

Predictive Modeling of Stress in the Healthcare Industry During COVID-19: A Novel Approach Using XGBoost, SHAP Values, and Tree Explainer

Pooja Gupta, DIT University, India*

Srabanti Maji, DIT University, India

Ritika Mehra, Dev Bhoomi Uttarakhand University, India

 <https://orcid.org/0000-0002-2785-5856>

ABSTRACT

There was a substantial medicine shortage and an increase in morbidity due to the second wave of the COVID-19 pandemic in India. This pandemic has also had a drastic impact on healthcare professionals' psychological health as they were surrounded by suffering, death, and isolation. Healthcare practitioners in North India were sent a self-administered questionnaire based on the COVID-19 Stress Scale (N = 436) from March to May 2021. With 10-fold cross-validation, extreme gradient boosting (XGBoost) was used to predict the individual stress levels. XGBoost classifier was applied, and classification accuracy was 88%. The results of this research show that approximately 52.6% of healthcare specialists in the dataset exceed the severe psychiatric morbidity standards. Further, to determine which attribute had a significant impact on stress prediction, advanced techniques (SHAP values), and tree explainer were applied. The two most significant stress predictors were found to be medicine shortage and trouble in concentrating.

KEYWORDS

COVID-19, CSS, Decision Making, Healthcare, Mindfulness, Psychological Well-Being, SHAP, XGBoost

INTRODUCTION

The COVID-19 epidemic is a one-of-a-kind hazard in terms of psychological distress. The psychological stress of healthcare experts was assessed at the time of COVID-19's second wave (March–May 2021). According to research on previous pandemics, stress is a primary driver of behavior (Taylor S, 2019). To date, there has been very less empirical research available on this topic. Fear or anxiety-related signs can be noticed among the general population through a pandemic (Salari

DOI: 10.4018/IJDSST.315758

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

et al., 2021; Rodriguez-Hidalgo et al., 20220; Roy et al., 2020). The available literature on other transmissible diseases, for example, severe acute respiratory syndrome, the Ebola viral disease, and the Middle East Respiratory Syndrome imply that many healthcare workers felt worried, stressed, and despair amidst and after the outbreak, which influenced their ability and psychological health, negatively (Chong et al., 2004; Sim et al., 2004; Tam et al., 2004; Khalid et al., 2016; Lee et al., 2018). COVID-19 has engulfed India's entire healthcare structure. India noted a daily figure topping 0.35 million new cases, with the second wave's peak happening in mid-May 2021 (Ranjan *et al.*, 2020). Throughout the crisis, healthcare workers on the front lines met threats to their bodily and mental health (Lai *et al.*, 2020). Health care practitioners and the general public were facing unprecedented challenges as an outcome of the COVID-19 pandemic. The procurement of appropriate personal protective equipment to protect doctors' and staff's physical health receives a lot of attention, but resources to preserve their mental health are just as vital. Since the well-being of healthcare professionals is of utmost importance and very few studies have been conducted to understand their stress levels as this phenomenon is new (Rajkumar, 2020; Pfefferbaum & North, 2020; Cullen *et al.*, 2020; Usher *et al.*, 2020), therefore, the current study has been planned to measure the damage that the COVID-19 has done to the mental health of healthcare practitioners during the second wave, in India, and therefore what factors play a role in stress. Individual stress levels have been predicted using Extreme Gradient Boosting (XGBoost). SHAP values, provided by (Lundberg and Lee, 2017), were used to give a better understanding of the model's interpretability as this method is an excellent way to reverse-engineer the prediction made by any machine learning algorithm. This paper provides the global interpretability of the model by showing how much each feature has contributed, either positively or negatively, to the prediction of stress. Further local interpretability was calculated by explaining each observation's prediction and contribution of features. The study's findings will have ramifications for healthcare policymakers and practitioners, as they will help in crafting policies to increase well-being, minimize burnout, and unhappiness among healthcare professionals, and therefore improve their ability to properly serve patients.

