

Prediction of Global Psychological Stress and Coping Induced by the COVID-19 Outbreak: A Machine Learning Study

ABSTRACT

Background: Artificial intelligence and machine learning have enormous potential to deal efficiently with a wide range of issues that traditional sciences may be unable to address. Neuroscience, particularly psychiatry, is one of the domains that could potentially benefit from artificial intelligence and machine learning. This study aims to predict *Stress* and assess *Coping* with stress mechanisms during the COVID-19 pandemic and, therefore, help establish a successful intervention to manage distress.

Methods: COVIDiSTRESS global survey data was used in this study and comprised 70 652 respondents after pre-processing. Binary classification is performed for predicting *Stress* and *Coping* with stress, while 2 ensemble machine learning algorithms, deep super learner and cascade deep forest, and state-of-the-art methods are explored for classification. Correlation attribute evaluation is used for feature significance. Statistical analysis, such as Cronbach's alpha, demographic statistics, Pearson's correlation coefficient, independent sample *t*-test, and 95% CI, is also performed.

Results: Globally, females, the younger population, and those in COVID-19 risk groups are observed to possess higher levels of *stress*. *Trust*, *Loneliness*, and *Distress* are found to be the primary predictors of *Stress*, whereas the significant predictors for *coping* with stress are identified as *Social Provision*, *Extroversion*, and *Agreeableness*. Deep super learner and cascade deep forest outperformed the state-of-the-art methods with an accuracy of up to 88.42%.

Conclusions: By comparing different classifiers, we can conclude that multi-layer ensemble outperforms all. Another aim of this study, is the ability to regulate demographic and negative psychological states with a goal of medical interventions and to work towards building multiple coping strategies to reduce stress and promote resilience and recovery from COVID-19.

Keywords: COVID-19, stress, coping, public health, machine learning

Introduction

The unanticipated eruption of COVID-19 in December 2019 forced people around the world to make drastic changes to their daily lives. To limit the spread of the virus and the overloading of healthcare systems, national governments imposed strict rules and restrictions on freedom of movement either through lockdowns or mandatory isolation. In addition to the fear of contracting the virus, people around the world experienced increased levels of stress and anxiety due to money worries, social distancing, and isolation amid orders to stay at home.¹

A crisis like the COVID-19 pandemic is a high stressor as it is a unique and rare event with a potentially fatal effect on health, encountered both directly or indirectly through friends and family and accompanied by societal constraints.^{2,3} Early identification of stress-vulnerable groups is crucial to initiate early treatment aimed at minimizing stress-borne psychological

Neha Prerna Tigga¹ 

Shruti Garg² 

¹Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, India

²Department of Computer Science and Engineering, Birla Institute of Technology, Mesra, India

Corresponding author:

Neha Prerna Tigga
✉ phdcs10052.18@bitmesra.ac.in

Received: December 7, 2021

Accepted: March 4, 2022

Cite this article as: Tigga NP, Garg S. Prediction of global psychological stress and coping induced by the COVID-19 outbreak: A machine learning study. *Alpha Psychiatry*. 2022;23(4):193-202.



Copyright©Author(s) - Available online at alpha-psychiatry.com.
Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

consequences. Therefore, identifying coping strategies is important. Coping is a mechanism for handling stress. In the literature, various studies have established a link between effective coping with stress and lower psychological distress in COVID-19 times.⁴⁻⁶ These studies mostly employ statistical tools to assess data.⁷ Machine learning (ML) techniques are becoming popular in clinical psychology and psychiatry in identifying vulnerable groups. Early helps determine the treatment needed for alleviating stress-related psychological consequences.⁸ Therefore, ML models are found to be suitable for predicting pandemic-induced *Stress* and identifying *Coping* with stress traits. Although *Stress* and *Coping* with stress have been assessed in several research papers, the studies were confined to certain groups and regions.⁹⁻¹¹ Research has not been carried out on a global population in terms of predicting *Stress* and *Coping* with stress mechanisms during the COVID-19 pandemic. ML models are particularly suited because they allow for large data samples.¹²

Machine learning techniques, such as boosting and decision tree (DT), were used to identify stress levels among working IT professionals.¹³ Similarly, various basic ML techniques were applied to predict levels of anxiety, depression, and stress in a modern lifestyle.¹⁴ However, in certain traditional ML models, like support vector machine (SVM), performance decreases when using a large amount of data and execution time increases compared to other ML models.¹⁵ Thus, 2 ensemble ML models, i.e., Deep Super Learner (DSL)¹⁶ and Cascade Deep Forest (CDF),¹⁷ were applied to predict *Stress* and *Coping* with stress on a global data sample.

This study first employed statistical methods to infer relationships between the variables. Subsequently, ML analysis was performed to make predictions, the same was being done in previous research.²⁶ The workflow of this research is shown in Figure 1.

The objectives of this study are as follows.

Statistical tools are employed to

1. Identify significant factors influencing a high-risk individual's *Stress*.
2. Identify significant attributes that contribute toward *Coping* with stress.

Application of ensemble ML methods yields the following

1. Prediction of *Stress* during the COVID-19 pandemic among a global population.
2. Prediction of *Coping* with stress during the COVID-19 pandemic among a global population.

MAIN POINTS

- Impact of sociodemographic and psychological variables on *Stress* and *Coping* with stress in the pandemic environment on global survey data is assessed by a statistical toolkit.
- Causative factors for *Stress* and coping with stress are extracted using correlation attribute evaluation.
- Prediction of high-risk *Stress* individuals using the ensemble machine learning method.
- Prediction of *Coping* with stress attributes using the ensemble machine learning method.

Methods

Dataset Acquisition

The dataset used is a publicly available global survey dataset. It was collected by Yamada et. al. in 2021.¹⁸ The entire survey, which may be extracted from <https://osf.io/mhszp/>, was conducted between March 30, 2020, and May 30, 2020, and included n=173 426 participants across 179 countries. This cross-cultural survey aimed to identify the psychological impact of COVID-19 on participants.

Measures and Their Scales

Perceived Stress Scale (PSS): This is a psychometric tool for estimating the impression of stress. It estimates the degree to which an individual's circumstances are assessed as being stressful. Participants were asked 10 questions relating to the prevalence of stressful situations on a 5-point scale ranging from 1 (never) to 5 (often). Participants were divided into 3 categories of perceived stress: low, moderate, and high. Perceived Stress Scale is also referred to as *Stress* in this study.

