



Screening for Psychological Distress in Healthcare Workers Using Machine Learning: A Proof of Concept

Steve Geoffrion^{1,2} · Catherine Morse^{1,2} · Marie-Michèle Dufour² · Nicolas Bergeron^{3,4} · Stéphane Guay^{1,3,5} · Marc J. Lanovaz^{1,2}

Received: 21 March 2023 / Accepted: 30 October 2023 / Published online: 16 November 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

The purpose of this study was to train and test preliminary models using two machine learning algorithms to identify healthcare workers at risk of developing anxiety, depression, and post-traumatic stress disorder. The study included data from a prospective cohort study of 816 healthcare workers collected using a mobile application during the first two waves of COVID-19. Each week, the participants responded to 11 questions and completed three screening questionnaires (one for anxiety, one for depression, and one for post-traumatic stress disorder). Then, the research team selected two questions (out of the 11), which were used with biological sex to identify whether scores on each screening questionnaire would be positive or negative. The analyses involved a fivefold cross-validation to test the accuracy of models based on logistic regression and support vector machines using cross-sectional and cumulative measures. The findings indicated that the models derived from the two questions and biological sex accurately identified screening scores for anxiety, depression, and post-traumatic stress disorders in 70% to 80% of cases. However, the positive predictive value never exceeded 50%, underlining the importance of collecting more data to train better models. Our proof of concept demonstrates the feasibility of using machine learning to develop novel models to screen for psychological distress in at-risk healthcare workers. Developing models with fewer questions may reduce burdens of active monitoring in practical settings by decreasing the weekly assessment duration.

Keywords Machine learning · Healthcare workers · Anxiety · Depression · Post-traumatic stress disorder

Introduction

When under strain, healthcare workers (HCWs) may develop a state of sustained or high work-related stress [1–3]. Work-related stress refers to adverse physical and emotional

effects, which occur when work expectations are inconsistent with the available resources and the needs of workers [3, 4]. High exposure to potentially traumatic events coupled with work-related stress may trigger episodes of psychological distress (e.g., anxiety, depression, PTSD) [5–7]. As a result, HCWs may develop mental health disorders that increase sick leaves and turnover rates [8, 9]. Studies conducted in healthcare settings have often underlined the resource scarcity (e.g., human, material) prevalent in these environments before COVID-19, which has only been worsened by its impact [5, 10–15]. This resource scarcity may contribute to psychological distress among HCWs [5, 6, 12, 13, 16]. Hence, the healthcare community needs interventions that work efficiently in targeting anxiety, depression, and PTSD in work settings [14, 17]. In the event of a crisis, providing interventions to all HCWs regardless of their mental health state may lead to a waste of resources [11, 16, 18]. Identifying HCWs most likely to need assistance is therefore critical when implementing effective preventive interventions to avoid staff shortages.

✉ Steve Geoffrion
s.geoffrion@umontreal.ca

¹ Research center of l'Institut universitaire en santé mentale de Montréal (CR-IUSMM), Montréal, Québec, Canada

² School of Psychoeducation, University of Montreal, Université de Montréal Pavillon Marie-Victorin École de psychoéducation, C. P. 6128, succursale Centre-ville, Montréal, Québec H3C 3J7, Canada

³ Department of Psychiatry, Université de Montréal, Montréal, Québec, Canada

⁴ Research Center of Centre Hospitalier de l'Université de Montréal (CRCHUM), Montréal, Québec, Canada

⁵ School of Criminology, Université de Montréal, Montréal, Québec, Canada

In the first months of a crisis, the guidelines of the National Institute of Health and Care Excellence (NICE) suggest using active monitoring as a method to identify individuals at risk of psychological distress [19–21]. However, studies show that continuous screening and active monitoring with traditional methods (e.g., self-report diaries, telephone follow-ups, in-person visits by professionals, supervision by management in the workplace) and questionnaires may be time-consuming and redundant, and may lead to reduced adherence [17, 22]. Ecological momentary assessment of mental health shows some similar limitations when administered daily [23, 24].

One potential solution to this issue involves the use of machine learning algorithms to train novel models to screen for distress. Machine learning may reduce the burden of active monitoring by decreasing the number of questions asked to HCWs. This analytic approach may allow for simultaneous testing of multiple factors and their complex interactions to identify the best performing algorithm [25]. By using machine learning algorithms, researchers can narrow down the number of questions that need to be asked on a weekly basis to more efficiently identify HCWs at risk of anxiety, depression, and PTSD. Reducing the number of questions may increase the likelihood that active monitoring remains a part of HCWs practices [21].

In recent a systematic review of machine learning studies on mental health prediction outcomes, using labeled examples to produce predictions (i.e., supervised learning) emerged as the primary approach [26]. Notably, Chung and Teo (2021) highlighted support vector machine (SVM) models as prominent for their accuracy in anxiety and depression prediction [26]. Le Nguyen et al. (2023) employed SVM models, achieving 99.32% accuracy when simulating distress experiences in HCWs [27]. However, the prevailing focus on cross-sectional methodology using retrospective data collection versus prospective may have introduced a blind spot in mental health research with respect to long term outcomes and trajectory [26].