RELATED WORK

In recent times, some researchers have worked on the detection of stress during COVID-19. Numerous machine learning methods have been used to predict stress. (Eder et al, 2021) collected data from 533 participants over seven weeks using the Perceived Vulnerability to Disease Scale. They applied two machine learning models: LASSO for the linear model and ERT for the non-linear model. The most important factors that contributed to fear were identified as worrying about food shortage and concerning the outbreak and its ramifications. The focus of this study was primarily on fear and perceived health. Another study proposed by (Praveen S.V. et al, 2021) analyzed the feelings of Indian citizens who had worry, stress, and trauma as a result of COVID-19. For this research, they gathered 840000 tweets. COVID-19 lockdown and Death were determined as the most critical variables that induce stress using natural language processing.

(Nooripour, Roghieh, et al, 2021) conducted an online survey during the COVID-19 outbreak to measure the psychological states of the Iranian population and collected the data from 755 people using different scales like The Adult Hope Scale, The Resilience Scale, and Paloutzian & Ellison's Spiritual Well-being Scale. Their findings showed that there is no relation between high spiritual well-being and the amount of stress. A person with high spiritual well-being can have severe stress. Spiritual well-being, hope, and resiliency could be significant stress contributors. In another study (Jha et al., 2021), data from 17764 individuals were collected in the United States. Authors used Bayesian network inference to identify the vital predictor of stress at the time of the pandemic, it was found that work from home, chronic mental illness, and, lack of communication induce mental stress. (Delgado-Gallegos et al., 2020) measured the stress levels of healthcare professionals in Mexico, they collected the data using Covid Stress Scale for 6 weeks, and information from 112

individuals was obtained. Results showed high levels of Xenophobia and Compulsive stress are the most important factors. In (Flesia et al., 2020), data from 2053 Italian individuals was collected using the 10-item Perceived Stress Scale. The authors used a diversity of classic machine learning approaches to predict stress, with Support Vector Machine, Random Forest, and Naïve Bayes, and found that women, those with less income, and persons living in a communal setting had increased levels of anxiety. A comparison of available literature on stress prediction using machine learning is offered in Table 1. This comparison is performed according to the scale used for Data collection, the number of participants, the main features of stress, and the methods used. But no work has used the SHAP explanation for highlighting the individual feature contribution to the prediction made by the machine learning model. This motivated the authors of the current study to dig deep into the machine learning model and find out the degree of each predictor's contribution to the target variable, either positively or negatively.

MATERIALS AND METHODS

Task Definition

Long-term stress is known to cause issues with both physical and mental health. According to a significant amount of pertinent research, "Stress level prediction" is a classification problem that calls for the extraction of potential features from "labeled data" as an input and the construction of a prediction model for "unlabeled data" in conjunction with a classification algorithm. In order to determine what factors, contribute to stress, the current study was designed to assess the harm that COVID-19 has caused to the mental health of healthcare professionals in India during the second wave. The prediction of a person's felt stress can be described as a classification problem, The major goal of this work is to use XGBoost to model individual COVID-19 stress levels into five classes (normal, mild, moderate, severe, and, extremely severe) of stress based on the Covid-19 Stress Scale. The SHAP (Shapley) values technique is added to measure feature contribution because the process by which the input values affect the output is unknown, which is a restriction of machine learning algorithms. By demonstrating how much each attribute has contributed, either favorably or

Table 1. Summarized Related Work

Authors	The scale used for Data Collection	Number of Participants	Main Features of Stress Identified	Method Used
Eder et al, (2021).	Perceived Vulnerability to Disease Scale	533	Worrying about shortage of food supplies	LASSO & ERT (Extremely Randomized Trees)
Praveen S. V. et al, (2021)	NA	840,000 Tweets	Death And Lockdown caused by the COVID-19	Natural Language Processing
Nooripour et al (2021)	The Resilience Scale, The Adult Hope Scale, Paloutzian & Ellison's Spiritual Wellbeing Scale	755	High spiritual well-being does not always have low stress, and this factor can Be a good stress predictor along with hope and resilience.	Adaptive Neuro-Fuzzy Inference System
Jha et al, (2021)	NA	17,764	Working and studying from home, digital communication, and social isolation were identified as contributing factors to mental illness during the COVID-19 pandemic.	Bayesian Network Inference, Inference, Random Forest, SVM: Support Vector Machine
Delgado-Gallegos et al (2020).	COVID-19 Stress Scale	112	High levels of Xenophobia and Compulsive stress	C5.0 Decision Tree Algorithm
Flesia et al (2020)	Perceived Stress Scale (PSS-10)	2053	Women participants with lower incomes, and people living with others all displayed higher levels of discomfort.	SVM, Naïve Bayes, RF

adversely, to the stress prediction, this research gives the model's overall interpretability and feature contribution of each observation.