Short 15-Item Big Five Inventory (BFI-S): The BFI-S is a popular concept that uses 15 items for expressing the 5 most significant aspects of personality: openness to experience (OTE), conscientiousness (CON), extroversion (EXT), agreeableness (AGR), and neuroticism (NEU). Each personality trait is measured by 3 items, and the scores range from 1 (never) to 6 (often).

Short Self-Reported Scale of Loneliness (SLON): This is a short version of the University of California, Los Angeles loneliness scale, which aims to measure loneliness and social isolation. It consists of 3 questions based on a 3-point scale where 1= hardly ever and 3= often.

Social Provision Scale (SPS): It is a 10-item questionnaire with a 6-point scale ranging from "strongly disagree" to "strongly agree" and is a shortened version of the 24-item social provision scale. It is intended to measure the perception of social support. By reducing the SPS from 24 to 10 questions, researchers can create a more reliable, timely, and valid method for assessing the availability of social support.

Distress: This is a 24-item questionnaire with a 7-point scale ranging from "strongly disagree" to "strongly agree," which is used to measure distress. It includes several key items relating to people's distress and fear during the coronavirus pandemic (e.g., accessibility to utilities, job loss, adapting job, schooling, social connections on online platforms, and the societal strains of isolation with children).

Coping: This is a 16-item questionnaire with a 6-point scale ranging from "strongly disagree" to "strongly agree" that measures the effectiveness of coping mechanisms during the COVID-19 outbreak (e.g., maintaining social contact, keeping updated, devoting one's time to preparation, hobbies, and spirituality).

Corona Concern (CC): This is a 5-item questionnaire with a 6-point scale ranging from "strongly disagree" to "strongly agree" for measuring the degree of concern relating to the repercussions of COVID-19 (e.g., concern for yourself, family, friends, the country and globally).

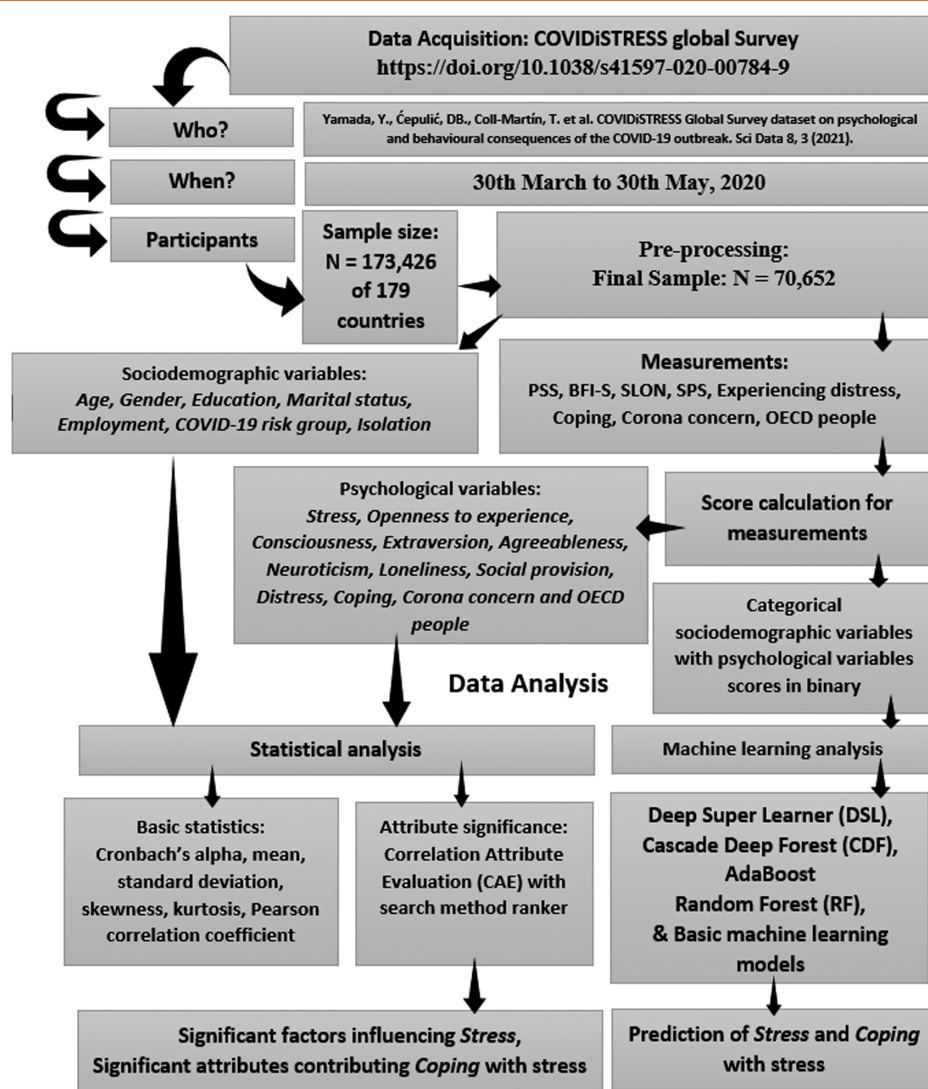


Figure 1. Workflow of research.

Organisation for Economic Cooperation and Development (OECD) People: This has 2 items with a scale ranging from 0 to 10 for measuring the degree of interpersonal trust based on the 2017 OECD guidelines (e.g., trust in most people, trust in people you know), where 0 is “not at all” and 1 is “completely.” *OECD people* is referred to as “trust” in this study.

Correlation Attribute Evaluation for Attribute Importance

Attribute importance refers to methods of calculating a score for each ML model's input variable; the scores describe the significance of each variable. A higher score indicates that a certain feature will have a greater impact on the model used to forecast a given variable. Correlation Attribute Evaluation (CAE)¹⁹ with a search method ranker was used. A CAE evaluates the worth of an attribute by measuring Pearson's correlation coefficient.

Machine Learning Method

Deep Super Learner: Super learning is a collection of algorithms that determines the best combination. Deep Super Learner is a method for achieving log loss and accuracy outcomes that are comparable

with deep neural networks (DNN) while using basic ML algorithms in a unified framework. With strong performance across diverse tasks using similar hyper-parameter settings, the DSL is robust, adaptive, and straightforward to train. Classical ML uses fewer hyper-parameters, provides more transparency in its findings, and has a faster convergence rate on smaller datasets. According to test findings, the DSL outperforms individual base learners, single-layer ensembles, and DNNs in some circumstances.¹⁶ With task-specific customization, the DSL's performance can be improved even further. Figure 2 shows a graphical representation of DSL.

