960 [2.17, 7.21] ^{***}	124 (43.1)	164 (56.9)	7.863 [3.05, 20.3] ^{***}
Retirees	45 (9.1)	451 (90.9)	0.656 [0.35, 1.24]	37 (11.8)	277 (88.2)	1.389 [0.52, 3.70]
Sickness leave	12 (28.6)	30 (71.4)	2.629 [1.10, 6.30] [*]	4 (28.6)	10 (71.4)	4.160 [0.95, 18.3] [†]
Disability	63 (31.0)	140 (69.0)	2.957 [1.57, 5.59] ^{**}	48 (32.2)	101 (67.8)	4.943 [1.86, 13.2] ^{**}
Others	87 (25.5)	254 (74.5)	2.251 [1.22, 4.15] ^{**}	77 (29.5)	184 (70.5)	4.352 [1.67, 11.3] ^{**}
Essential worker						
No	589 (24.0)	1861 (76.0)	Ref	419 (25.2)	1241 (74.8)	Ref
Yes	127 (24.2)	398 (75.8)	1.008 [0.81, 1.26]	77 (23.3)	254 (76.7)	0.898 [0.68, 1.19]
Work-life conflict	1.84 ± 0.06	1.61 ± 0.03	1.474 [1.32, 1.65] ^{***}	1.70±0.06	1.59±0.04	1.231 [1.07, 1.41] ^{**}
Social epidemiological predictors						
Number of ...						
...Times outdoor	2.69 ± 0.09	2.96 ± 0.05	0.805 [0.74, 0.87] ^{***}	2.95±0.10	3.19±0.05	0.794 [0.72, 0.88] ^{***}

(continued)

Table 1. (continued)

	Distressed: (2020)			Distressed: (2021)		
	Yes (%)	No (%)	OR [95% CI]★	Yes (%)	No (%)	OR [95% CI]★
...People <1 m	2.49 ± 0.10	2.49 ± 0.05	1.004 [0.94, 1.07]	2.51±0.11	2.55±0.07	0.977 [0.90, 1.06]
Willingness stayhome	4.02 ± 0.08	3.96 ± 0.05	1.050 [0.97, 1.13]	3.71±0.11	3.81±0.06	0.935 [0.86, 1.01]
NPI Practice	5.04 ± 0.15	4.81 ± 0.08	1.065 [1.02, 1.11]**	5.66±0.18	5.43±0.10	1.058 [1.01, 1.11]*
Contact with...						
...Family/friends	3.23 ± 0.08	2.93 ± 0.04	1.322 [1.22, 1.44]***	3.13±0.10	2.83±0.05	1.291 [1.17, 1.42]***
...colleagues./work	2.67 ± 0.10	2.74 ± 0.06	0.960 [0.90, 1.02]	3.03±0.12	3.13±0.07	0.950 [0.88, 1.02]
Self-rated exposure	2.31 ± 0.08	1.83 ± 0.04	1.574 [1.45, 1.71]***	1.94±0.10	1.50±0.05	1.423 [1.30, 1.56]***
Transmission risk						
Underlying risk	185 (21.4)	678 (78.6)	0.812 [0.67, 0.98]*	129 (19.6)	528 (80.4)	0.644 [0.51, 0.81]***
Contacted sick	364 (30.0)	851 (70.0)	1.711 [1.44, 2.03]***	190 (31.9)	406 (68.1)	1.665 [1.34, 2.06]***
Quarantine	83 (28.2)	211 (71.8)	1.272 [0.97, 1.67]†	109 (31.6)	236 (68.4)	1.503 [1.17, 1.94]**
Isolation	41 (36.0)	73 (64.0)	1.819 [1.23, 2.69]**	49 (44.1)	62 (55.9)	2.533 [1.72, 3.74]***
Transmitted	6 (85.7)	1 (14.3)	19.082 [2.29, 158.8]**	5 (26.3)	14 (73.7)	1.077 [0.39, 3.01]
Pregnant	6 (22.2)	21 (77.8)	0.901 [0.36, 2.24]	3 (14.3)	18 (85.7)	0.499 [0.15, 1.70]
None	221 (20.8)	844 (79.2)	0.749 [0.63, 0.90]**	159 (23.4)	521 (76.6)	0.882 [0.71, 1.10]
Corona symptom						
Cough	102 (34.7)	192 (65.3)	1.788 [1.38, 2.31]***	34 (35.1)	63 (64.9)	1.673 [1.09, 2.57]**
Heavy breath	85 (48.0)	92 (52.0)	3.173 [2.33, 4.32]***	29 (36.7)	50 (63.3)	1.795 [1.12, 2.87]*
Fever	52 (49.5)	53 (50.5)	3.260 [2.20, 4.83]***	8 (47.1)	9 (52.9)	2.707 [1.04, 7.05]*
Muscle pain	119 (42.5)	161 (57.5)	2.597 [2.02, 3.35]***	71 (47.3)	79 (52.7)	2.994 [2.13, 4.20]***
Diarrhoea	76 (46.6)	87 (53.4)	2.965 [2.15, 4.08]***	29 (50.0)	29 (50.0)	3.139 [1.86, 5.30]***
Tired	207 (46.7)	236 (53.3)	3.486 [2.82, 4.30]***	115 (47.3)	128 (52.7)	4.174 [3.16, 5.51]***
None	387 (18.0)	1759 (82.0)	0.334 [0.28, 0.40]***	322 (20.2)	1271 (79.8)	0.326 [0.26, 0.41]***

Note. *** $p < .001$; ** $p < .01$; * $p < .05$; † $p < .10$. ★ Odds ratio and 95% confidence interval are calculated based on bivariate logistic regressions.

among couples living with their children was significantly lower than in any other household type. When looking at occupational groups, even higher proportions of depressed people could be found among jobseekers, people working within service, sale, or elementary occupations, and those with sickness leave or receiving disability benefits.

Predicting depression

Although the factors contributing to depression have varied across the different stages of the pandemic, the distribution of depression scores (the SCL-scores) looks rather similar in 2020 and 2021 (Supplemental Appendix D). For both years, most people scored lower than two, meaning that majority of people did not suffer from depression and anxiety. The original data was imbalanced and faced a classification problem where the distribution of depressed people was biased. The proportion depressed in these 2 years can be seen in Table 2. The percentage was high in both years, with about one in four being categorised as depressed following the SCL-5 checklist. Still, around 75% of the respondents were not diagnosed as being depressed. After oversampling the training set, the proportion of distressed and not distressed people was significantly more balanced.

Figure 2 shows the comparison of the ROC curves of different models in 2020 and 2021. The ROC curve plots true positive rates (sensitivity) and false positive rates (1-specificity). The y -axis represents the probability of detecting a positive test given that a person had depression, while the x -axis shows the fall-out rate, which is the probability of falsely identify a healthy person as depressed. When predicting depression in 2020, random forest had the highest AUC value (AUC = 0.753),

followed by elastic net (0.752), logistic regression with lasso (0.748) and ridge (0.745) regularisation, and GBM (0.744). AUC was lowest for the classification tree-model (AUC = 0.690).