COVID-19 Stress Scale (CSS)

To measure stress in medical personnel, CSS was used. CSS was designed in May 2020 to identify the stress and to better comprehend and evaluate COVID-19-related mental suffering (Taylor *et al.*, 2020). CSS measures stress by employing 36 elements with a valid 5-factor solution. This instrument may be applied to evaluate stress and anxiety symptoms associated with COVID-19. These symptoms are related to (1) danger and contamination fears, (2) fears about economic consequences, (3) xenophobia (4) compulsive checking and reassurance-seeking, and (5) traumatic stress symptoms. On a scale of 5-point, items were rated from 0 (not at all) to 4 (extremely). Scores will range from 0 to 144. Normal is indicated by a zero (0), mild by a one (1), moderated by a two (2), severe by a three (3), and extremely severe by a four (4). Overall score of less than 50 indicates normal, 50-66 indicates mild, 66-82 indicates moderate, 82-90 indicates severe, and more than 90 indicates extremely severe stress. The scoring system was followed by (Mauder *et al.*, 2006; Almutairi, *et al.*, 2018).

Participants

The proposed study used information from 436 healthcare professionals (250 doctors and 186 nurses). These medical professionals were working for various public and private institutions in north India at the time of the study. (Delhi, Uttar Pradesh, and Uttarakhand). 262 (60.09%) men and 174 (39.90%) women made up the total number of health personnel.

Data Collection

There are many datasets available in the literature, Eder *et al.* (2021) used the Perceived Vulnerability to Disease Scale and collected data from the 533 participants, Nooripour *et al.* (2021) used The Resilience Scale, The Adult Hope Scale, Paloutzian & Ellison's Spiritual Wellbeing Scale and collected data from 755 participants. Delgado-Gallegos *et al.* (2020) used COVID-19 Stress Scale and collected data 112. But all these datasets were prepared using different scales and at different times and since the objective of the proposed study is to measure the stress of healthcare professionals during the second wave of COVID-the 19, the authors decided to collect their data using Covid Stress Scale (CSS). CSS was specially designed in May 2020 to identify the stress during COVID19 and to better comprehend and evaluate COVID-19-related mental suffering.

From March to May 2021, data was gathered via a questionnaire that was distributed as a Google Form among health professionals via different online channels such as Linked In, E-mail, and WhatsApp groups. There were two sections to the questionnaire. The first section contains demographic information, while the second section contains CSS items for stress measurement. A total of 450 responses were collected when the questionnaire was delivered to 1000 medical personnel in north India (Delhi, Uttar Pradesh, and, Uttarakhand). Received responses were double-checked for data correctness. Due to irregularities or incompleteness, a few of the questionnaires were eliminated. Finally, 436 replies were found suitable for data analysis. Participants were requested to give their consent before the study, and they were given the option to leave the survey at any point on the form.

Dataset Description

There are 39 features in the dataset including demographic information and CSS's 36 items. A detailed dataset description is given in table 2.

Extreme Gradient Boost

Algorithm XGBoost is a machine learning method that is both innovative and scalable (Chen and Guestrin, 2016). XGBoost method is applied to measure stress because the sequential addition of new

base-learners to the ensemble, is the fundamental principle of gradient boosting, in contrast to popular ensemble techniques like the random forest, which rely on an averaging of models in the ensemble. By focusing on the training data that are challenging to estimate, the XGBoost approach improves the ensemble model's prediction performance using additive base-learners. XGBoost provides two crucial methods: column subsampling and shrinkage. The shrinkage technique scales the newly added weights, which helps to decrease the impact of each tree and over-fitting. Column subsampling only selects a random subset of input features for a specific tree, as a result, the XGBoost was chosen to accomplish the modeling task.

XGBoost may be used as a solution to classification and regression difficulties (Friedman, 2001). This algorithm's goal is to keep growing a tree by adding new trees and segregating features. XGBoost combines numerous weak classifiers for producing a powerful classifier (Wang, L. et al., 2020).