Cascade Deep Forest: Cascade Deep Forest, also known as gcForest, is a new DT ensemble ML approach.¹⁷ This method employs a deep-forest ensemble with a cascade structure to facilitate representation learning. The number of cascade stages can be selected flexibly, allowing model intricacy to be automatically configured and for gcForest to function well, even with smaller datasets. GcForest has fewer hyper-parameters than DNNs and its efficiency is robust to hyper-parameter configurations. Across most circumstances, with default settings, it can process data from several domains and produce outstanding results.²⁰

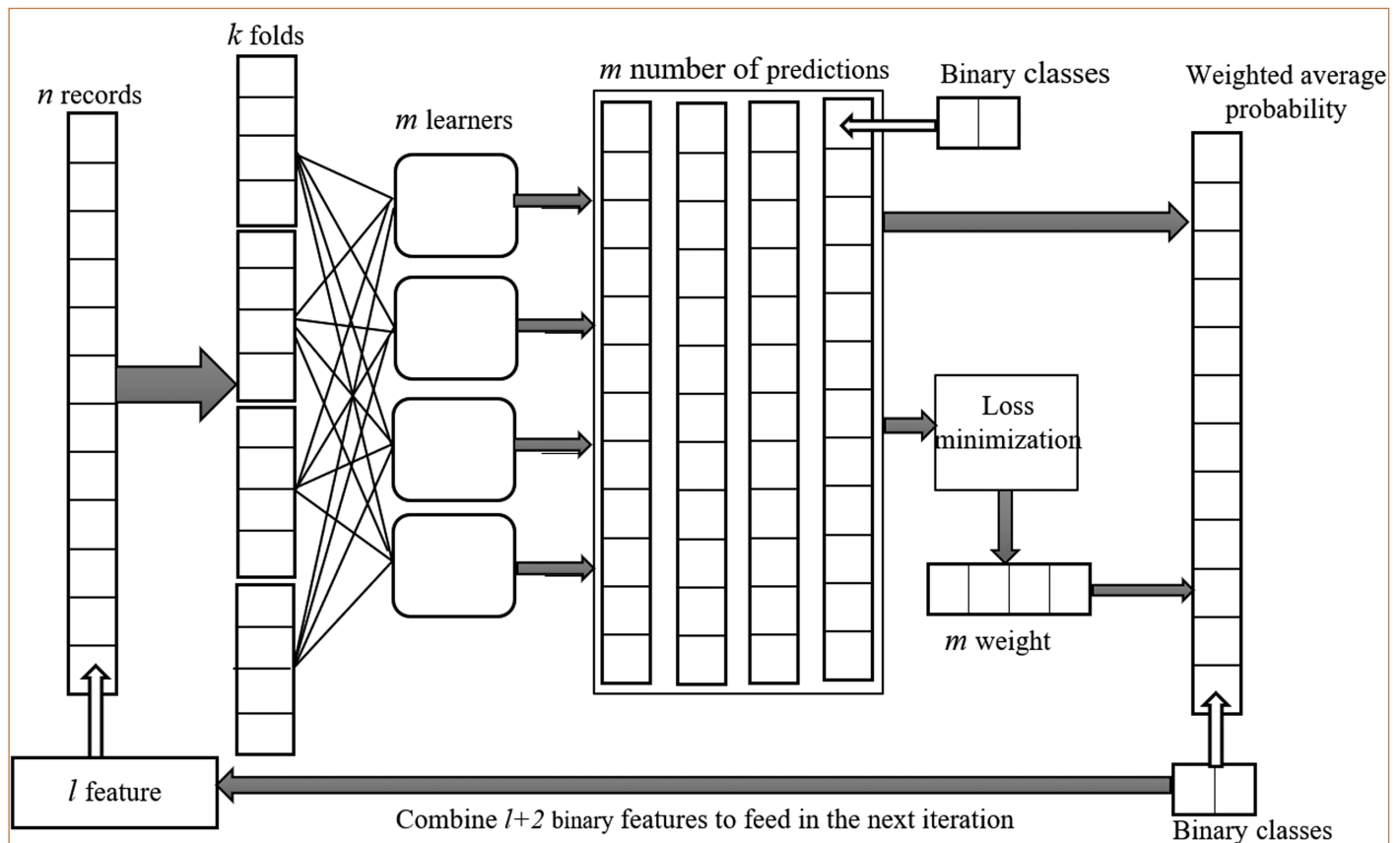


Figure 2. Overall approach followed in this study for the application DSL.

While DNN analyses raw features layer after layer, gcForest uses a cascade architecture (Figure 3), whereby each level of the cascade has features extracted by the level before it and transmits its processed output to the next level. Every random forest has 500 trees, and the candidate in this study was selected at random as a d set of features, while the segmentation was chosen from the feature with the best Gini score. The tree was constructed by selecting random features to split at every node, and the tree expanded until every leaf node exclusively contained instances of the same class or no more than 10 instances. A hyper-parameter was the number of trees for each forest. As this experiment involved binary classification, every forest output was a 2-dimensional class vector that was connected to the input feature to depict the next original input. The class vectors created from each forest were built using k-fold cross-validation to limit the danger of overfitting.

Experiments and Results

Dataset Preparation

To maintain consistency of the dataset, cases that provided incomplete and incorrect information were removed from the dataset. Therefore, $n=70\,652$ cases post-data-filtering are assessed in this study. The resulting dataset consisted of sociodemographic and psychological variables. The sociodemographic variables included Age, Gender, Education, Marital Status, Employment, Covid-19 risk group, and Isolation. The psychological variables included PSS, OTE, CON, EXT, AGR, NEU, SLON, SPS, Distress, Coping, CC, and OECD people (for more details, see Supplementary Table S1).

The scores were calculated for all psychological variables by averaging the scores from each question in the questionnaire. The BFI-S score was computed by taking the average scores of the 3 questions for each personality type. The average was then grouped into 2 classes: high and low (see Supplementary Table S2).

Sociodemographic variables like Age, Education, and Marital status were pre-processed as follows. The age variable was continuous and was categorized as "young," "middle-aged," and "old." Education and Marital status were merged to form 3 categories for each scale (see Supplementary Table S1). The final dataset dimension was $70\,652 \times 19$, where 70 652 represents the number of individual responses with 19 attributes each. The attributes are Age, Gender, Education, Marital Status, Employment, COVID-19 risk group, Isolation, PSS, OTE, CON, EXT, AGR, NEU, SLON, SPS, Distress, Coping, CC, and OECD people.