Given the complexity of machine learning, several concerns have arisen in the research literature [28]. A common challenge is the small sample size, influenced by data collection costs and ethical constraints [26, 27]. In fact, many machine learning studies are still in the early stages of demonstrating their feasibility due to small samples and limited external validation [26]. Data quality also influences machine learning predictability and underscores the need for representative samples, especially in HCW's studies [29]. Additionally, machine learning model performances vary according to data samples and preprocessing impacts, resulting in the need for using multiple models for optimal accuracy [26].

The current study

Given the limitations of current practices, the purpose of this study was to use machine learning to develop and test preliminary models with fewer questions to screen for the risk of developing anxiety, depression, and PTSD in HCWs.

Methods

Study design

We used the data collected from a prospective cohort study through a mobile application, in the province of Quebec, Canada, between May 8, 2020, and January 24, 2021, during the first and the second waves of COVID-19 [30]. The purpose of this original cohort study was to examine the evolution and the trajectory of psychological distress in HCWs during and after the first wave of COVID-19 [30]. Participants were prompted to complete several questionnaires through the Ethica app on a weekly basis for a period of 12 weeks. A notification reminded the participants to complete their assessment each week. Weekly reports remained confidential and voluntary. The research ethics board of the CRCHUM approved the research project. Each participant provided informed consent before their participation. The original study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines [31]. We applied a fivefold cross-validation approach to train and test our machine learning models using the data from the original prospective study.

Participants

A total of 816 HCWs from eight healthcare centers in Quebec, Canada, participated in the prospective cohort study. The number of respondents varied each week (i.e., not all participants responded to the questionnaire). No participants responded more than 12 times, with 39.9% of HCW responding 10 times or more, 28.7% between five and nine times, and 31.4% fewer than five times. Our sample involved HCWs from different sectors of health and social services, including hospitals (71%), long-term care (8%), and local community services centres (11%). We invited all workers to participate in the study regardless of the position they held within their organization. Our research protocol excluded participants on sick leave, for a reason unrelated to COVID-19 at the time of recruitment. Note that we did not include age beyond the initial assessment, as the maximum duration of participation was limited to five weeks. We removed any

participants with missing values (features or labels) before conducting our analyses.

Measures

The mobile application used for data collection included questions about psychological distress, which were quantified through anxiety, depression, and PTSD. We measured these indicators through the French versions of the General Anxiety Disorder-7 (7-items; GAD-7; range 0 to 21; cut-off score = 10) [32], the Patient Health Questionnaire (9-items; PHQ-9; range 0 to 27; cut-off score = 11) [33], and the short form of the Post-Traumatic Stress Disorder Checklist for *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition (8 items; PCL-5; range 0 to 32; cut-off score = 13) [34]. For each measure, the participant was asked to respond about the state of their mental health in the past seven days. If participants scores exceeded the clinical threshold, a message appeared at the end of the questionnaire to encourage them to contact support resources. These measures served as outcomes for the supervised machine learning algorithms.

As part of the prospective study, each participant had to respond to 11 additional questions related to their contact and concerns with COVID-19, their perceived support, their quality of life, and their perceived level of stress (see Supplementary material in Fig. A for original questions). To develop a model that required the least effort from the HCWs, we selected two of those questions to use as features (i.e., predictors) in our models. To select the two questions, we first removed the questions relating to COVID-19 ($n = 5$) so that our models could be used beyond the context of a pandemic. Second, our analysis involved identifying the remaining questions ($n = 6$) with the highest correlations, on average, with the three outcome measures anxiety ($M = 5.60$, $SD = 4.41$), depression ($M = 6.16$, $SD = 4.89$) and PTSD ($M = 6.61$, $SD = 5.88$) (see Correlation Matrix in Table 1).

Of the three remaining questions with the highest correlations, two involved quality of life and were highly correlated. Thus, we removed the quality of life question with the lower correlation and were left with two questions overall. The two questions were: (Q1) In the last seven days, what was your level of stress at work?, and (Q2) When reflecting on your life over the last seven days, how would you rate your personal quality of life? Each question involved a 10-point scale from lowest (1) to highest (10). The purpose was to examine whether responses to these two questions could identify scores on the GAD-7, PHQ-9, and PCL-5.

Analyses

To compare the cross-sectional and cumulative measures of anxiety, depression, and PTSD, our analyses applied two machine learning algorithms: logistic regression and support vector machines. Logistic regressions involve using the sigmoid function; it is akin to a linear regression, but its output is a probability from 0 to 1. Support vector machines allow the classification of the data in a nonlinear fashion by projecting them in a higher dimension and using a hyperplane to produce the separation between the binary labels. We chose to test two different algorithms to strengthen the conclusions that could be drawn from our analyses.