Shapley Additive Explanations (SHAP)

Machine learning techniques involve training a model on a dataset, after which the model generates predictions. But it is impossible to foresee the importance of traits in producing predictions. Although some attributes look possibly the most significant, the model's perspective may have been influenced by other factors. It's difficult to describe what happened within a model only based on predicted results. IG method can be used with logistic regression, support vector machines, and neural networks. On the other hand, Tree-based algorithms like boosted trees and random forests—make up the majority of non-differentiable model classes. At the leaves, they encode discrete values. These call for the use of a Tree SHAP-style Shapley-value-based technique. A Shapley-value-based approach necessitates computing the model's output on a sizable number of inputs drawn from the enormous subspace of all conceivable feature value combinations. Most machine learning frameworks provide efficient and robust support for computing gradients of differentiable models. But for IG, a differentiable model is necessary. A Shapley-value-based technique, however, does not assume these presumptions. SHAP is an open-source package that verifies whether a machine learning model is dependable. The Shapley values of Lloyd Shapley, which they developed in 1951 as a cooperative game theory solution paradigm, serve as the foundation of SHAP (Roth, A.E., 1988). With SHAP, one can display the worth of features and can also understand how these features affected any machine learning model's ability to predict using a range of visualizations (Lundberg and Lee, 2017). XGBoost, lightgbm, and decision trees all use the Tree Explainer algorithm, which is fast and accurate. Though some tree-based models ascribe results to feature status automatically. This is vital to understand that “feature importance” and “feature contribution” are not the same, the former emphasizes which elements influence model performance, whereas the latter not only emphasizes but also straightly quantifies each feature's role in the prediction outcome. Retrieving the valuable features through the trained model is quite simple because each node in a DT, for example, is a constraint on a single feature that splits the dataset. For classification, either Gini impurity or information gain/entropy is used to determine the locally optimal condition. As a result, the averaged impurity decline from every feature over all the trees in the ensemble may be used to rank feature relevance (Meng *et al.*, 2021). However, the model's ranking of feature relevance is insufficient to explain an individual prediction. It's crucial to understand how each component contributes to the ultimate result, for example. A feature attribution method is used to obtain insight into how individuals measure stress:

$$g(z') = \varphi_0 + \sum_{i=1}^M \varphi_i z'_i \quad (1)$$

where M denotes the overall number of features and g is the explanation model, φ_i is the feature attribution's value of feature i, and $z'_i \in \{0, 1\}$, A 1 in the coalition vector indicates that the associated

Table 2. Dataset Description

#	Feature	Scope	Scale	Abbreviation
1	Gender	M(male), F(female)	NA	NA
2	Age	Age of participant	NA	NA
3	I am worried about catching the virus	0 (not at all) to 4 (extremely)	Danger	worried_catching_virus
4	I am worried that I can't keep my family safe from the virus			family_not_safe
5	I am worried that our healthcare system won't be able to protect my loved ones			healthcare_not_able_to_protect_loved_ones
6	I am worried that our healthcare system is unable to keep me safe from the virus			healthcare_unable_to_keep_safe
7	I am worried that basic hygiene is not enough to keep me safe from the virus			basic_hygiene
8	I am worried that social distancing is not enough			socialdist_not_enough
9	I am worried about grocery stores running out of food		Socio-economic consequences	food_shortage
10	I am worried that grocery stores will close down			grocery_stores_closedown
11	I am worried about a shortage of disinfectant supplies			disinfectant_supplies_shortage
12	I am worried about grocery stores running out of cold/flu remedies			cold_flu_remedies_shortage
13	I am worried about grocery stores running out of water			water_shortage
14	I am worried about pharmacies running out of prescription medicines			medicines_shortage
15	I am worried that foreigners are spreading the virus in my country		Xenophobia	foreigners_spreading_virus_in_country
16	If I went to a restaurant that specialized in foreign foods, I'd be worried about catching the virus			If_went_to_restaurant_worried_about_catching_virus
17	I am worried about coming into contact with foreigners			contact_with_foreigners_because_they_might_virus
18	If I met a person from a foreign country, I'd be worried that they might have the virus			meeting_person_from_foreign_country
19	If I was in an elevator with a group of foreigners, I'd be worried they were infected with virus			foreigner_infect_with_virus_in_elevator
20	I am worried that foreigners are spreading the virus because they're not as clean as we are			foreigners_spreading_virus_because_they_not_clean
21	I am worried that if I touched something in a public space, I would catch the virus		Contamination	touching_something_publicspace
22	I am worried that if someone coughed or sneezed near me, I would catch the virus			coughed_sneezed_near_me
23	I am worried that people around me will infect me with the virus			people_around_infect
24	I am worried about taking change in cash transactions			cash_transactions
25	I am worried that I might catch the virus from handling money or using a debit machine			using_debit_machine
26	I am worried that my mail has been contaminated by mail handlers			mail_handlers
27	I had trouble concentrating because I kept thinking about the virus		Traumatic stress,	trouble_concentrating
28	Disturbing mental images about the virus popped into my mind against my will			disturbing_mental_images
29	I had trouble sleeping because I worried about the virus			trouble_sleeping
30	I thought about the virus when I didn't mean to			thinking_too_much
31	Reminders of the virus caused me to have physical reactions, sweating or a pounding heart			sweating_pounding_heart
32	I had bad dreams about the virus			bad_dreams
33	Searched the Internet for treatments for COVID-19		Compulsive checking	searching_Internet_for_treatments
34	Asking health professional			asking_health_professionals
35	YouTube videos about COVID-19			youTube_videos_about_COVID-19
36	Checking your own body for signs of infection			checking_body_for_infection
37	Seeking reassurance from friends or family about COVID-19			reassurance_from_friends
38	Social media posts concerning COVID-19			social_media_post_concerning_COVID-19
39	Stress	normal, mild, moderate, severe, extremely severe		