The 2021 AUCs were generally higher than the 2020 AUCs. When predicting depression in 2020, ridge regression had the highest AUC value (AUC = 0.783), followed by EN (0.781), RF (0.780), SVM (0.778) and lasso regularisation (0.786). The area under the ROC curve was lowest for SVM-model, with an AUC value of 0.624.

There were two possible outcomes to be predicted: whether a person was depressed or not. All individuals are assumed to belong to one of the two populations. The most pressing goal was to detect the risk of being depressed given various social and epidemiological conditions within the timeframe in which individuals were situated, either in an earlier or later stage of the pandemic. A key factor affecting the entries of the prediction is the cost assigned to different types of prediction errors. The consideration of relative costs can be understood as the relative preferences. The choice of cost depends on how willing we are to avoid a certain kind of prediction error at the expense of another error type (Berk, 2008).

The empirical cost ratios were realised by adjusting the probability thresholds for different models (see, e.g. Elkan (2001) for optimal cost-sensitive decisions). Threshold is the proportion of negative examples. Based on the threshold, we can calculate, given a certain false negative rate

Table 2. Percentage of Distressed in the Dataset: Original Data, Training Set and Test Set.

Dataset	2020		2021	
	Distressed	Not distressed	Distressed	Not distressed
Original data	716 (24.07%)	2259 (75.93%)	496 (24.91%)	1495 (75.09%)
Train (before SMOTE)	421 (23.76%)	1.351 (76.24%)	308 (25.77%)	887 (74.23%)
Train (after SMOTE)	1263 (48.32%)	1351 (51.68%)	616 (40.98%)	887 (59.02%)
Test set	295 (24.52%)	908 (75.48%)	188 (23.62%)	608 (76.38%)

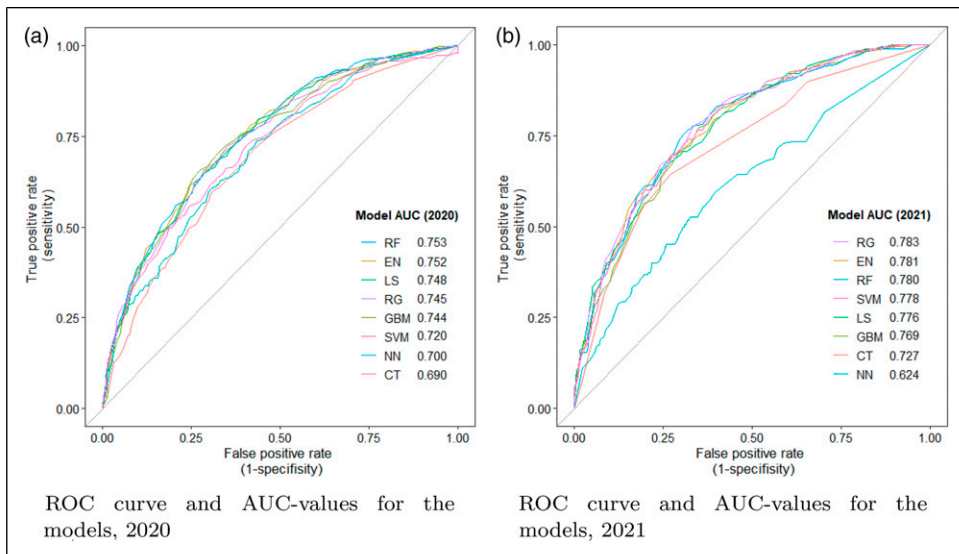


Figure 2. Predictive performance of machine learning models for depression, 2020 and 2021.

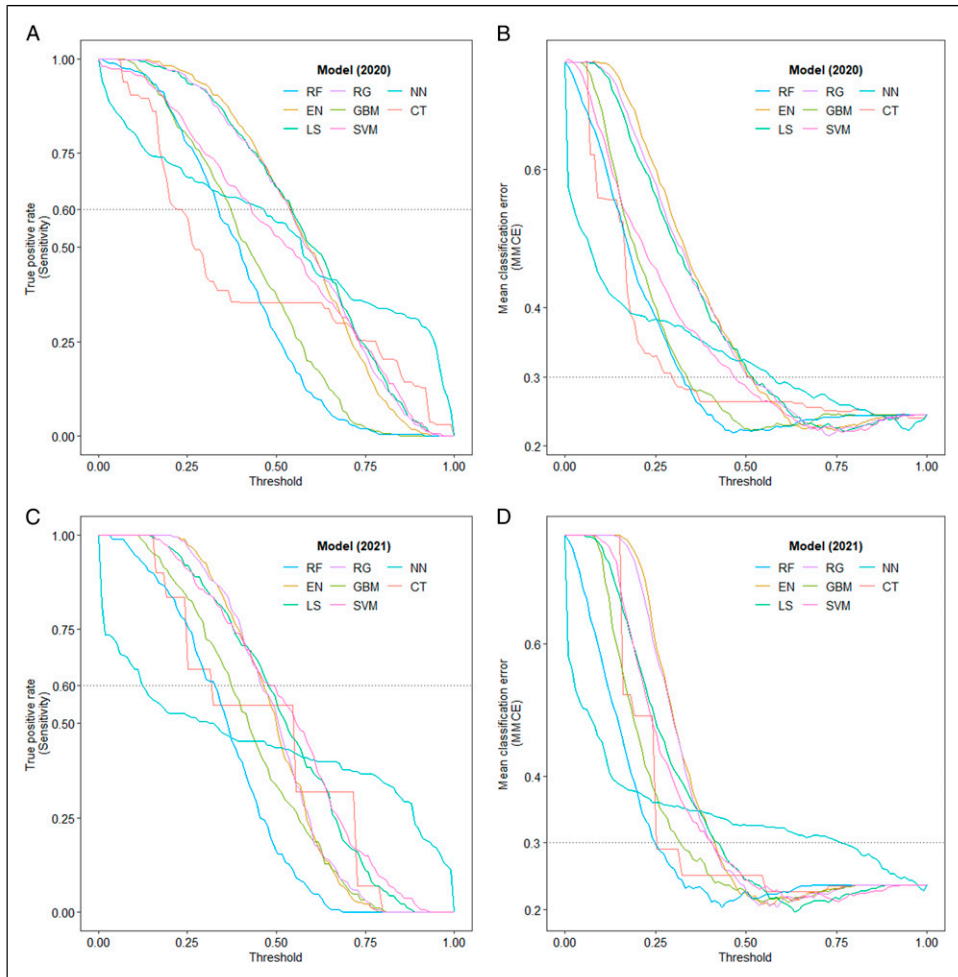


Figure 3. Sensitivity and mean classification error with varying threshold.

(falsely classifying someone as not depressed), how much more costly it would be to receive false positive results (wrongly predicting healthy people as being depressed). For example, a threshold of 0.5 equally weighs the two types of errors. When the threshold is reduced to 0.33, the cost of missing out an actual depressed person is twice as costly as predicting a healthy person as mentally distressed. For a threshold of 0.25, false negative errors weigh three times as high as the false positive errors. See [Supplemental Appendixes E and F](#) for classification thresholds for equal true positive rate and true negative rate in each model.