For both the logistic regression and support vector machine, we generally used the default values provided by the sklearn package in Python. More specifically, our logistic regression involved an L2 penalty term, a C regularization parameter of 1, a tolerance for stopping of 0.0001, the inclusion of an intercept, and the Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm as a solver. We set the number of maximum iterations to 200 and balanced our weight classes to minimize skew in the error patterns. Our support vector machine involved the following parameters: a C regularization parameter of 1, a radial basis

Table 1 Correlation Matrix of original six questions

Variables	Q1	Q2	Q3	Q4	Q5	Q6	GAD7	PHQ9	PCL5
Q1	1	-0.182	-0.34554	-0.08675	-0.08172	-0.12252	0.405083	0.342313	0.341443
Q2	-0.182	1	0.488477	0.349839	0.205723	0.217198	-0.45561	-0.52642	-0.46976
Q3	-0.34554	0.488477	1	0.226308	0.293329	0.434976	-0.43832	-0.46696	-0.45499
Q4	-0.08675	0.349839	0.226308	1	0.520047	0.425601	-0.18304	-0.23218	-0.20282
Q5	-0.08172	0.205723	0.293329	0.520047	1	0.571722	-0.17934	-0.20968	-0.20199
Q6	-0.12252	0.217198	0.434976	0.425601	0.571722	1	-0.1859	-0.21538	-0.21406
GAD7	0.405083	-0.45561	-0.43832	-0.18304	-0.17934	-0.1859	1	0.801238	0.798219
PHQ9	0.342313	-0.52642	-0.46696	-0.23218	-0.20968	-0.21538	0.801238	1	0.778129
PCL5	0.341443	-0.46976	-0.45499	-0.20282	-0.20199	-0.21406	0.798219	0.778129	1

Q1 In the last seven days, what was your level of stress at work?, *Q2* When reflecting on your life over the last seven days, how would you rate your personal quality of life?, *Q3* When reflecting on your life over the last seven days, how would you rate your professional quality of life?, *Q4* In the last 7 days, how available do you feel your friends and family were to offer you support?, *Q5* In the last 7 days, how available do you feel your colleagues were to offer you support?, *Q6* In the last 7 days, how available do you feel your organization was to offer you support?

function kernel, and a gamma value of 1 divided by the product of the number of features and the variance of x . As with the linear regression, we also used balanced class weights.

The three predictor variables, or features, for our models, were the response to the two questions (i.e., work stress and quality of life) and biological sex for each participant. The three outcomes, or labels, were binary variables extracted from the GAD-7, PHQ-9, and PCL-5. To screen for potential anxiety problems, the authors recommend using a cut-off of 10 [32]. Hence, a score of 10 or more was considered as a positive result, when screening for anxiety, whereas a score less than 10 was labelled as a negative result. Similarly, the cut-off value was set at 13, when screening for depression, as it has been recommended as a cut-off by some more recent studies [35, 36]. In this case, a value of 13 or more was scored as positive, when screening of depression, whereas a value of less than 13 was categorized as a negative result. Finally, we also set the cut-off at 13 for the PCL-5 as recommended by Price et al. [34].

Our analyses involved testing one cross-sectional and four cumulative models for each outcome (for a total of 10 models per algorithm). The cross-sectional model used the data from time 1 only as predictors and as outcomes. The initial cumulative model used the sum of predictors of times 1 and 2 as input and the outcomes at time 2 as output. Correspondingly, the next cumulative model used the sum of the predictors of times 1 to 3 and the outcomes at time 3. Our analyses applied the same procedures to times 4 and 5, whereas all preceding measures were included in the predictors, but only the last time point was used as an outcome. This manipulation allowed us to examine whether cumulative scores provided better screening for anxiety, depression, and PTSD.

For each model, our analyses produced four measures to examine the adequacy of the results: accuracy, sensitivity, specificity, and positive predictive value. Accuracy measured agreement between the outcome values produced by the model and the true outcome values by dividing the number of agreements by the total number of samples. Sensitivity, also referred to as the true positive rate, involves dividing the number of true positives correctly identified by the model by the total number of true positives. Specificity, also referred to as the true negative rate, computes the opposite: the number of true negatives correctly identified by the model by the total number of true negatives. Positive predictive value identifies the number of positive cases correctly identified by the model by dividing the number of true positives by the sum of true positives and negatives. We created confusion matrices to analyze the total error rate and its distribution (see Supplementary material in Figs. B et C).

A risk to consider with machine learning is overfitting the data, which would result in a failure of the models to generalize novel data. To prevent this issue, our analyses involved a fivefold cross-validation, wherein the models

were trained on 80% of the data and tested on the remaining 20%. Using a fivefold cross-validation balanced our need to keep the maximum amount of data in the training set while keeping at least 100 samples in our test set. The latter was necessary to control the margin of error of our outcome measures (e.g., accuracy, specificity). This process was conducted five times so that each sample (i.e., individual) was in the test set exactly once. Our results section reports the means across folds for given model parameters. We conducted all our analyses using Python (version 3.7) with the scikit-learn machine learning package (version 0.23.2). The raw (anonymized) data and code are available freely in an online repository at: https://osf.io/3ey8x/?view_only=c5a41aceaeaa4061995900fe7ac14631.