Dataset is available at: <https://drive.google.com/file/d/1sblYL5jKPLBYk6jTiiKJuH2twxbDaivn/view?usp=sharing>

feature value is “present,” whereas a 0 indicates that it is “missing”. Further, the calculation of the Shapley values includes only some feature values. The modeling of coalitions as a linear model is a method for computing the Φ ’s. The coalition vector x' for x , the instance of interest, is a vector of all 1s, indicating that all feature values are “present”. Equation (2) shows the formula:

$$g(x') = \varphi_0 + \sum_{j=1}^M \varphi_j \quad (2)$$

The term “feature contribution” was used to describe the value of feature attribution. According to the average feature attribution covering all observations, both a single prediction and the complete model can be interpreted using model. The Shapley value φ_i for a certain feature i (out of total N features) can be expressed as shown in equation (3), p is the prediction given by the model, and a set S containing non-zero indexes for certain feature i (out of total N features):

$$\varphi_i(p) = \sum_{S \in N \setminus \{i\}} \frac{|S|!(M - |S| - 1)!}{M!} [p(SU\{i\}) - p(S)] \quad (3)$$

SHAP has been extended to support tree-based machine learning models, tree SHAP value estimate algorithm was proposed to overcome the difficulty of predicting φ_i in equation (3) with traditional Shapley values. TreeSHAP was designed to be a model-specific replacement for KernelSHAP, but it has since been discovered that it might result in illogical feature attributions. Instead of utilizing the marginal expectation, TreeSHAP uses the conditional expectation to define the value function. There is an issue with the restricted expectation that features that do not impose any impact on the prediction function p may receive a TreeSHAP estimate other than zero (Sundararajan *et al.*, 2019; Janzing *et al.*, 2019). When a feature is linked with another feature that has an impact on the prediction, a non-zero estimate can occur.

SHAP Interaction Values

Because certain features become more predictive when combined with others, evaluating the interactions between them is an issue that cannot be overlooked. The contribution of feature interactions is referred to as combined feature contributions, as opposed to the contribution of an individual feature. According to the Shapley interaction index, SHAP interaction values are an extension of SHAP values built from the Shapley interaction index from game theory. The joint feature contribution value $\varphi_{i,j}$ between features i and j , which follows similar axioms as SHAP values, may be calculated as follows:

$$\varphi_{i,j} = \sum_{S \in N \setminus \{i,j\}} \frac{|S|!(M - |S| - 2)!}{Z(M - 1)!} \nabla_{i,j}(S), \quad (4)$$

Where $i \neq j$, and

$$\nabla_{i,j}(S) = p(SU\{i,j\}) - p(SU\{i\}) - p(SU\{j\}) + p(S) \quad (5)$$

Although the pairwise relationship between joint features is very easy to capture in GBT models, still the feature interactions were further quantified using equations (4) and (5), and this allowed us to estimate the joint contribution of interactive features to the stress prediction model. The SHAP interaction rate between two features i and j are divided, evenly so $\varphi_{i,j} = \varphi_{j,i}$, as shown in equation (4). Further the total interaction effect is $\varphi_{i,j} + \varphi_{j,i}$. The difference between a feature's SHAP value and the SHAP interaction value as shown in equation (6) can be defined as a major effect of a prediction.

$$\varphi_{i,j} = \varphi_i - \sum_{j \neq i} \varphi_{i,j} \quad (6)$$

SHAP interaction values are derived from the same axioms as SHAP values, and they allow for the assessment of main and interaction effects separately for individual model predictions.