In this study, we illustrated results that were tuned to achieve a sensitivity level of at least 60% and a mean classification error below 30%. [Figure 3](#) shows how sensitivity and mean classification error changed with the thresholds for the different models. The higher the threshold, the lower the sensitivity and classification error. The horizontal reference lines were also the baselines for 60% sensitivity and 30% classification error. When predicting depression in 2020, sensitivity for the penalised regressions (RG, LS and EN) was relatively high even with a threshold of around 0.5. In other words, these models did not require that more weight be given to false negative rate to achieve a desired proportion of true positive rate. Furthermore, for the SVM-models, sensitivity

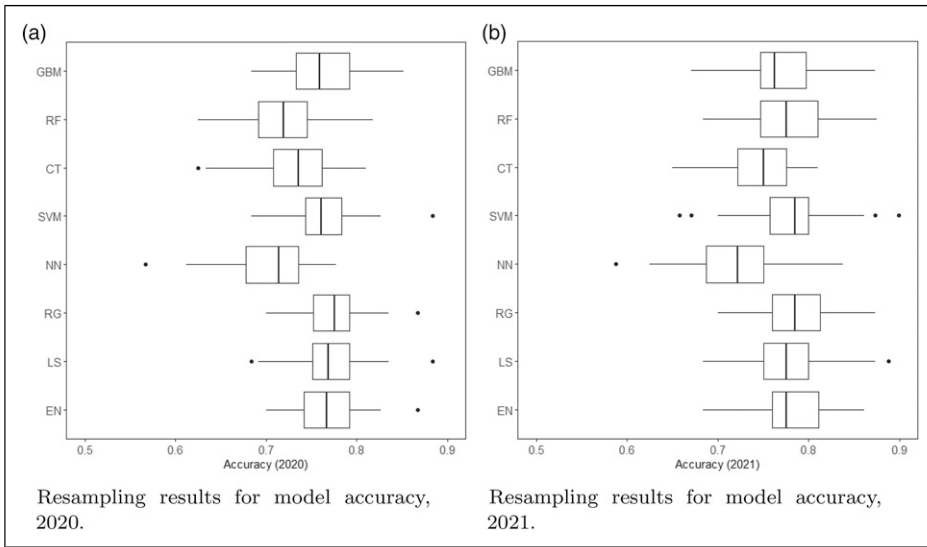


Figure 4. Resampling of accuracies when predicting the test set with chosen hyperparameters.

Table 3. Prediction Performance of Depression, 2020 and 2021.

	Measure	Decision Trees			SVM		Panelised Regression		
		GBM	RF	CT	Radial	NN	Ridge	Lasso	ENet
2020	Accuracy	.720	.711	.666	.681	.675	.696	.688	.691
	Sensitivity	.603	.610	.600	.603	.607	.661	.658	.661
	Specificity	.758	.743	.687	.706	.697	.707	.698	.700
	Precision	.447	.436	.384	.400	.394	.423	.415	.418
	NPV	.855	.854	.841	.846	.845	.865	.863	.864
	AUC	.744	.753	.690	.720	.709	.745	.748	.752
	Threshold	.372	.332	.235	.430	.461	.500	.500	.500
2021	Accuracy	.722	.764	.710	.739	.600	.741	.740	.758
	Sensitivity	.601	.601	.644	.606	.590	.606	.601	.601
	Specificity	.760	.814	.730	.780	.604	.783	.783	.806
	Precision	.447	.500	.425	.460	.315	.463	.461	.489
	NPV	.860	.868	.868	.865	.827	.865	.864	.867
	AUC	.769	.780	.727	.778	.624	.783	.776	.781
	Threshold	.374	.327	.317	.470	.461	.455	.477	.462

fell right below 60% with a threshold of 0.5. Thus, the thresholds needed to be slightly lower than 0.5. Finally, for the decision tree models (GBM, RF and CT), the thresholds needed to be much lower than 0.5 to achieve acceptable results for sensitivity.

A similar pattern can be observed in the 2021 models: Thresholds for the tree models were the lowest to achieve a reasonable true positive rate of at least 60%. For the SVM-models and the penalised regressions, the threshold did not need to be significantly lower than 0.5.

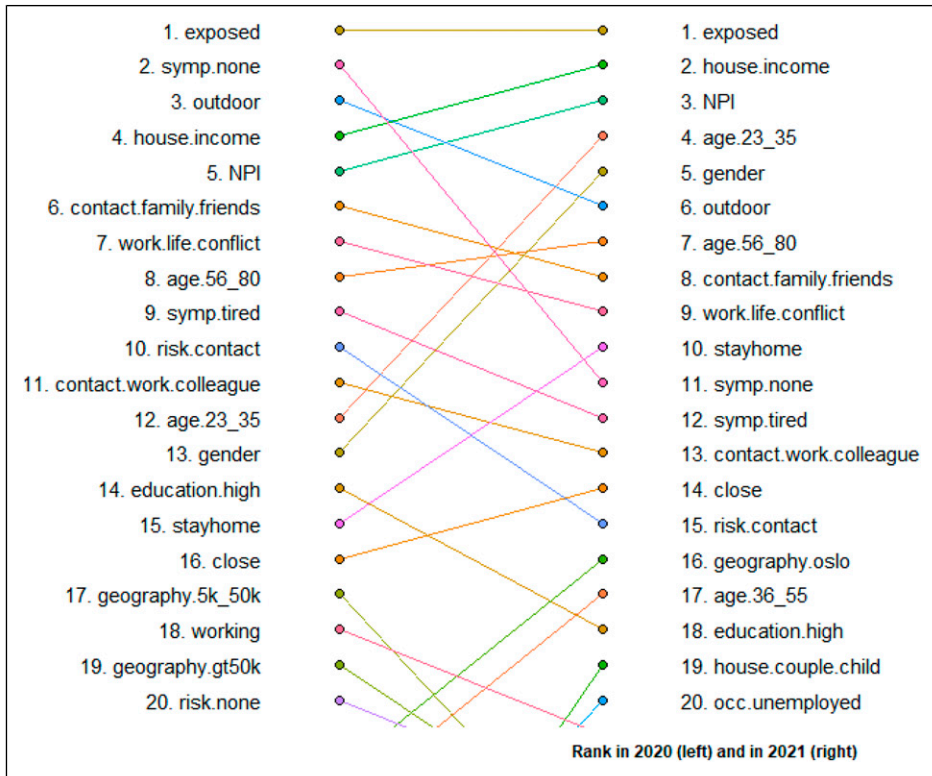


Figure 5. Ranking of the 20 most important variables in 2020 and 2021.

In addition to threshold tuning, the models were optimised to achieve higher accuracies. When using models with selected hyperparameter and threshold, most of the machine learning models had an ACC rate of around or above 70% when predicting the test set. Figure 4 shows the results from resampling model accuracies based on the final models. With the optimised thresholds, regularised logistic regressions, GBM and SVM-radial had the highest median accuracies in 2020, whereas regularised regressions, SVM-radial and tree models had the highest median accuracies in 2021. While NN had the lowest ACC and the highest mean classification errors in both 2020 and 2021.