Results

Table 2 presents the profile of the cohort at each time point. The female to male ratio, the age and the number of years of experience, remained consistent across each time point, suggesting that participant loss was random. The scores on the GAD7, the PHQ9, and PCL5 all decreased over time. Two potential explanations for this observation were that participant distress decreased over time or that those participants who were distressed were more likely to drop out of the study. To identify an explanation, we conducted a post hoc analysis wherein we compared changes in scores for participants that were included in both times 1 and 5. The mean within-participant change in scores for each of the three measures closely matched the ones observed in our Table. This observation indicates that our participants dropped out at random, but that distress did decrease over time in those who remained in the study.

Table 3 presents the four outcome measures for the models using logistic regressions (upper half) and support vector machines (lower half). For the logistic regression, the most accurate and sensitive model to screen for anxiety as measured by GAD-7 was cross-sectional. That is, cumulatively measuring anxiety did not appear to improve the screening for anxiety using the two questions and biological sex. Consistently, the most sensitive model was also the cross-sectional when applying the support vector machine. However, the support vector machine produced the highest overall accuracy and specificity with three weeks of cumulative measurements. The problem is that sensitivity was at its lowest in this model. Given the importance of sensitivity for screening tools, these results suggest that cumulative measurement does not produce more adequate screening for anxiety.

In contrast, cumulative models were better at screening depressive symptoms (i.e., PHQ-9) than cross-sectional models. Notably, the logistic regression model had over

Table 2 Sociodemographic profile of the cohort at each time point according to sex, healthcare centers, age, years of experience, scores of the two questions selected and scores of the GAD7, PHQ9 and PCL5

Variables	Time Point				
	T1 (n=816)	T2 (n=736)	T3 (n=693)	T4 (n=635)	T5 (n=592)
Sexe (n)					
Female	720	648	611	560	521
Male	96	88	82	75	71
Agency (n)					
1	171	160	156	147	136
2	204	176	165	150	142
3	41	40	37	33	31
4	89	84	80	75	71
5	125	115	110	101	98
6	93	82	79	73	71
7	18	17	15	14	14
8	75	62	51	42	29
Age	39.8 (10.1)	39.5 (9.9)	39.7 (9.9)	39.8 (9.8)	39.9 (9.9)
Yrs of Experience	12.6 (9.2)	12.7 (9.1)	12.8 (9.1)	12.9 (9.1)	13.0 (9.2)
Q1	5.2 (2.7)	5.3 (2.6)	5.1 (2.6)	5.1 (2.6)	5.2 (2.6)
Q2	6.4 (2.1)	6.6 (2.0)	6.7 (2.0)	6.7 (2.1)	6.7 (2.0)
GAD7	6.5 (4.7)	5.8 (4.3)	5.4 (4.3)	5.1 (4.2)	5.0 (4.2)
PHQ9	7.1 (5.2)	6.4 (4.8)	6.0 (4.9)	5.6 (4.6)	5.3 (4.6)
PCL5	8.1 (6.4)	7.0 (5.8)	6.2 (5.5)	5.8 (5.5)	5.5 (5.5)

Yrs Years, *Agency (n)* participant per healthcare center, *Q1* In the last seven days, what was your level of stress at work?, *Q2* When reflecting on your life over the last seven days, how would you rate your personal quality of life?, Mean(SD)

0.80 accuracy, sensitivity, and specificity only when a cumulative 3-week measures were used as predictors. Table 1 indicates that a consistent pattern was observable with support vector machines. Specifically, the cumulative 3-week measures produced the best accuracy and sensitivity to screen for depression screening in the workers. Yet, having more weekly measures did not further improve the measures. Moreover, the positive predictive value decreased with more measures for both models, which is an issue as it would increase the number of HCWs that need to be further tested. The 5-week cumulative predictors led to the worst screening with logistic regression, and similar results to the cross-sectional model were found for the support vector machines.

We observed similar patterns when using the PCL-5 to screen for PTSD. The most accurate model overall was the cumulative measure at three weeks for both algorithms, but the cross-sectional model followed closely behind and had much better positive predictive values. The worst performing models were the cumulative five weeks of the logistic regression and the cumulative two weeks for the support vector machines. Taken together, these results underline the limited utility of using cumulative weekly measures to screen for anxiety, but support a 3-week running measure for potentially detecting depression and PTSD in workers.

Discussion

The results of this study showed that our preliminary models accurately screened for anxiety, depression, and post-traumatic stress disorder in 70% to 80% of cases using two questions and biological sex. These results highlight the ability of machine learning to reduce questions to screen for psychological distress, from 24 items for the GAD-7, PHQ-9 and PCL-5, to two questions and biological sex. Our comparison of cross-sectional and cumulative measures also revealed that the cross-sectional model provides the best balance between accuracy and predictive positive values.