EXPERIMENTAL SETUP

The experiment was performed using Python 3. The model was trained on Windows 10 with a 2.30 GHz Intel Core I3-8145U processor and 8 GB RAM.

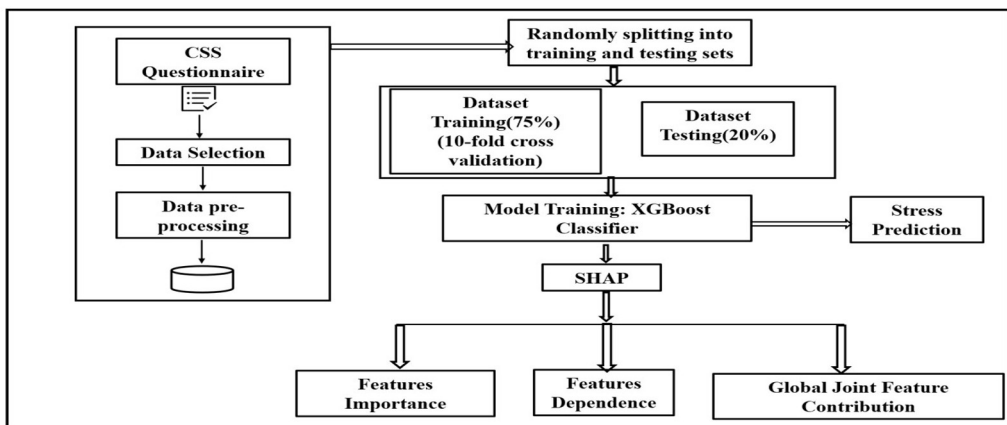
The scikit-learn machine learning component and Python 3 were used to implement XGBoost. The variables X and y were given the roles of predictors and goals at the first step of the analysis. Figure 1 represents the workflow of the proposed model. The dataset was parted into standard 75/5/20 train/validation/test ratio.

RESULTS AND FINDINGS

Model Comparison Results

K-fold cross-validation is performed to avoid any over-fitting problems. Due to the limited size of the dataset, there is not much difference in the test accuracies using 5 cross-validation and 10 cross-validation. Authors have also tested different values of k but there was not any significant variation as shown in table 3. Overall, 10-fold cross-validation is giving the best accuracy.

Figure 1. Workflow of the Proposed Model



The XGBoost model was implemented with 10-fold cross-validation with a training accuracy of 88.89. The specificity was 1.0, and the sensitivity was 0.8127. Figure 2 depicts the ROC curve. The rate of severe stress in male doctors is 52%, compared to female doctors, who have a higher rate of moderate stress (65%).

The performance of logistic regression, random forest, and XGB models was compared. Logistic regression was performed with 64.34% accuracy, second, the RF model worked superbly, and its findings were superior to those of logistic regression. With ACC and F1 increasing by 3.79, and 2.78, XGBoost outperformed RF. The performance comparison findings in table 4, show that the XGBoost model outperformed Logistic Regression and Random Forest in terms of stress problem prediction.

Feature Ablation Study

The dataset contains 39 attributes; thus, a feature selection method (correlation coefficient) was applied to select the most significant features. Finally, the customized dataset contained 6 features and