Data Analysis

Statistical Analysis

This study consists of 70 652 global population data aged 18-110, with a mean age of 38.57 (standard deviation (SD)=13.33). All of the statistical analysis on sociodemographic and psychological descriptive data was performed in SPSS. First, Cronbach's alpha was calculated to check the reliability and consistency of each measurement scale (Table 1). Almost all the scales are under the acceptable range of internal consistency. The normality was then checked using skewness and kurtosis statistics after assuming the distribution is normal. Although the skewness and kurtosis values in a normal distribution are both zero,

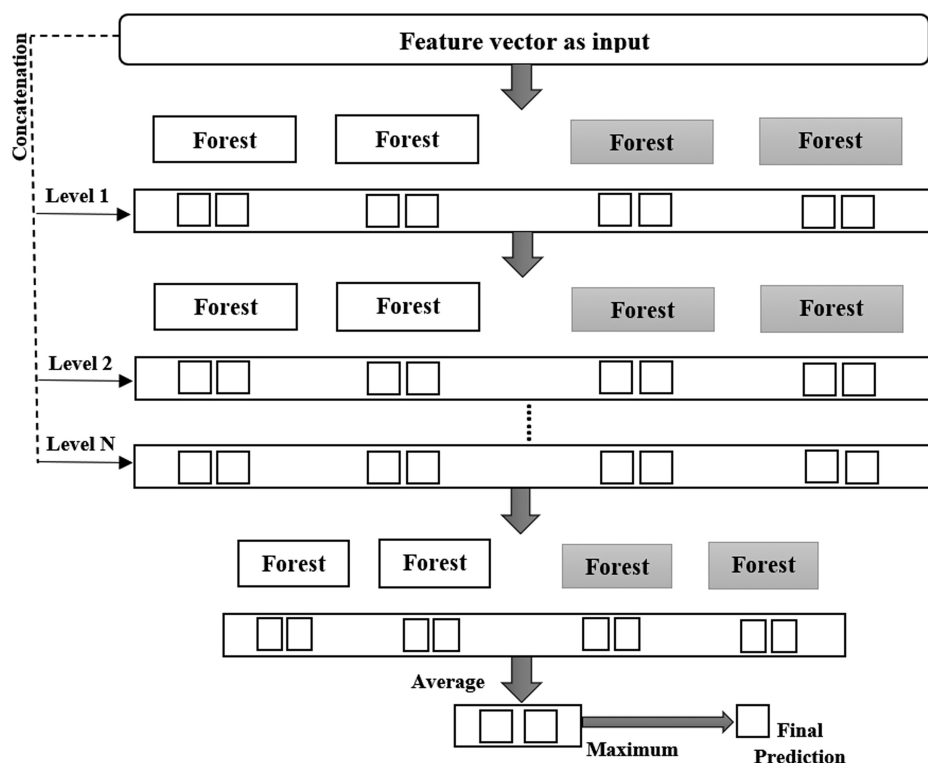


Figure 3. The structure of a cascade forest is seen in this graphic. Assume that every level of the cascade has 2 entirely random tree forests (in gray) and 2 random forests (in white). Here we have binary classes to predict, so every forest will generate a 2-dimensional class vector, which will be concatenated to re-represent the original input data.

skewness and kurtosis values between -2 and $+2$ are acceptable for psychometric applications. In this study, the skewness (min = -1.191 and max = 0.346) and kurtosis (min = -0.559 and max = 1.190) values are within the permissible range of -2 to $+2$ (Table 2).

Pearson's correlation coefficient determines the association between *Stress*, *coping* with stress and the independent variables (Tables 3 and 4). *Stress* has a positive low to moderate correlation with *SLON* ($r=0.40$, $P < .001$), *Distress* ($r=0.39$, $P < .001$) and *NEU* ($r=0.28$, $P < .001$). *Coping* with stress has a positive low to moderate

correlation with *SPS* ($r=0.33$, $P < .001$), *AGR* ($r=0.21$, $P < .001$), *EXT* ($r=0.20$, $P < .001$) and *CC* ($r=0.20$, $P < .001$). Significance level α is taken as 0.001 . An independent sample *t*-test is carried out to check the significance between groups for *Stress* and *Coping* with stress and their independent variables. It was determined that all the independent variables were statistically significant ($P < .001$) with regard to *Stress* and *Coping* with stress. The complete breakdown of the dataset regarding "high" and "low" *Stress* and *Coping* with stress is reported in Supplementary Table S1. The "others" category from gender was excluded for statistical interpretation because of a lack of specificity, even though it is included for ML training.

Table 1. Cronbach Alpha for all the Measurements

Characteristics	Cronbach's Alpha
PSS	0.312
OTE	0.666
CON	0.612
EXT	0.766
AGR	0.560
NEU	0.708
SLON	0.777
SPS	0.920
Distress	0.872
Coping	0.747
CC	0.818
OECD people	0.754

PSS, Perceived Stress Scale; OTE, Openness to Experience; CON, Conscientiousness; EXT, Extraversion; AGR, Agreeableness; NEU, Neuroticism; SLON, Short Self-reported Scale of Loneliness; SPS, Social Provision Scale; CC, Corona Concern; OECD, Organisation for Economic Cooperation and Development.

Attribute Significance

The Weka toolkit was used to find important attributes for *Stress* and *Coping* with stress using CAE with a search method ranker, the results of which are shown in Figure 4. For predicting *Stress*; *OECD people*, *Loneliness*, *Distress*, *Isolation*, and *Neuroticism* were the top 5 predictor variables, while *Social provision*, *Extraversion*, *Agreeableness*, *Corona concern*, and *Openness to experience* are the top 5 predictors for *coping* with stress.

Machine Learning Analysis

The entire ML study is carried out on Google Colab Pro using Python. The dataset is divided into a ratio of $80:20$. The 2 ML models, DSL and CDF, are used to predict *Stress* and *Coping* with stress.

Most of the attributes in the dataset are imbalanced (Table 5). This occurs when the distribution between the classes is biased or skewed. The proportion can range from a little skewed to a significant

Table 2. Descriptive Statistics of the Study Subjects (n = 70 652)

Characteristics	Minimum	Maximum	Mean (SD)		Skewness	Kurtosis
PSS	1	5	2.9798	0.39828	-0.132	0.842
OTE	1	6	4.5063	0.92017	-0.516	0.042
CON	1	6	4.3339	0.87265	-0.326	-0.093
EXT	1	6	3.9203	1.12530	-0.228	-0.559
AGR	1	6	4.4293	0.81654	-0.444	0.144
NEU	1	6	3.3281	1.05101	0.073	-0.469
SLON	1	5	2.5628	0.98940	0.346	-0.521
SPS	1	6	4.9152	0.83890	-1.191	1.950
Distress	1	7	3.7486	0.86819	-0.141	-0.137
Coping	1	6	3.7623	0.63704	-0.236	0.585
CC	1	6	4.5968	0.93562	-0.814	0.817
OECD people	0	10	6.7680	1.67174	-0.834	0.682

PSS, Perceived Stress Scale; OTE, Openness to Experience; CON, Conscientiousness; EXT, Extraversion; AGR, Agreeableness; NEU, Neuroticism; SLON, Short Self-reported Scale of Loneliness; SPS, Social Provision Scale; CC, Corona Concern; OECD, Organisation for Economic Cooperation and Development.