The summarised results with the selected thresholds are shown in Table 3. For 2020 data, GBM and RF had the highest ACC above 70% after cost-sensitive tuning for sensitivity. To achieve a true positive rate of around 60%, a threshold of 0.37 for GBM and 0.33 for RF were chosen. This corresponds to a weight that was around twice as much on the false positive than the false negative error. Sensitivity was highest (66%) for the regularised regression models without the need to penalise false positive errors (threshold = 0.5). Accuracies (around 70%) for these models were also relatively high. For 2021 predictions, the highest ACC was obtained with random forest and elastic net (both around 76%), closely followed by RG and LS (both around 74%). Classification trees had a sensitivity of 64% and relatively high ACC of 71%. NN displayed the worst performance, in which neither sensitivity nor ACC was higher than 60%.

In sum, the best models for predicting depression for both years of the pandemic were algorithms based on a set of decision trees and the regularised logistic regressions.

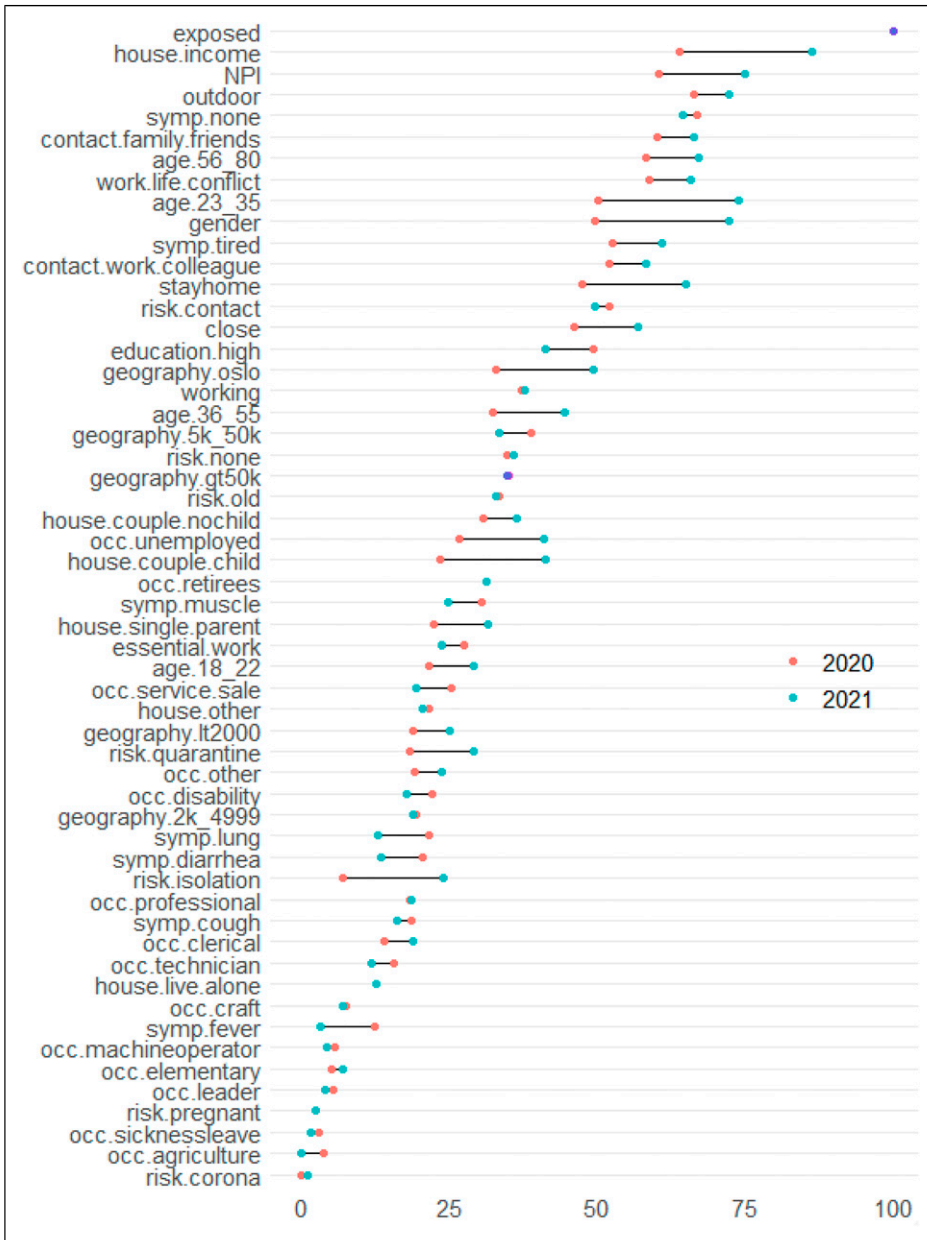


Figure 6. Variable importance in 2020 and 2021 measured by mean decrease Gini.

Importance of Predictive Variables

Due to its overall positive performance in both years, we used random forest to map out how significantly each feature contributes to predicting depression. Figure 5 shows the top 20 most important variables in 2020 and 2021. For both years, self-reported risk for disease exposure was ranked as the most important variable in predicting depression. Other important variables that

were stable over the 2 years were: household income, how often a person went outdoors, the extent to which one has followed nonpharmaceutical measures, the frequency of contacting family and friends, age between 56 and 80 years, and work–life conflict. The rank of variable importance also changed over the 2 years. For example, not having any COVID-19-related symptoms was the second most important feature in predicting depression in 2020, but became less important in 2021. The youngest age group (between 23 and 35 years) and gender became more important in predicting depression in 2021, compared to their relative rank in input importance in 2020. See ranking of all 55 input variables in [Supplemental Appendix G](#).

Finally, [Figure 6](#) shows feature importance measured by mean decrease Gini using random forest. The mean decrease in Gini measures how each variable contributes to the homogeneity of the nodes and leaves in the RF model. The higher the value of mean decrease Gini, the more important the variable in predicting depression. The red and blue dots represent variable importance in 2020 and 2021, respectively. The figure also shows the difference between the 2 years in mean decrease Gini scores for each variable, illustrated by the distance between the red and blue points. Variables that have a large distance are household income, NPI, age between 23 and 35, gender, willingness to stay at home, living in Oslo, household type of being a couple with children, unemployment and being isolated due to COVID-symptoms/transmission. All the mentioned variables have contributed more to predicting depression in 2021, compared to in 2020.

Discussion and Conclusion

This study is based on datasets from 2020 to 2021 that include demographics, social aspects and epidemiological and behavioural factors. We used machine learning models to diagnose depression in different pandemic phases. The results show that the net benefits of the models differ depending on threshold. The decision tree models (gradient boosting machines and random forest algorithms) and regularised regressions (elastic net and ridge and lasso) offer the best performance for predicting depression when evaluating the cost-sensitive balance between different error types, ACC and AUC.

The tree models and penalised regressions offer relatively high performances. With both accuracies and AUCs fall between 0.7 and 0.8, we can conclude that, even without considering psychometric instruments or the psychopathological causes of anxiety and depression, predictive models with social determinants achieve desirable results in predicting psychological distress. This supports the view that beyond the psychological and psychiatric causes of depression, social and environmental factors are important for predicting and explaining mental illnesses and disorders ([Shim & Compton, 2020](#)).