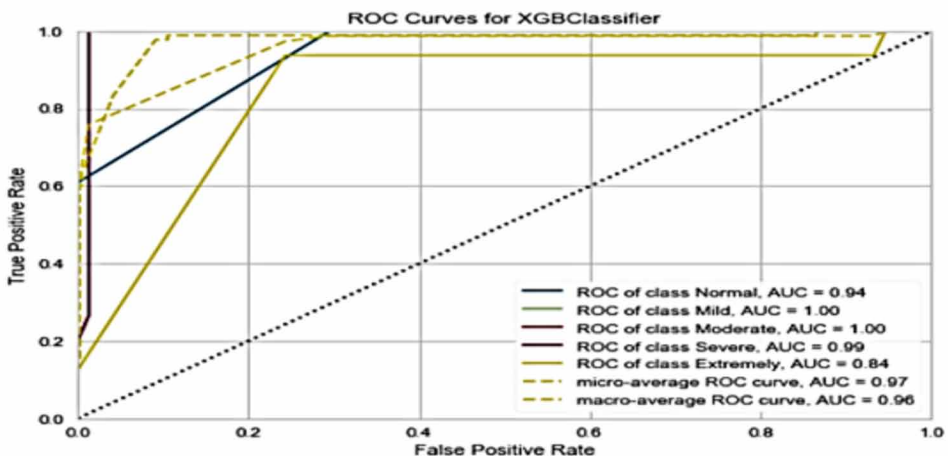
Table 3. K-Fold Cross Validation

#	K-fold CV	Test accuracy
1	K=3	86.28
2	K=5	88.11
3	K=7	88.67
4	K=10	88.89

Table 4. Model Comparison Results

Model	ACC (%)	F1(%)	AUC
LR	64.34	64.41	0.6731
RF	85.10	85.52	0.8444
XGB	88.89	88.30	0.8714

Figure 2. ROC Curves for XGB Classifier



one label. The ablation study was performed using a customized dataset. Seven trials were scheduled (providing a base trial with all the features). After running the experiment five times, features can be ranked based on their average influence on test accuracy, as shown in Table 5. Initially, every input data was included as a reference then each feature was dropped from the dataset one by one, and model accuracy was recorded. The model performs best when the `thinking_too_much` feature is removed from the training dataset, while the model's test accuracy is lowest in the base trial.

EXPLAINABILITY OF THE MODEL WITH SHAP VALUES

SHAP Force Plot (Local interpretability)

Figure 3 illustrates a sample's SHAP Force graphs from a row-wise SHAP examination. Force plot considers a single observation and displays in rank order the contribution of each feature to the prediction; this greatly increases its transparency (Nemesure *et al.*, 2021). SHAP values get assigned to each observation individually. Features that raise the forecast (to the right) are emphasized in red, and those that reduce it are painted in blue (to the left). Conventional variable significance algorithms display results for the entire population rather than for each individual. The local interpretability pinpoints and contrasts the impacts of the features. "Trouble_concentrating," is the feature that contributes the maximum to the stress prediction. This implies that trouble concentrating at the workplace along with the shortage of medicines grows the stress levels higher due to COVID-19. This prediction has an output score of 2.20 and a base score of 1.56. The score for feature "trouble concentrating" is 3, which is lower than the average score of 3.48, hence the prediction is pushed towards the left side. The anticipated stress value is less than usual; additionally, the key contributing element is the individual's "reassurance from friends," as well as "socialdist_not enough." If these factors are eliminated, the outcome result would be higher than 2.20, indicating that they are more intensely stressed than the normal individual. Every observation has its exclusive SHAP force plot. Further, a collective force plot can be generated after integrating every individual force plot, as illustrated in Figure 4.

Feature Importance SHAP Plot (Global interpretability)

The SHAP value method can demonstrate the degree of each predictor's contribution to the target variable, either positively or negatively. This plot is similar to the feature importance plot, but it can display if each variable has association with the target. Figure 5 depicts the plot, which shows the top twenty most important features for stress prediction. With a higher SHAP value, a more important feature was in informing the model's prediction. Thus, "medicines_shortage" and "trouble_concentrating" were the top two CSS elements in stress prediction. It's worth noting that these two characteristics also seemed the most important in a single SHAP force map. The X-axis

Table 5. Ablation study using XGBoost

#	Dropped Feature	Test accuracy
1	All Features (base trial)	0.888
2	trouble_concentrating	0.892
3	medicines_shortage	0.897
4	Age	0.912
3	touching_something_publicspace	0.906
5	family_not_safe	0.918
6	thinking_too_much	0.946