Table 3. Pearson's Correlation Coefficient for High and Low Composition of Stress with Respect to 18 Independent Variables

Independent Variables	High Stress	Low Stress	Pearson's Correlation		95% CI for the Difference	
	Mean (SD)	Mean (SD)	Coefficient	P	Lower	Upper
Age	38.21 (13.14)	41.09 (14.31)	-0.15	<.001 ^a	-3.17	-2.58
Openness to experience (OTE)	4.54 (0.89)	4.21 (1.01)	0.16	<.001 ^a	0.31	0.35
Conscientiousness (CON)	4.35 (0.86)	4.15 (0.91)	0.09	<.001 ^a	0.18	0.22
Extraversion (EXT)	3.94 (1.11)	3.75 (1.17)	0.06	<.001 ^a	0.16	0.21
Agreeableness (AGR)	4.44 (0.80)	4.32 (0.85)	0.04	<.001 ^a	0.09	0.13
Neuroticism (NEU)	3.38 (1.04)	2.93 (1.03)	0.28	<.001 ^a	0.42	0.46
Loneliness (SLON)	2.64 (0.98)	2.00 (0.81)	0.40	<.001 ^a	0.61	0.66
Social provision (SPS)	4.94 (0.81)	4.72 (0.98)	0.05	<.001 ^a	0.19	0.23
Distress	3.80 (0.85)	3.33 (0.85)	0.39	<.001 ^a	0.45	0.49
Coping	3.78 (0.62)	3.60 (0.69)	0.10	<.001 ^a	0.16	0.19
Corona concern (CC)	4.63 (0.92)	4.31 (0.99)	0.20	<.001 ^a	0.30	0.34
OECD people	6.75 (1.66)	6.83 (0.72)	-0.09	<.001 ^a	-0.11	-0.03

^aPearson's correlation coefficient $P < .001$

Table 4. Pearson's Correlation Coefficient for High and Low Composition of Coping with Respect to 18 Independent Variables

Independent Variables	High Coping	Low Coping	Pearson's Correlation		95% CI for the Difference	
	Mean (SD)	Mean (SD)	Coefficient	P	Lower	Upper
Age	38.44 (13.36)	39.53 (13.05)	-0.04	<.001 ^a	-1.39	-0.78
Stress (PSS)	2.98 (0.38)	2.91 (0.40)	0.10	<.001 ^a	0.06	0.07
Openness to experience (OTE)	4.54 (0.89)	4.23 (1.06)	0.18	<.001 ^a	0.28	0.33
Conscientiousness (CON)	4.36 (0.85)	4.12 (0.96)	0.15	<.001 ^a	0.22	0.26
Extraversion (EXT)	3.98 (1.09)	3.47 (1.21)	0.20	<.001 ^a	0.47	0.52
Agreeableness (AGR)	4.47 (0.79)	4.12 (0.92)	0.21	<.001 ^a	0.33	0.36
Neuroticism (NEU)	3.32 (1.03)	3.37 (1.16)	-0.01	<.001 ^a	-0.07	-0.02
Loneliness (SLON)	2.55 (0.97)	2.62 (1.11)	-0.05	<.001 ^a	-0.09	-0.048
Social provision (SPS)	4.99 (0.76)	4.35 (1.11)	0.33	<.001 ^a	0.62	0.65
Distress	3.78 (0.85)	3.51 (0.94)	0.18	<.001 ^a	0.24	0.28
Corona concern (CC)	4.64 (0.90)	4.27 (1.08)	0.20	<.001 ^a	0.34	0.38
OECD people	6.85 (1.61)	6.15 (1.93)	0.14	<.001 ^a	0.66	0.73

^aPearson's correlation coefficient $P < .001$

imbalance, with hundreds of instances in the minority class and thousands in the majority class. Because most ML algorithms for classification are created with the notion of an equivalent number

of samples for every class, imbalanced classifications are problematic for predictive analysis. Consequently, models emerge with poor prediction accuracy, particularly for the minority class. This poses a

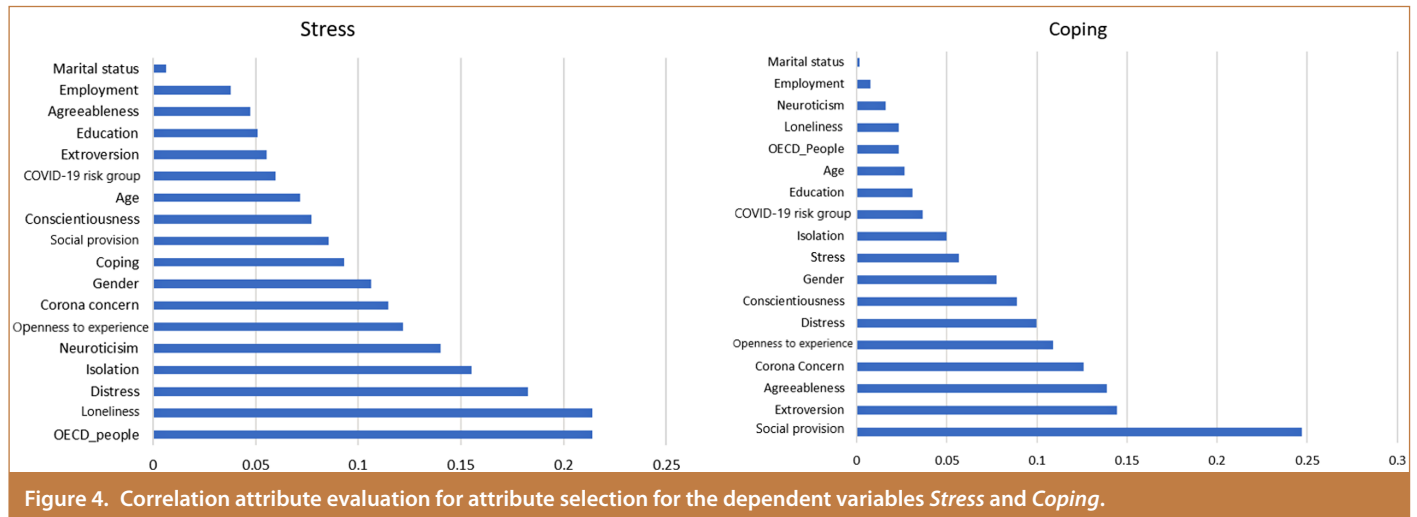


Figure 4. Correlation attribute evaluation for attribute selection for the dependent variables *Stress* and *Coping*.