The dynamic interaction of the three items included in the model may explain the observed results. The first question assessed the level of subjective work-related stress on a 10-point scale. Stress underlies anxiety, depression, and PTSD; anxiety is a result of anticipatory stress, depression develops through chronic stress and PTSD is life-threatening reaction to stress [37–39]. As such, this question may capture the influence of a transdiagnostic symptom. The second question related to quality of life may act as a protective factor by moderating different work-related stress components. A higher quality of life seems to impart an increased resistance to negative consequences of stress, much like a buffer, which provides to an individual a more secure distance from

Table 3 Accuracy, Sensitivity and Specificity in Identifying a Severe Score on the GAD-7, PHQ-9, and PCL-5 Using Logistic Regressions and Support Vector Machines: Training Results

Predictors	GAD-7			PHQ-9			PCL-5					
	Accuracy	Sensitivity	Specificity	PPV	Accuracy	Sensitivity	Specificity	PPV	Accuracy	Sensitivity	Specificity	PPV
<i>Logistic Regression Algorithm</i>												
Time 1 only (n = 816)	0.743	0.797	0.726	0.484	0.76	0.784	0.757	0.374	0.726	0.744	0.724	0.443
Times 1 and 2 (n = 736)	0.726	0.71	0.727	0.349	0.772	0.78	0.769	0.303	0.712	0.716	0.71	0.318
Times 1 to 3 (n = 693)	0.739	0.74	0.739	0.331	0.82	0.821	0.818	0.34	0.734	0.737	0.734	0.293
Times 1 to 4 (n = 635)	0.735	0.719	0.739	0.282	0.737	0.761	0.736	0.214	0.715	0.681	0.721	0.243
Times 1 to 5 (n = 592)	0.679	0.695	0.677	0.266	0.737	0.728	0.737	0.196	0.696	0.673	0.699	0.214
<i>Support Vector Machine</i>												
Time 1 only (n = 816)	0.733	0.764	0.722	0.472	0.723	0.807	0.711	0.34	0.722	0.752	0.719	0.438
Times 1 and 2 (n = 736)	0.739	0.707	0.742	0.361	0.762	0.79	0.757	0.295	0.697	0.695	0.693	0.301
Times 1 to 3 (n = 693)	0.772	0.659	0.794	0.359	0.801	0.835	0.796	0.318	0.743	0.701	0.747	0.297
Times 1 to 4 (n = 635)	0.746	0.731	0.75	0.294	0.798	0.691	0.81	0.253	0.702	0.657	0.708	0.229
Times 1 to 5 (n = 592)	0.642	0.706	0.631	0.244	0.73	0.75	0.728	0.194	0.716	0.767	0.71	0.243

PPV Positive predictive value

the stress itself [1, 40, 41]. Lastly, biological sex contribute to the model inasmuch as women are at higher risk of developing anxiety, depression, and PTSD [42]. In sum, these three items allow the screening of psychological distress while considering biological sex differences.

One of the implications of this study is that it may be possible to mitigate attrition of active monitoring by reducing the number of questions and the assessment duration to screen for psychological distress. Not only could this decrease the burden on organizations to identify at-risk workers, but it would also take less time away from workers who need to focus on patient services, especially during a crisis. Such an approach may also allow more rapid identification of workers in distress, as early interventions are more likely to prevent the emergence of mental health problems (in contrast to late interventions) [43]. Finally, proper screening allows for a better distribution and organization of resources instead of systematically offering aid to all HCWs, which may avoid wasting limited resources.

This research has limitations that should be noted. The main limitation of our analyses is that we did not conduct hyperparameter tuning for our models, which could have further improved their accuracy. Concerns with overfitting and the size of our sample guided our decision not to apply hyperparameter tuning with the current dataset. A second limitation is that participants did not respond to all measurement points. Considering that the HCWs participated in the context of the COVID-19 pandemic, we recognize that the reality of workers made it difficult to take part in active monitoring consistently. The third limitation lies in the fact that various sectors within healthcare services, including hospitals, long-term care facilities, and local community services centers, each present unique challenges that could potentially confound our model's results. This limitation is underscored by the clustered nature of the data, which comprises workers from 8 distinct Quebec healthcare centers, where the influence of internal management and environmental factors could lead to correlated residuals among the workers. However, it remains that data was compiled with the intention of limiting the number of questions. A fifth limitation is the self-reported nature of the data, where social desirability and self-representation distortion may produce bias. However, the use of a mobile app in this context reduced the probability of bias because of its anonymous nature. Finally, the positive predictive models remained too low for our models to be adopted in work settings. Further tuning with more data needs to be conducted to produce models that produce higher percentages.

That said, our proof of concept clearly shows that machine learning is a promising tool to develop more efficient screening procedures for psychological distress at work. These results provide further support for the potential of machine learning [44–46]. Our machine learning approach could pave

the way for future research to develop more accurate models. To increase the power and positive prediction values of our models, one avenue may be to increase the range in our response scale. For example, passing from 0–10 to 0–15 might increase the specificity and sensitivity of our models. Combined with a larger dataset, machine learning algorithms may eventually train models capable of identifying the correct cut-off score in a higher proportion of cases.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10916-023-02011-5>.