Psychiatrists and mental health professionals are proficient in identifying various driving factors for mental illness, such as poor choices and adverse health behaviour. However, when they attempt to intervene at the level of these factors, it is often too late ([Shim & Compton, 2020](#)). Scholars have thus advocated moving interventions further upstream and seeking the social origins of the risk factors, namely, the social determinants of mental illness ([Shim & Compton, 2020](#)). Pandemic context, life situation and social position can jointly influence individual choice, behavioural patterns and mental health. Therefore, our predictive models based on social factors are congruent with the widely adopted biopsychosocial model for understanding mental disorders and social circumstances.

A crucial point is the discussion concerning how to weigh different kinds of prediction errors: to miss out a true depressive individual, or to falsely diagnose a healthy person. On the one hand, false positives can be harmful as it may lead to unnecessary public expense and financial burdens, and sometimes higher levels of distress ([Klinkman et al., 1998](#)). On the other hand, false negatives may delay people's diagnosis, which lead to significantly more psychiatric disorders and

functional impairment (Leon et al., 1999). The relative costs assigned to type of errors affect the predictions made, and the cost ratios can be a policy matter and should be decided by relevant stakeholders. These discussions merit more attention and should be subject to future studies.

Important Predictors of Mental Health Problems

Among the predictor features, the self-perceived risk of contracting the coronavirus was the most influential factor for predicting depression. This is in line with previous research showing that fear increases anxiety in healthy individuals and intensifies mental illness for those with pre-existing disorders (Elhai et al., 2021; Fitzpatrick et al., 2020). Self-perceived exposure risk is naturally connected with certain circumstances during the pandemic – whether or not a person needs to travel by collective transport or has the possibility of working remotely from home. Therefore, the current mental health cascading from the COVID-19 pandemic has exposed biological and mental expressions of social inequalities.

The study reveals highly ranked predictors of mental health that remain stable over time, including income, compliance with NPI, frequency of being outdoors, contact with family and friends and work–life conflict. These findings support studies that have suggested that social distancing interventions and home shielding have different impacts on mental health: people with a higher level of social and family support are less vulnerable to mental health problems (Grey et al., 2020), while those who were not able to follow protective measures or were exposed to the coronavirus to a greater extent exhibited higher levels of anxiety and distress (cf. Harris & Sandal, 2021).

Variables such as age, gender, education, employment status and geographical location were also among the 20 most important predictors of mental distress and depression in both 2020 and 2021. This is in line with previous studies that found that, during the pandemic, younger adults and women are more vulnerable to stress and depression (Ausín et al., 2021; Varma et al., 2021), and that people having lower educational attainment, who were unemployed or who experienced job loss have higher overall mental distress and anxiety (Ganson et al., 2021; Mazzi et al., 2020).

Many important predictors of depression may be intertwined with a complex social reality. For example, (in)ability to comply with NPIs and social distancing measures may reflect structural and socioeconomic inequalities. Some people have work or life situations in which they are more likely to contract the disease than others, and not all had digital competence or could afford to buy food and daily necessities online. Women are likelier to carry a heavier share of childcare and were therefore under greater pressure from work–life conflict; students who rent apartments in big cities but only received remote lectures are likelier to feel lonely; the unemployed or those who lost their jobs may experience lower self-esteem, greater economic challenges and be more worried about the future. These inequalities may, in turn, lead to more psychological distress. Thus, social inequalities may be translated and reinforced by the pandemic and made visible by emotional burdens and mental illnesses. This also confirms the importance of social risk factors in differentiating mental health outcomes among people with different socioeconomic statuses.

Changes Between Early and Later Pandemic Stages

Certain epidemiological factors (i.e. not having COVID-19-related symptoms and the transmission risk of having close contact with the infected) and living in a small or medium-sized city noticeably declined in importance between 2020 and 2021 as predictors of mental distress. At the same time, the importance of demographic, socioeconomic and behavioural variables such as younger age groups, gender, employment status, couples living with children, living in Oslo and willingness to stay at home increased substantially.

While self-perceived exposure was consistently important for people's psychological wellbeing, epidemiological features such as having COVID symptoms or was in close contact with the infected influenced the level of psychological distress to a larger extent in the relatively early stage of the pandemic. When the pandemic was an unknown event and local governments were experimenting with different measures, people became anxious because the disease seemed close and the situation uncertain. After the pandemic evolved and progressed through the first year, new norms were settled and social differences represented by gender, age, region, household type and employment situation started to emerge. Younger age groups were more exposed to social isolation, couples with children struggled more with balancing life, work and free time, gender inequality became more visible, the unemployed were put under even greater pressure, and long-term social distancing led to loneliness and depression. The lagged effects of social inequalities are embodied in mental health status, and thus, the mentioned social determinants became relatively more important later in the pandemic.

Residential regions implemented different preventive measures and structural conditions during the pandemic, which may impact people's stress levels and mental health status. In 2020, the preventive measures were rather similar between larger cities, while they were more relaxed in smaller places due to lower transmission rates. In 2021, while lives gradually became rather normal in other regional areas, Oslo implemented stricter lockdown policies and preventive measures. Therefore, the pandemic might have a larger impact on feelings of isolation, anxiety and distress for citizens living in Oslo than in other parts of Norway.

From Predictive Models to 'First-Hand' Screening

Predictions made by machine learning algorithms are often flexible, and their performance criteria are often easy to modify with respect to optimising parameters for ACC and adjusting for certain types prediction error. The present study considers the trade-offs of different types of errors and considers false negative errors as being costlier than false positive errors. Applications of these models in future studies may consider the cost benefits of certain prediction outcomes with the involvement of stakeholders such as health professionals and decision makers.

Additionally, traditional methods, such as linear or non-linear regressions, are often unsuitable for including many explanatory variables, as they are prone to overfitting and are challenged by poor predictive performance. Machine learning models are advantageous for handling numerous variables. This study overcomes these challenges and includes a wide range of social, demographic, behavioural and epidemiological factors when predicting depression. By leveraging studies on identifying the social determinants of mental health to predict outcomes, machine learning models can lead to rapid diagnosis and treatment, as well as translation of data for decision makers to bridge process gaps and promote overall preventative public health. Stakeholders will be able to forecast vulnerability and risk factors that have central roles in aggravating depressive symptoms, and the prediction models can help them implement interventions that prevent persistent mental health problems. See detailed information about usability of the machine learning models in [Supplemental Appendix H](#).

The study has some limitations. First, it does not include important psychometric, genetic or epigenetic markers in predicting mental health. Studies with such variables may achieve higher ACC and better model performance. Second, the study only used a measure of depressive diagnosis. It would be interesting to look at the type of and the severity of depressive symptoms. Furthermore, the survey focuses on the pandemic situation in Norway, and can therefore not be generalised to forecast mental health for other countries. However, the predictive variables used in this study are relatively convenient to access. In addition to a standard screening process, the models can contribute in situations when information is less accessible, for example in situations where people are severely depressed, or when interview meetings may not be suitable. It may also help

stakeholders and mental health practitioners to perform a simple and fast ‘first-hand’ screening to diagnose people who are more likely to suffer from depression in the pandemic situation.