Table 5. Dataset Obtained after ADASYN Resampling Technique for the Imbalanced Original Dataset on Training Set

	Deep Super Learner (DSL)	Cascade Deep Forest (CDF)
<i>Stress</i>	Original dataset ({High:39500, Low:5716}) Resampled dataset({High:39500, Low:40788})	Original dataset ({High:49438, Low:7083}) Resampled dataset ({High: 49438, Low:50326})
<i>Coping</i>	Original dataset ({High: 49863, Low: 6658}) Resampled dataset ({High:49863, Low:50225})	Original dataset ({High:49880, Low:6641}) Resampled dataset ({High:49880, Low:49942})

ADASYN, Adaptive Synthetic Sampling.

difficulty because the minority class is usually more significant than the majority equivalent, so the problem is more susceptible to classification inaccuracies for the minority class than for the majority. To overcome this problem, an adaptive synthetic sampling oversampling technique was used to balance the classes using the imblearn package of Python. Adaptive synthetic sampling is a simulated data generating technique that has the benefit of not replicating minority data and creating additional data for “difficult to learn” examples.

The hyper-parameters used for the ML methods are provided in Supplementary Table S3. Machine Learning predictive performance, based on the evaluation metric of accuracy, precision, recall, and F1 score is provided in Table 6. The formulas for accuracy, precision, recall, and F1 score are given below:

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}} \quad (1)$$

$$\text{Precision}(P) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

$$\text{Recall}(R) = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (3)$$

$$\text{F1 score}(F1) = 2 * \frac{P * R}{P + R} \quad (4)$$

However, Python programming language computes weighted average values for precision, recall, and F1 score. The formula for weighted average is as follows:

$$\text{weighted avg Precision} = \frac{|y_{class1}|}{|y|} * P_{class1} + \frac{|y_{class2}|}{|y|} * P_{class2} \quad (5)$$

$$\text{weighted avg Recall} = \frac{|y_{class1}|}{|y|} * R_{class1} + \frac{|y_{class2}|}{|y|} * R_{class2} \quad (6)$$

$$\text{weighted avg F1 score} = \frac{|y_{class1}|}{|y|} * F1_{class1} + \frac{|y_{class2}|}{|y|} * F1_{class2} \quad (7)$$

Where, $|y|$ is the total number of testing sample and $|y_{class1}|$ and $|y_{class2}|$ are the samples for each class, “high” and “low,” respectively. Where

$\frac{|y_{class1}|}{|y|}$ and $\frac{|y_{class2}|}{|y|}$ are class weights assigned for each class.

The accuracies obtained by both algorithm DSL and CDF are not much different. The 2 algorithms when compared in terms of execution time, DSL was faster than CDF (Table 6). The accuracies of DSL and CDF methods are also compared with other base learners, i.e., logistic regression, multilayer perceptron, k-nearest neighbor and SVM as well as single-layer ensembles, i.e., AdaBoost and random forest (Figure 5). From Figure 5, it is determined that DSL and CDF outperformed the other classifiers. The average accuracy of the multilayer ensemble (DSL and CDF) was obtained at 88.07% for the test data.

Table 6. Performance Metrics of Ensemble Machine Learning Methods.

Predictor	Time (minutes)	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
Deep Super Learner (DSL)					
Stress	4:54	88.32	86.08	88.32	83.90
Coping	5:34	88.42	86.31	88.42	84.19
Cascade Deep Forest (CDF)					
Stress	8:44	87.14	82.75	87.13	82.56
Coping	7:50	88.39	83.89	88.38	83.76

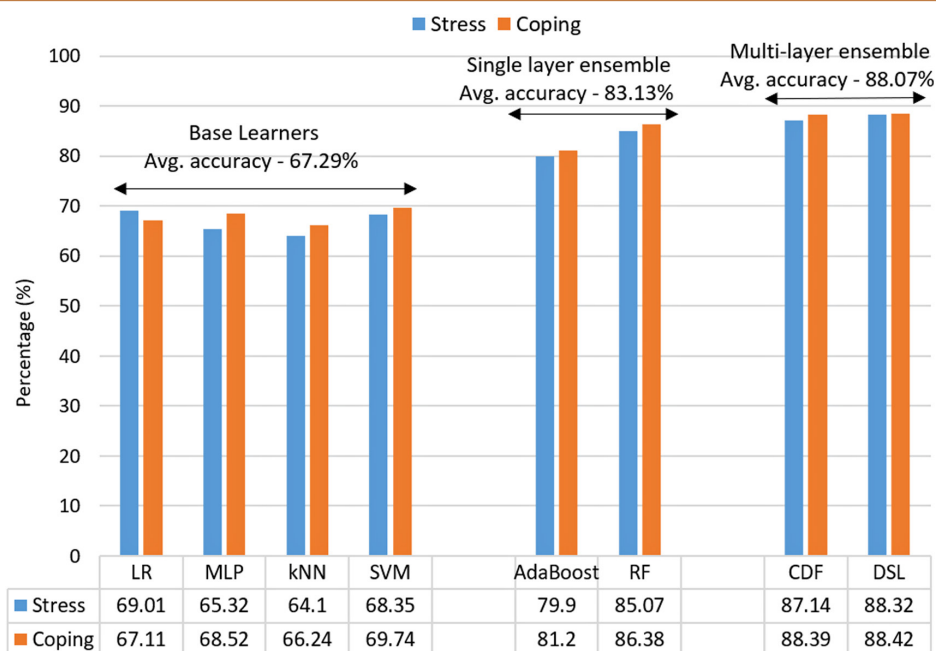


Figure 5. Performance comparison of based learners, single-layer ensembles, DSL, and CDF methods. (LR, Logistic Regression; MLP, Multi-layer Perceptron; kNN, k-Nearest Neighbour; SVM, Support Vector Machine; RF, Random Forest; DSL, Deep Super Learner; CDF, Cascade Deep Forest).