Authors' Contributions S.G. and M.L supervised the project, and collaborated with C.M. to write the main manuscript text. M.-M.B. contributed in terms of data management and analysis. N.B. and S.G. were responsible for conceptualization and received funding for the project. All authors reviewed the manuscript.

Funding The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research project was supported by a grant (#2020–2023-COVID19-PSOv2a-51476) from the Fonds de Recherche du Québec—Ministère de l'Économie et Innovation as well as a salary award (#268274) from the Fonds de Recherche du Québec—Institut de Recherche Robert Sauvé en Santé et Sécurité au Travail to the last author.

Data Availability Raw data and code are available freely in an online repository, the URL is provided in the methods section of the manuscript.

Declarations

Ethical Approval The research ethics board of the CRCHUM approved the research project. The original study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines [31].

Consent to Participate Each participant provided informed consent before their participation.

Consent to Publish Each participant provided informed consent to publish results.

Competing Interests The Authors declare that there is no conflict of interest.

References

1. S. Johnson, C. Cooper, S. Cartwright, I. Donald, P. Taylor, and C. Millet, "The experience of work-related stress across occupations," *J. Manag. Psychol.*, vol. 20, no. 2, pp. 178–187, Jan. 2005, <https://doi.org/10.1108/02683940510579803>.
2. K. Santarone, M. McKenney, and A. Elkbuli, "Preserving mental health and resilience in frontline healthcare workers during COVID-19," *Am. J. Emerg. Med.*, vol. 38, no. 7, pp. 1530–1531, Jul. 2020, <https://doi.org/10.1016/j.ajem.2020.04.030>.
3. P. Trivellas, P. Reklitis, and C. Platis, "The Effect of Job Related Stress on Employees' Satisfaction: A Survey in Health Care," *Proc. 2nd Int. Conf. Integr. Inf. IC-ININFO 2012 Bp. Hung. August 30 – Sept. 3 2012*, vol. 73, pp. 718–726, Feb. 2013, <https://doi.org/10.1016/j.sbspro.2013.02.110>.