A major clinical challenge in mental health is that traditional diagnostic systems perform poorly in terms of the detection of the early stages of mental disorders (Hickie et al., 2019). There is a large group of low to moderate cases of mental illness that are often underserved due to constrained resources – the healthcare system often prioritises people with severe mental disorders. Therefore, the dimension of inequality in treatment and screening does not only originate in individuals’ socio-economic positions; it is also rooted in the practical decisions that have been adopted in the traditional mental healthcare system, as well as unequal resource allocation and structural deficiencies. Instead of conventional health service models of ‘majority in-clinic and minority web-delivered’ care, a flipped digital solution may deliver better mental health services in response to public health crises during the pandemic (Davenport et al., 2020). In addition, the use of technology-enabled care coordination contributes to improving mental health outcomes and reducing potential deaths by suicide (Iorfino et al., 2021). Furthermore, this type of care increases the capacity of primary care and healthcare practitioners (Young & Nesbitt, 2017). Therefore, the predictive models in our study may potentially be integrated with digital tools to provide faster and easier accessible mental health prediction for people with low to moderate mental distress. As early detection of depression is crucial, such a first-hand screening tool may help with the early diagnosis of mental health problems and contribute to timely preventative interventions.

Acknowledgments

Thanks to Torbjørn Skardhamar, Svenn-Erik Mamelund and Åke Brännström for discussion and feedbacks, and thanks to the editor and anonymous referees for valuable comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This paper is part of the project ‘CorRisk: Early COVID-19 wave in Norway: Social inequality in morbidity, compliance to nonpharmaceutical interventions and labour market consequences’, funded by a grant from the Research Council of Norway, grant no. 312716. Oslo Metropolitan University and Consumption Research Norway (SIFO) has funded data collection for this study.

Supplemental Material

Supplemental material for this article is available online.

References

- Allen, J., Balfour, R., Bell, R., & Marmot, M. (2014). Social determinants of mental health. *International Review of Psychiatry*, 26(4), 392–407. <https://doi.org/10.3109/09540261.2014.928270>
- Alvarez-Galvez, J. (2016). Discovering complex interrelationships between socioeconomic status and health in Europe: A case study applying Bayesian networks. *Social Science Research*, 56, 133–43. <https://doi.org/10.1016/j.ssresearch.2015.12.011>
- Ausín, B., González-Sanguino, C., Castellanos, M. Á., & Muñoz, M. (2021). Gender-related differences in the psychological impact of confinement as a consequence of COVID-19 in Spain. *Journal of Gender Studies*, 30(1), 29–38. <https://doi.org/10.1080/09589236.2020.1799768>

- Bakkeli, N. Z. (2021). Health, work, and contributing factors on life satisfaction: a study in Norway before and during the COVID-19 pandemic. *SSM - Population Health*, 14, 100804. <https://doi.org/10.1016/j.ssmph.2021.100804>
- Bamman, D., Underwood, T., & Smith, N. A. (2014). A bayesian mixed effects model of literary character. *ACL*, 1, 370–379. <https://doi.org/10.3115/v1/p14-1035>
- Berk, R. A. (2008). Statistical learning from a regression perspective. In: *Springer series in statistics*. Springer-Verlag.
- Bo, H. X., Li, W., Yang, Y., Wang, Y., Zhang, Q., Cheung, T., Wu, X., & Xiang, Y. T. (2021). Posttraumatic stress symptoms and attitude toward crisis mental health services among clinically stable patients with COVID-19 in China. *Psychological Medicine*, 51(6), 1052–1053. <https://doi.org/10.1017/S0033291720000999>
- Bocchi, L., Coppini, G., Nori, J., & Valli, G. (2004). Detection of single and clustered microcalcifications in mammograms using fractals models and neural networks. *Medical Engineering & Physics*, 26(4), 303–312. <https://doi.org/10.1016/j.medengphy.2003.11.009>
- Bryant-Genevier, J., Rao, C. Y., Lopes-Cardozo, B., Kone, A., & Brykit, R. (2021) *Symptoms of depression, anxiety, post-traumatic stress disorder, and suicidal ideation among state, tribal, local, and territorial public health workers during the COVID-19 pandemic — United States, March–April 2021*. Morbidity and Mortality Weekly Report (70). <https://doi.org/10.15585/mmwr.mm7026e1>
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, 5, 8869–8879. <https://doi.org/10.1109/access.2017.2694446>
- Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics*, 214(1), 325–346. <https://doi.org/10.1140/epjst/e2012-01697-8>
- Davenport, T. A., Cheng, V. W. S., Iorfino, F., Hamilton, B., Castaldi, E., Burton, A., Scott, E. M., & Hickie, I. B. (2020). Flip the clinic: a digital health approach to youth mental health service delivery during the COVID-19 pandemic and beyond. *JMIR mental health*, 7(12), e24578. <https://doi.org/10.2196/24578>. JMIR mental health distributor: JMIR mental health institution: JMIR mental health label: JMIR mental health publisher JMIR publications inc, Toronto, Cannada
- Ebrahimi, O. V., Hoffart, A., & Johnson, S. U. (2021). Physical distancing and mental health during the COVID-19 pandemic: factors associated with psychological symptoms and adherence to pandemic mitigation strategies. *Clinical Psychological Science*, 9(3), 489–506. <https://doi.org/10.1177/2167702621994545>
- Elhai, J. D., Yang, H., McKay, D., Asmundson, G. J. G., & Montag, C. (2021). Modeling anxiety and fear of COVID-19 using machine learning in a sample of Chinese adults: associations with psychopathology, sociodemographic, and exposure variables. *Anxiety, Stress, and Coping*, 34(2), 130–144. <https://doi.org/10.1080/10615806.2021.1878158>
- Elkan, C. (2001). The foundations of cost-sensitive learning. In: Proceedings of the 17th international joint conference on Artificial intelligence - Volume 2, *IJCAI'*, San Francisco, CA, USA, August 4 - 10, 2001. Morgan Kaufmann Publishers Inc, 973–978.
- Eurostat. (2020). Labour market in the light of the COVID 19 pandemic - quarterly statistics. In *Eurostat, Statistics Explained*. Luxembourg: Eurostat.
- Fana, M., Torrejón Pérez, S., & Fernández-Macías, E. (2020). Employment impact of Covid-19 crisis: From short term effects to long terms prospects. *Journal of Industrial and Business Economics*, 47(3), 391–410. <https://doi.org/10.1007/s40812-020-00168-5>
- Fitzpatrick, K. M., Drawve, G., & Harris, C. (2020). Facing new fears during the COVID-19 pandemic: The state of America's mental health. *Journal of Anxiety Disorders*, 75(1), 102291. <https://doi.org/10.1016/j.janxdis.2020.102291>