Discussion

This study examined how the COVID-19 crisis has affected levels of stress among a global population along with their coping mechanism. The findings confirmed that individuals viewed the crisis as a stressful event, with the degree of stress in the current sample being higher than that of the overall population in a non-emergency situation. Almost 87.43% of the entire population had high stress, with 88.13% showing high coping mechanisms (see supplementary materials Table S1). These findings are consistent with current research on the psychological effects of COVID-19.^{9,21–23}

Role of Sociodemographic and Psychological Variables in Predicting Stress and Coping with Stress Using Statistical Methods

Statistical analysis revealed that 87.42% of the total population of the dataset belonged to a high-stress level category, while 89.07% of the total young population reported stress levels that were 5–8% higher than the middle-aged and older populations. The same finding was reported by Shanahan *et al.* (2020) and Emery *et al.* (2021).^{24,25} Consistent with other research, females experienced higher levels of stress, but they had better-coping mechanisms compared to males—a finding that aligns with existing research.^{26,27} Education plays an important role in Coping with stress, thus, individuals who completed their education up to degree level or higher had better-coping capabilities during the outbreak. Self-employed individuals had higher Coping with stress compared to others. Those belonging to the COVID-19 high-risk group and isolated showed higher stress levels. (see Supplementary Table S1 for details).

An independent sample *t*-test, Pearson's correlation, and confidence interval statistics are calculated for high and low levels of stress and

coping for all of the psychological variables and are found to be significant at $P < .001$. Stress has a positive moderate to low correlation with the psychological factors' *Loneliness*, *Distress*, and *Neuroticism* (Table 3); these findings are supported by earlier studies.²⁸ *Social provision*, *Extroversion*, *Agreeableness*, and *Corona concern* had positive moderate to low correlations with Coping with stress (Table 4). Agbaria and Mokh (2021) found that positive psychological factors help in Coping.²⁹

Identification of Significant Attributes for Stress and Coping with Stress

From the attribute selection methods, the 5 topmost factors for Stress were *OECD people*, *Loneliness*, *Distress*, *Isolation*, and *Neuroticism* (Figure 4). *OECD people* or *Trust* has been associated with stress in the past, while chronic stress is linked to a reduction in overall trust.³⁰ Other negative factors, such as *Loneliness*, *Distress*, *Isolation*, and *Neuroticism*, can also increase the intensity of stress.³¹ For predicting coping with stress, 5 significant positive factors were *Social provision*, *Extroversion*, *Agreeableness*, *Corona concern*, and *Openness to experience* (Figure 4). The association of *Social provision* has been reported for safeguarding psychological health by Labrague (2021).³²

Role of Sociodemographic and Psychological Variables in Predicting Stress and Coping with Stress Using ML Classification Methods

Considering the high risks inherent in stress, individuals may suffer major psychological issues if it is not dealt with in a timely fashion. This study aimed to develop an intervention to predict Stress and Coping with stress so that those at risk during the pandemic could seek appropriate help. In this regard, ML methods were employed to highlight individuals with "high" and "low" levels of Stress and Coping with stress. DSL and CDF have been used as good ML models for

making predictions across various fields of research.^{33,34} The ensemble methods, DSL and CDF, outperformed the state-of-the-art base learners and single-layer ensembles. The accuracies obtained by DSL and CDF were almost 5% and 20% better than single-layer ensemble and state-of-art base learners, respectively (Figure 5).

Although the existing literature supports the relationships explored in this study, care should be exercised because of the following limitations. When interpreting these findings as significant, it is important to note that the data was gathered from a global population. Therefore, the general global population and diversity factors influenced the outcomes more universally. Further research can be carried out on this global data regarding countries, gender, and age to produce outcomes in a more contextually specific manner.

Another limitation is that the data was collected via an online survey. While this ensured huge samples, sample representativeness was not guaranteed. As a result, extremely vulnerable populations, such as the homeless, the poor, or those with no internet access, may be underrepresented in this study. Also, due to the use of self-reporting assessments, this study was unable to check the reliability of the responses or confirm that the respondents understood the questions correctly. These flaws should be addressed in future studies.

Finally, it has been acknowledged that the COVID-19 pandemic has had a significant influence on mental health. Adopting multiple coping strategies, such as behavioral activation, adaptation-based coping, and mindful and compassionate practices, could help reduce stress while promoting resilience and healing. In the wake of the COVID-19 outbreak, these tactics may be extremely fruitful because they will help people find purpose, develop endurance for distress, improve social assistance, develop a sense of profound psychological connectivity, and adopt target value-driven behaviors.

In conclusion, this study aimed to develop ML models to predict *Stress and Coping* with stress related to COVID-19 using sociodemographic and psychological variables. The process of ML will reduce the need for skill and increase reliance on data to make precise predictions to identify individuals more susceptible to the risk of developing serious psychological issues and to develop timely interventions and support. Additionally, it can be used to create broad data-driven analytical frameworks that can be applied to several domains of interest.

Peer-review: Externally peer reviewed.

Author Contributions: Concept – N.P.T., S.G.; Design – N.P.T.; Supervision – S.G.; Resource – N.P.T., S.G.; Materials – N.P.T., S.G.; Data Processing – N.P.T.; Analysis and/or Interpretation – N.P.T.; Literature Search – N.P.T., S.G.; Writing – N.P.T.; Critical Reviews – S.G.

Acknowledgments: The second author acknowledges research extended by the University of Delhi under Institute of Eminence (IoE).

Declaration of Interests: The authors have no conflicts of interest to declare.

Funding: This work was supported by the University of Delhi under IoE with grant ref. no./IoE/2021/12/FRP.