4. E. Moustaka and T. C. Constantinidis, "Sources and effects of Work-related stress in nursing," *Health Sci. J.*, vol. 4, no. 4, p. 7, 2010.
5. J. Lai *et al.*, "Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019," *JAMA Netw. Open*, vol. 3, no. 3, pp. e203976–e203976, Mar. 2020, <https://doi.org/10.1001/jamanetworkopen.2020.3976>.
6. A. Lasalvia *et al.*, "The Sustained Psychological Impact of the COVID-19 Pandemic on Health Care Workers One Year after the Outbreak-A Repeated Cross-Sectional Survey in a Tertiary Hospital of North-East Italy," *Int. J. Environ. Res. Public Health*, vol. 18, no. 24, 2021, <https://doi.org/10.3390/ijerph182413374>.
7. F. C. Troglia da Silva and M. L. R. Neto, "Psychiatric disorders in health professionals during the COVID-19 pandemic: A systematic review with meta-analysis," *J. Psychiatr. Res.*, vol. 140, pp. 474–487, Aug. 2021, <https://doi.org/10.1016/j.jpsychires.2021.03.044>.
8. L. N. Dyrbye *et al.*, "Burnout Among Health Care Professionals: A Call to Explore and Address This Underrecognized Threat to Safe, High-Quality Care," *NAM Perspect.*, Jul. 2017, <https://doi.org/10.31478/201707b>.
9. A. Zaka, S. E. Shamloo, P. Fiorente, and A. Tafuri, "COVID-19 pandemic as a watershed moment: A call for systematic psychological health care for frontline medical staff," *J. Health Psychol.*, vol. 25, no. 7, pp. 883–887, Jun. 2020, <https://doi.org/10.1177/1359105320925148>.
10. H. Blake, F. Bermingham, G. Johnson, and A. Tabner, "Mitigating the Psychological Impact of COVID-19 on Healthcare Workers: A Digital Learning Package," *Int. J. Environ. Res. Public Health*, vol. 17, no. 9, Art. no. 9, Jan. 2020, <https://doi.org/10.3390/ijerph17092997>.
11. W. El-Hage *et al.*, "Les professionnels de santé face à la pandémie de la maladie à coronavirus (COVID-19) : quels risques pour leur santé mentale ?," *L'Encéphale*, vol. 46, no. 3, Supplement, pp. S73–S80, Jun. 2020, <https://doi.org/10.1016/j.encep.2020.04.008>.
12. P. Galanis, I. Vraka, D. Fragkou, A. Bilali, and D. Kaitelidou, "Nurses' burnout and associated risk factors during the COVID-19 pandemic: A systematic review and meta-analysis," *J. Adv. Nurs.*, vol. 77, no. 8, pp. 3286–3302, Aug. 2021, <https://doi.org/10.1111/jan.14839>.
13. A. E. Muller *et al.*, "The mental health impact of the covid-19 pandemic on healthcare workers, and interventions to help them: A rapid systematic review," *Psychiatry Res.*, vol. 293, p. 113441, 2020, <https://doi.org/10.1016/j.psychres.2020.113441>.
14. W. Wilson *et al.*, "Prevalence and Predictors of Stress, anxiety, and Depression among Healthcare Workers Managing COVID-19 Pandemic in India: A Nationwide Observational Study," *Indian J. Psychol. Med.*, vol. 42, no. 4, pp. 353–358, 2020, <https://doi.org/10.1177/0253717620933992>.
15. D. Zaçe, I. Hoxhaj, A. Orfino, A. M. Viteritti, L. Janiri, and M. L. D. Pietro, "Interventions to address mental health issues in healthcare workers during infectious disease outbreaks: A systematic review," *J. Psychiatr. Res.*, vol. 136, pp. 319–333, 2021, <https://doi.org/10.1016/j.jpsychires.2021.02.019>.
16. J. Maben 1 2, 3, and J. Bridges, "Covid-19: Supporting nurses' psychological and mental health," *J. Clin. Nurs.*, vol. 29, no. 15–16, pp. 2742–2750, 2020, <https://doi.org/10.1111/jocn.15307>.
17. A. C. McFarlane, P. Williamson, and C. A. Barton, "The impact of traumatic stressors in civilian occupational settings," *J. Public Health Policy*, vol. 30, no. 3, pp. 311–327, Sep. 2009, <https://doi.org/10.1057/jphp.2009.21>.
18. M. Luo, L. Guo, M. Yu, W. Jiang, and H. Wang, "The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public – A systematic review and meta-analysis," *Psychiatry Res.*, vol. 291, p. 113190, Sep. 2020, <https://doi.org/10.1016/j.psychres.2020.113190>.
19. T. Halms, M. Strasser, M. Kunz, and A. Hasan, "How to Reduce Mental Health Burden in Health Care Workers During COVID-19?—A Scoping Review of Guideline Recommendations," *Front. Psychiatry*, vol. 12, 2022, Accessed: Jun. 27, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fpsy.2021.770193>
20. S. Rose, J. Hartnett, and S. Pillai, "Healthcare worker's emotions, perceived stressors and coping mechanisms during the COVID-19 pandemic," *PLOS ONE*, vol. 16, no. 7, p. e0254252, Jul. 2021, <https://doi.org/10.1371/journal.pone.0254252>.
21. National Institute for Health and Care Excellence, "Post-traumatic stress disorder," p. 54, Dec. 2018, [Online]. Available: <https://www.nice.org.uk/guidance/ng116/resources/posttraumatic-stress-disorder-pdf-66141601777861>
22. A. Ehlers *et al.*, "A Randomized Controlled Trial of Cognitive Therapy, a Self-help Booklet, and Repeated Assessments as Early Interventions for Posttraumatic Stress Disorder," *Arch. Gen. Psychiatry*, vol. 60, no. 10, pp. 1024–1032, Oct. 2003, <https://doi.org/10.1001/archpsyc.60.10.1024>.
23. D. S. Krantz and A. S. Baum, *Technology and Methods in Behavioral Medicine*. Psychology Press, 1998.
24. S. Shiffman and M. Hufford, "Ecological momentary assessment," *Appl. Clin. Trials*, vol. 10, Jan. 2008.
25. J. J. McArdle and G. Ritschard, *Contemporary issues in exploratory data mining in the behavioral sciences*. in Contemporary issues in exploratory data mining in the behavioral sciences. New York, NY, US: Routledge/Taylor & Francis Group, 2014, pp. xxii, 473.
26. J. Chung and J. Teo, "Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges," *Appl. Comput. Intell. Soft Comput.*, vol. 2022, p. e9970363, Jan. 2022, <https://doi.org/10.1155/2022/9970363>.
27. K. Le Nguyen, H. Thi Trinh, T. T. Nguyen, and H. D. Nguyen, "Comparative study on the performance of different machine learning techniques to predict the shear strength of RC deep beams: Model selection and industry implications," *Expert Syst. Appl.*, vol. 230, p. 120649, Nov. 2023, <https://doi.org/10.1016/j.eswa.2023.120649>.
28. D. Bzdok and A. Meyer-Lindenberg, "Machine Learning for Precision Psychiatry: Opportunities and Challenges," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, vol. 3, no. 3, pp. 223–230, Mar. 2018, <https://doi.org/10.1016/j.bpsc.2017.11.007>.
29. T. Ching *et al.*, "Opportunities and obstacles for deep learning in biology and medicine," *J. R. Soc. Interface*, vol. 15, no. 141, p. 20170387, Apr. 2018, <https://doi.org/10.1098/rsif.2017.0387>.
30. M.-M. Dufour, N. Bergeron, A. Rabasa, S. Guay, and S. Geoffrion, "Assessment of Psychological Distress in Health-care Workers during and after the First Wave of COVID-19: A Canadian Longitudinal Study: Évaluation de la Détresse Psychologique Chez Les Travailleurs de la Santé Durant et Après la Première Vague de la COVID-19: une étude longitudinale canadienne," *Can. J. Psychiatry*, vol. 66, no. 9, pp. 807–814, Sep. 2021, <https://doi.org/10.1177/070674372111025217>.
31. E. von Elm *et al.*, "The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies," *J. Clin. Epidemiol.*, vol. 61, no. 4, pp. 344–349, Apr. 2008, <https://doi.org/10.1016/j.jclinepi.2007.11.008>.
32. R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe, "A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7," *Arch. Intern. Med.*, vol. 166, no. 10, pp. 1092–1097, May 2006, <https://doi.org/10.1001/archinte.166.10.1092>.
33. K. Kroenke, R. L. Spitzer, and J. B. Williams, "The PHQ-9: validity of a brief depression severity measure," *J. Gen. Intern. Med.*, vol. 16, no. 9, pp. 606–613, Sep. 2001, <https://doi.org/10.1046/j.1525-1497.2001.016009606.x>.
34. M. Price, D. D. Szafranski, K. van Stolk-Cooke, and D. F. Gros, "Investigation of abbreviated 4 and 8 item versions of the PTSD Checklist 5," *Psychiatry Res.*, vol. 239, pp. 124–130, May 2016, <https://doi.org/10.1016/j.psychres.2016.03.014>.
35. C. Beard, K. J. Hsu, L. S. Rifkin, A. B. Busch, and T. Björgvinsson, "Validation of the PHQ-9 in a psychiatric sample," *J. Affect. Disord.*,