- Frone, M. R. (2000). Work-family conflict and employee psychiatric disorders: The national comorbidity survey. *The Journal of Applied Psychology, 85*(6), 888–95. <https://doi.org/10.1037/0021-9010.85.6.888>
- Ganson, K. T., Tsai, A. C., Weiser, S. D., Benabou, S. E., & Nagata, J. M. (2021). Job insecurity and symptoms of anxiety and depression among u.s. young adults during COVID-19. *Journal of Adolescent Health, 68*(1), 53–56. <https://doi.org/10.1016/j.jadohealth.2020.10.008>
- Grey, I., Arora, T., Thomas, J., Saneh, A., Tohme, P., & Abi-Habib, R. (2020). The role of perceived social support on depression and sleep during the COVID-19 pandemic. *Psychiatry Research, 293*(1), 113452. <https://doi.org/10.1016/j.psychres.2020.113452>
- Haibo He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. <https://doi.org/10.1109/tkde.2008.239>
- Halfin, A. (2007). Depression: The benefits of early and appropriate treatment. *The American Journal of Managed Care, 13*(4 Suppl), S92-7. PMID: 18041868.
- Hamilton, M. (1980). Rating depressive patients. *The Journal of Clinical Psychiatry, 41*(12 Pt 2), 21–4. PMID: 7440521.
- Harper, C. A., Satchell, L. P., Fido, D., & Latzman, R. D. (2021). Functional fear predicts public health compliance in the COVID-19 pandemic. *International Journal of Mental Health and Addiction, 19*, 1875–1888. DOI: <https://doi.org/10.1007/s11469-020-00281-5>
- Harris, S. M., & Sandal, G. M. (2021). COVID-19 and psychological distress in Norway: The role of trust in the healthcare system. *Scandinavian Journal of Public Health, 49*(1), 96–103. <https://doi.org/10.1177/1403494820971512>
- Hawes, M. T., Szenczy, A. K., Klein, D. N., Hajcak, G., & Nelson, B. D. (2021). Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. *Psychological Medicine, 1*–9. <https://doi.org/10.1017/S0033291720005358>
- He, H., & Ma, Y. (2013). *Imbalanced learning: Foundations, algorithms, and applications*. Wiley-IEEE Press.
- Hellevik, O. (2016). Extreme nonresponse and response bias. *Quality & Quantity, 50*(5), 1969–1991. <https://doi.org/10.1007/s11135-015-0246-5>
- Hertz-Palmor, N., Moore, T., Gothelf, D., DiDomenico, G., Dekel, I., Greenberg, D., Brown, L., Matalon, N., Visoki, E., White, L., Himes, M., Schwartz-Lifshitz, M., Gross, R., Gur, R., Gur, R., Pessach, I., & Barzilay, R. (2020). *Association among income loss, financial strain and depressive symptoms during COVID-19: Evidence from two longitudinal studies*. medRxiv.
- Hickie, I. B., Scott, E. M., Cross, S. P., Iorfino, F., Davenport, T. A., Guastella, A. J., Naismith, S. L., Carpenter, J. S., Rohleder, C., Crouse, J. J., Hermens, D. F., Koethe, D., Markus Leweke, F., Tickell, A. M., Sawrikar, V., & Scott, J. (2019). Right care, first time: A highly personalised and measurement-based care model to manage youth mental health. *The Medical Journal of Australia, 211* Suppl 9(S9), S3–S46. <https://onlinelibrary.wiley.com/doi/pdf/10.5694/mja2.50383>. eprint
- Horwitz, A. V. (2009). An overview of sociological perspectives on the definitions, causes, and responses to mental health and illness. In T. L. Scheid, & T. N. Brown (Eds), *A Handbook for the Study of Mental Health: Social Contexts, Theories, and Systems* 2 edition. Cambridge University Press, 6–19.
- Hossain, M. M., Tasnim, S., Sultana, A., Faizah, F., Mazumder, H., Zou, L., McKyer, E. L. J., Ahmed, H. U., & Ma, P. (2020). Epidemiology of mental health problems in COVID-19: A review. *F1000 Research, 9*, 636. <https://doi.org/10.12688/f1000research.24457.1>
- Iorfino, F., Occhipinti, J. A., Skinner, A., Davenport, T., Rowe, S., Prodan, A., Sturgess, J., & Hickie, I. B. (2021). The impact of technology-enabled care coordination in a complex mental health system: a local system dynamics model. *Journal of Medical Internet Research, 23*(6)e25331. <https://doi.org/10.2196/25331>. Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher JMIR Publications Inc

- Kang, L., Li, Y., Hu, S., Chen, M., Yang, C., Yang, B. X., Wang, Y., Hu, J., Lai, J., Ma, X., Chen, J., Guan, L., Wang, G., Ma, H., & Liu, Z. (2020). The mental health of medical workers in Wuhan, China dealing with the 2019 novel coronavirus. *The Lancet Psychiatry*, 7(3), e14. [https://doi.org/10.1016/S2215-0366\(20\)30047-X](https://doi.org/10.1016/S2215-0366(20)30047-X)
- Klinkman, M. S., Coyne, J. C., Gallo, S., & Schwenk, T. L. (1998). False positives, false negatives, and the validity of the diagnosis of major depression in primary care. *Annals of Family Medicine*, 7(5), 451–61. <https://doi.org/10.1001/archfami.7.5.451>
- Leon, A. C., Portera, L., Olfson, M., Kathol, R., Farber, L., Lowell, K. N., & Sheehan, D. V. (1999). Diagnostic errors of primary care screens for depression and panic disorder. *The International Journal of Psychiatry in Medicine*, 29(1), 1–11. <https://doi.org/10.2190/7AMF-D1JL-8VHA-APGJ>
- Lewis, K. (2020). COVID-19: Preliminary data on the impact of social distancing on loneliness and mental health. *Journal of Psychiatric Practice*, 26(5), 400–404. <https://doi.org/10.1097/PRA.0000000000000488>
- Li, G., Miao, J., Wang, H., Xu, S., Sun, W., Fan, Y., Zhang, C., Zhu, S., Zhu, Z., & Wang, W. (2020). Psychological impact on women health workers involved in COVID-19 outbreak in Wuhan: A cross-sectional study. *Journal of Neurology, Neurosurgery, and Psychiatry*, 91(8), 895–897. <https://doi.org/10.1136/jnnp-2020-323134>
- Li, J., Yang, Z., Qiu, H., Wang, Y., Jian, L., Ji, J., & Li, K. (2020b). Anxiety and depression among general population in China at the peak of the COVID-19 epidemic. *World Psychiatry*, 19(2), 249–250. <https://doi.org/10.1002/wps.20758>
- Lu, W., Wang, H., Lin, Y., & Li, L. (2020). Psychological status of medical workforce during the COVID-19 pandemic: A cross-sectional study. *Psychiatry Research*, 288(2), 112936. <https://doi.org/10.1016/j.psychres.2020.112936>
- Mazzi, M. C., Iavarone, A., Musella, C., De Luca, M., de Vita, D., Branciforte, S., Coppola, A., Scarpa, R., Raimondo, S., Sorrentino, S., Lualdi, F., & Postiglione, A. (2020). Time of isolation, education and gender influence the psychological outcome during COVID-19 lockdown in caregivers of patients with dementia. *European Geriatric Medicine*, 11(6), 1095–1098. <https://doi.org/10.1007/s41999-020-00413-z>
- Moussavi, S., Chatterji, S., Verdes, E., Tandon, A., Patel, V., & Ustun, B. (2007). Depression, chronic diseases, and decrements in health: Results from the world health surveys. *Lancet*, 370(9590), 851–8. [https://doi.org/10.1016/S0140-6736\(07\)61415-9](https://doi.org/10.1016/S0140-6736(07)61415-9)
- Muñoz, R. F., Mrazek, P. J., & Haggerty, R. J. (1996). Institute of medicine report on prevention of mental disorders: Summary and commentary. *American Psychologist*, 51(11), 1116–1122. <https://doi.org/10.1037//0003-066x.51.11.1116>
- Nelson, L. K. (2020). Computational grounded theory: A methodological framework. *Sociological Methods & Research*, 49(1), 3–42. <https://doi.org/10.1177/0049124117729703>
- OECD (2020). *OECD employment outlook 2020: Worker security and the COVID-19 crisis*. Technical report. OECD Publishing.
- Proto, E., & Quintana-Domeque, C. (2021). COVID-19 and mental health deterioration by ethnicity and gender in the UK. *Plos One*, 16(1), e0244419. <https://doi.org/10.1371/journal.pone.0244419>
- Prout, T. A., Zilcha-Mano, S., Aafjes-van Doorn, K., Békés, V., Christman-Cohen, I., Whistler, K., Kui, T., & Di Giuseppe, M. (2020). Identifying predictors of psychological distress during COVID-19: A machine learning approach. *Frontiers in Psychology*, 11(1), 586202. <https://doi.org/10.3389/fpsyg.2020.586202>
- Radford, J., & Joseph, K. (2020). Theory in, theory out: The uses of social theory in machine learning for social science. *Frontiers in Big Data*, 3(1), 18. <https://doi.org/10.3389/fdata.2020.00018>
- Salari, N., Hosseini-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulpoor, S., Mohammadi, M., Rasoulpoor, S., & Khaledi-Paveh, B. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health*, 16(1), 57. <https://doi.org/10.1186/s12992-020-00589-w>
- Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopez, M., Arabia, G., Morelli, M., Gilardi, M. C., & Quattrone, A. (2014). Machine learning on brain MRI data for differential diagnosis of