References

1. Cullen W, Gulati G, Kelly BD. Mental health in the COVID-19 pandemic. *QJM An Int J Med.* 2020;113(5):311-312. [\[CrossRef\]](#)
2. Pierce M, Hope H, Ford T, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. *Lancet Psychiatry.* 2020;7(10):883-892. [\[CrossRef\]](#)
3. Yarıbeygi H, Panahi Y, Sahraei H, Johnston TP, Sahebkar A. The impact of stress on body function: a review. *Excli J.* 2017;16:1057-1072. [\[CrossRef\]](#)
4. Yu H, Li M, Li Z, et al. Coping style, social support and psychological distress in the general Chinese population in the early stages of the COVID-19 epidemic. *BMC Psychiatry.* 2020;20(1):426. [\[CrossRef\]](#)
5. Shechter A, Diaz F, Moise N, et al. Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic. *Gen Hosp Psychiatry.* 2020;66:1-8. [\[CrossRef\]](#)
6. Rahman MA, Islam SMS, Tungpunkom P, et al. COVID-19: factors associated with psychological distress, fear, and coping strategies among community members across 17 countries. *Global Health.* 2021;17(1):117. [\[CrossRef\]](#)
7. Dukart J, Weis S, Genon S, Eickhoff SB. Towards increasing the clinical applicability of machine learning biomarkers in psychiatry. *Nat Hum Behav.* 2021;5(4):431-432. [\[CrossRef\]](#)
8. Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: opportunities and challenges. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2018;3(3):223-230. [\[CrossRef\]](#)
9. Kar N, Kar B, Kar S. Stress and coping during COVID-19 pandemic: result of an online survey. *Psychiatry Res.* 2021;295:113598. [\[CrossRef\]](#)
10. Rose S, Hartnett J, Pillai S. Healthcare worker's emotions, perceived stressors and coping mechanisms during the COVID-19 pandemic. *PLoS One.* 2021;16(7):e0254252. [\[CrossRef\]](#)
11. Patias ND, Von Hohendorff J, Cozzer AJ, Flores PA, Scorsolini-Comin F. Mental health and coping strategies in undergraduate and graduate students during COVID-19 pandemic. *Trends Psychol.* 2021;29:414-433.
12. Kuan PF, Yang X, Clouston S, et al. Cell type-specific gene expression patterns associated with posttraumatic stress disorder in World Trade Center responders. *Transl Psychiatry.* 2019;9(1):1. [\[CrossRef\]](#)
13. Reddy US, Thota AV, Dharun A. Machine learning techniques for stress prediction in working employees. In 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), Madurai, India. IEEE; 2018:1-4.
14. Priya A, Garg S, Tigga NP. Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Comput Sci.* 2020;167:1258-1267. [\[CrossRef\]](#)
15. Karamizadeh S, Abdullah SM, Halimi M, Shayan J, Javad Rajabi M. Advantage and drawback of support vector machine functionality. In 2014 International Conference on Computer, Communications, and Control Technology (I4CT), Langkawi, Malaysia. IEEE; 2014:63-65.
16. Young S, Abdou T, Bener A. Deep super learner: a deep ensemble for classification problems. In *Canadian Conference on Artificial Intelligence*. Springer, Cham. 2018;(84-95). [\[CrossRef\]](#)
17. Zhou ZH, Feng J. Deep forest. *Natl Sci Rev.* 2019;6(1):74-86. [\[CrossRef\]](#)
18. Yamada Y, Čepulić DB, Coll-Martin T, et al. COVIDiSTRESS Global Survey dataset on psychological and behavioural consequences of the COVID-19 outbreak. *Sci Data.* 2021;8(1):1-23. [\[CrossRef\]](#)
19. Gnanambal S, Thangaraj M, Meenatchi VT, Gayathri V. Classification algorithms with attribute selection: an evaluation study using Weka. *Int J Adv Netw Appl.* 2018;9(6):3640-3644.
20. Chu Y, Kaushik AC, Wang X, et al. DTI-CDF: a cascade deep forest model towards the prediction of drug-target interactions based on hybrid features. *Brief Bioinform.* 2021;22(1):451-462. [\[CrossRef\]](#)
21. Babore A, Lombardi L, Viceconti ML, et al. Psychological effects of the COVID-2019 pandemic: perceived stress and coping strategies among healthcare professionals. *Psychiatry Res.* 2020;293:113366. [\[CrossRef\]](#)

22. Chaney C. Family stress and coping among African Americans in the age of COVID-19. *J Comp Fam Stud*. 2020;51(3-4):254-273. [\[CrossRef\]](#)
23. Saravanan C, Mahmoud I, Elshami W, Taha MH. Knowledge, anxiety, fear, and psychological distress about COVID-19 among university students in the United Arab Emirates. *Front Psychiatry*. 2020;11:582189. [\[CrossRef\]](#)
24. Shanahan L, Steinhoff A, Bechtiger L, et al. Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study. *Psychol Med*. 2020;1-10. [\[CrossRef\]](#)
25. Emery RL, Johnson ST, Simone M, Loth KA, Berge JM, Neumark-Sztainer D. Understanding the impact of the COVID-19 pandemic on stress, mood, and substance use among young adults in the Greater Minneapolis-St. Paul area: findings from project EAT. *Soc Sci Med*. 2021;276:113826. [\[CrossRef\]](#)
26. Flesia L, Monaro M, Mazza C, et al. Predicting perceived stress related to the Covid-19 outbreak through stable psychological traits and machine learning models. *J Clin Med*. 2020;9(10):3350. [\[CrossRef\]](#)
27. García-Fernández L, Romero-Ferreiro V, Padilla S, David López-Roldán P, Monzó-García M, Rodríguez-Jimenez R. Gender differences in emotional response to the COVID-19 outbreak in Spain. *Brain Behav*. 2021;11(1):e01934. [\[CrossRef\]](#)
28. Gubler DA, Makowski LM, Troche SJ, Schlegel K. Loneliness and well-being during the Covid-19 pandemic: associations with personality and emotion regulation. *J Happiness Stud*. 2020;22(5):1-20. [\[CrossRef\]](#)
29. Agbaria Q, Mokh AA. Coping with stress during the coronavirus outbreak: the contribution of big five personality traits and social support. *Int J Ment Health Addiction*. 2021;1-9. [\[CrossRef\]](#)
30. Potts SR, McCuddy WT, Jayan D, Porcelli AJ. To trust, or not to trust? Individual differences in physiological reactivity predict trust under acute stress. *Psychoneuroendocrinology*. 2019;100:75-84. [\[CrossRef\]](#)
31. Jurblum M, Ng CH, Castle DJ. Psychological consequences of social isolation and quarantine: issues related to COVID-19 restrictions. *Aust J Gen Pract*. 2020;49(12):778-783. [\[CrossRef\]](#)
32. Labrague LJ. Psychological resilience, coping behaviours and social support among health care workers during the COVID-19 pandemic: a systematic review of quantitative studies. *J Nurs Manag*. 2021;29(7):1893-1905. [\[CrossRef\]](#)
33. Ai Z, Luktarhan N, Zhou A, Lv D. Webshell attack detection based on a deep super learner. *Symmetry*. 2020;12(9):1406. [\[CrossRef\]](#)
34. Liu X, Wang R, Cai Z, Cai Y, Yin X. Deep multigrained cascade forest for hyperspectral image classification. *IEEE Trans Geosci Remote Sensing*. 2019;57(10):8169-8183. [\[CrossRef\]](#)