- vol. 193, pp. 267–273, Mar. 2016, <https://doi.org/10.1016/j.jad.2015.12.075>.
36. A. S. Moriarty, S. Gilbody, D. McMillan, and L. Manea, “Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis,” *Gen. Hosp. Psychiatry*, vol. 37, no. 6, pp. 567–576, Nov. 2015, <https://doi.org/10.1016/j.genhosppsych.2015.06.012>.
 37. N. H. Kalin, “The Critical Relationship Between Anxiety and Depression,” *Am. J. Psychiatry*, vol. 177, no. 5, pp. 365–367, May 2020, <https://doi.org/10.1176/appi.ajp.2020.20030305>.
 38. B. N. Rusli, B. A. Edimansyah, and L. Naing, “Working conditions, self-perceived stress, anxiety, depression and quality of life: A structural equation modelling approach,” *BMC Public Health*, vol. 8, p. 48, Feb. 2008, <https://doi.org/10.1186/1471-2458-8-48>.
 39. Mardi J. Horowitz, *Stress Response Syndromes: PTSD, Grief, Adjustment, and Dissociative Disorders, 5th Edition*. Accessed: Oct. 03, 2022. [Online]. Available: <https://universitypressofamerica.com/ISBN/9780765710079/Stress-Response-Syndromes-PTSD-Grief-Adjustment-and-Dissociative-Disorders-5th-Edition>
 40. S. M. Horowitz, R. T. Blackburn, D. W. Edington, and D. M. Kloss, “Association between Job Stress and Perceived Quality of Life,” *J. Am. Coll. Health*, vol. 37, no. 1, pp. 29–35, Jul. 1988, <https://doi.org/10.1080/07448481.1988.9939038>.
 41. E. Kabir-Mokamelkhah, A. Bahrani-Ahmadi, and N. Aghili, “Work-related stress and quality of life among Iranian blue-collar workers with self-reported low back pain,” *Med. J. Islam. Repub. Iran*, vol. 30, p. 474, Dec. 2016, Accessed: Oct. 03, 2022. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5419238/>
 42. N. Breslau, R. C. Kessler, H. D. Chilcoat, L. R. Schultz, G. C. Davis, and P. Andreski, “Trauma and posttraumatic stress disorder in the community: the 1996 Detroit Area Survey of Trauma,” *Arch. Gen. Psychiatry*, vol. 55, no. 7, pp. 626–632, Jul. 1998, <https://doi.org/10.1001/archpsyc.55.7.626>.
 43. R. A. Bryant, A. C. McFarlane, D. Silove, M. L. O’Donnell, D. Forbes, and M. Creamer, “The Lingering Impact of Resolved PTSD on Subsequent Functioning,” *Clin. Psychol. Sci.*, vol. 4, no. 3, pp. 493–498, May 2016, <https://doi.org/10.1177/2167702615598756>.
 44. T. A. Prout, “Identifying Predictors of Psychological Distress During COVID-19: A Machine Learning Approach,” *Front. Psychol.*, vol. 11, p. 14, 2020.
 45. D. A. Adler, F. Wang, D. C. Mohr, and T. Choudhury, “Machine learning for passive mental health symptom prediction: Generalization across different longitudinal mobile sensing studies,” *PLOS ONE*, p. 20.
 46. K. Aafjes-van Doorn, C. Kamsteeg, J. Bate, and M. Aafjes, “A scoping review of machine learning in psychotherapy research,” *Psychother. Res.*, vol. 31, no. 1, pp. 92–116, Jan. 2021, <https://doi.org/10.1080/10503307.2020.1808729>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.