- Parkinson's disease and progressive supranuclear palsy. *Journal of Neuroscience Methods*, 30(2), 222–230. <https://doi.org/10.1016/j.jneumeth.2013.11.016>
- Shim, R. S., & Compton, M. T. (2020). The social determinants of mental health: Psychiatrists' roles in addressing discrimination and food insecurity. *Focus: a Guide To AIDS Research*, 18(1), 25–30. <https://doi.org/10.1176/appi.focus.20190035>
- Silva, M., Loureiro, A., & Cardoso, G. (2016). Social determinants of mental health: A review of the evidence. *The European Journal of Psychiatry*, 30(4), 259–292, ISSN 0213-6163.
- Singh, C. B. P. (2020). Perceptual threat, social distancing and will-being of people during COVID19 pandemic. *International Journal of Human Potentials Management*, 2(2), 35–50. <http://www.abmjournal.org/index.php/ijhpm/article/view/34>
- Sommerlad, A., Marston, L., Huntley, J., Livingston, G., Lewis, G., Steptoe, A., & Fancourt, D. (2021). Social relationships and depression during the COVID-19 lockdown: Longitudinal analysis of the COVID-19 social study. *Psychological Medicine*, 13, 1–10. <https://doi.org/10.1017/s0033291721000039>
- Stoye, G., Sibieta, L., Rasul, I., Farquharson, C., & Davenport, A. (2020). *The geography of the COVID-19 crisis in England*The institute for fiscal studies. Technical report
- Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry*, 57(2), 113–118. <https://doi.org/10.1080/08039480310000932>
- Tambs, K., & Moum, T. (1993). How well can a few questionnaire items indicate anxiety and depression? *Acta Psychiatrica Scandinavica*, 87(5), 364–7. <https://doi.org/10.1111/j.1600-0447.1993.tb03388.x>
- Vahratian, A., Blumberg, S. J., Terlizzi, E. P., & Schiller, J. S. (2021). Symptoms of anxiety or depressive disorder and use of mental health care among adults during the COVID-19 pandemic - United States, August 2020-february 2021. *MMWR. Morbidity and Mortality Weekly Report*, 70(13), 490–494. <https://doi.org/10.15585/mmwr.mm7013e2>
- Varma, P., Junge, M., Meaklim, H., & Jackson, M. L. (2021). Younger people are more vulnerable to stress, anxiety and depression during COVID-19 pandemic: A global cross-sectional survey. *Progress in Neuro-psychopharmacology & Biological Psychiatry*, 109, 110236. <https://doi.org/10.1016/j.pnpbp.2020.110236>
- Victor, E., Aghajan, Z. M., Sewart, A. R., & Christian, R. (2019). Detecting depression using a framework combining deep multimodal neural networks with a purpose-built automated evaluation. *Psychological Assessment*, 31(8), 1019–1027. <https://doi.org/10.1037/pas0000724>
- Wegner, M. (2021). New approaches to regulatory innovation emerging during the crucible of COVID-19. *Therapeutic Innovation & Regulatory Science*, 55(2), 463–466. <https://doi.org/10.1007/s43441-020-00239-8>
- Wilder, B., Charpignon, M., Killian, J., Ou, H.-C., Mate, A., Jabbari, S., Perrault, A., Desai, A., Tambe, M., & Majumder, M. S. (2020). *The role of age distribution and family structure on Covid-19 dynamics: A preliminary modeling assessment for hubei and lombardy*. Technical report. Harvard John A. Paulson School of Engineering and Applied Science.
- Williams, T., Davis, J., Figueira, C., & Vizard, T. (2021). *Coronavirus and depression in adults, great britain: January to March 2021. Technical report*. Office of National Statistics.
- Young, H. M., & Nesbitt, T. S. (2017). Increasing the capacity of primary care through enabling technology. *Journal of General Internal Medicine*, 32(4), 398–403. <https://doi.org/10.1007/s11606-016-3952-3>
- Yuan, R., Xu, Q. H., Xia, C. C., Lou, C. Y., Xie, Z., Ge, Q. M., & Shao, Y. (2020). Psychological status of parents of hospitalized children during the COVID-19 epidemic in China. *Psychiatry Research*, 288(1), 112953. <https://doi.org/10.1016/j.psychres.2020.112953>
- Zamarro, G., & Prados, M. J. (2021). Gender differences in couples' division of childcare, work and mental health during COVID-19. *Reviews on Environmental Health*, 19(1), 11–40. <https://doi.org/10.1007/s11150-020-09534-7>

- Zandifar, A., Badrfam, R., Yazdani, S., Arzaghi, S. M., Rahimi, F., Ghasemi, S., Khamisabadi, S., Mohammadian Khonsari, N., & Qorbani, M. (2020). Prevalence and severity of depression, anxiety, stress and perceived stress in hospitalized patients with COVID-19. *Journal of Diabetes & Metabolic Disorders*, 19(2), 1431–1438. <https://doi.org/10.1007/s40200-020-00667-1>
- Zhao, S. Z., Wong, J. Y. H., Wu, Y., Choi, E. P. H., Wang, M. P., & Lam, T. H. (2020). Social distancing compliance under COVID-19 pandemic and mental health impacts: A population-based study. *International Journal of Environmental Research and Public Health*, 17(18), 6692. <https://doi.org/10.3390/ijerph17186692>

Author Biography

Nan Zou Bakkeli is a senior researcher at Consumption Research Norway, Oslo Metropolitan University. She is also a part of the research team at Centre for Research on Pandemics & Society (PANSOC). Bakkeli has a PhD in sociology from the University of Oslo, and her research fields are social inequality, medical sociology, social epidemiology, health studies, comparative welfare studies and statistics.