Figure 3. SHAP Individual Force Plot

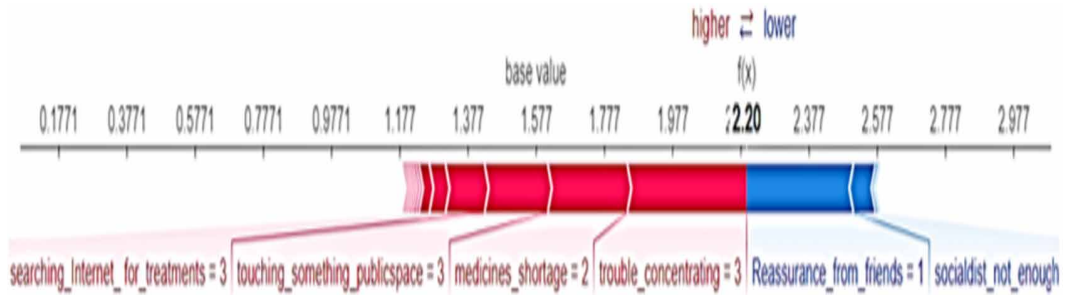
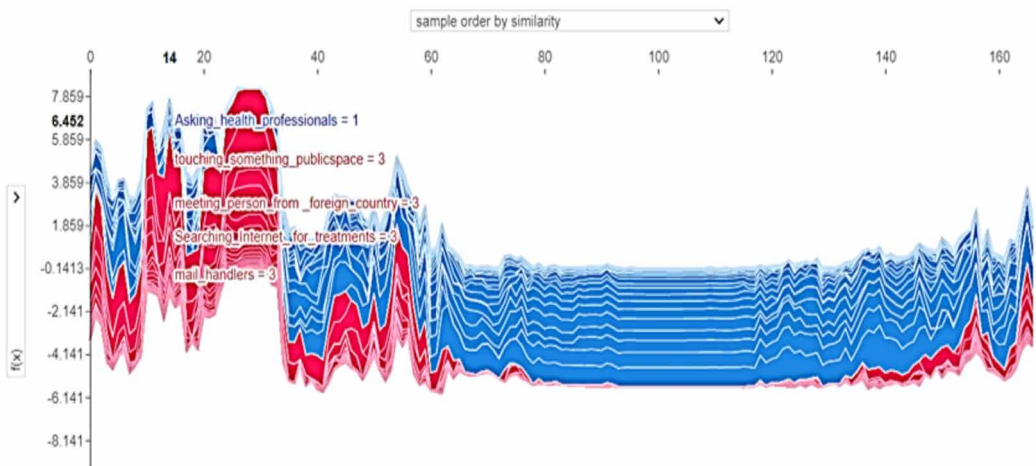


Figure 4. SHAP Collective Force Plot



in figure 5 represents the average magnitude change in the model's performance when a feature is removed from the model. It was originally deduced that feature contributions differed amongst stress classes, with some specific characteristics contributing significantly more than others. The results reflect that "medicines shortage" and trouble_concentrating" were more prominent on the whole, it is consistent with the observation that there was an unexpected lack of medications and medical supplies in India during COVID-19's second wave. Figure 6 shows similar important features evaluated using XGBoost. Figure 7 shows the most important features in the prediction of class 1 (severe). Here predictors' positive and negative associations with the target variable are shown. This summary plot, unlike a standard bar chart, displays the SHAP values for each feature. This is a collection of scatter plots, one for each feature, arranged by importance. The feature names are shown on the y-axis in decreasing order of relevance, while the SHAP values for each characteristic are listed on the x-axis from lowest to highest. Each dot represents one of the data set's samples. Like the feature importance summary plot, as shown in Figure 5, it can be deduced that "medicines shortage" was the most significant feature trailed by the "touching_something_publicspace", and "trouble_concentrating" for class 1. Similarly, it can be observed from the feature importance plot for class 2 (extremely severe) as shown in figure 8 that "trouble_concentrating" and "medicines shortage" are the two most important features in predicting stress. In another finding, it was noticed that contributions of individual features varied across stress levels, and a broad range of impacts on the model output had been seen even with the same feature value as shown in figure 7, such as

some high-contribution features like “medicines shortage”, “touching_something_publicspace”, “trouble_concentrating”, “asking_health_professionals” and “meeting_person_from_foreign_country” or “trouble_concentrating”, “medicines shortage”, “social_media_post_concerning_covid-19”, “disinfectant_supplies_shortage” and “touching_something_publicspace” in figure 8. This suggests that other variables may have an impact on the effects of these high-contribution features; as a result, the model must capture the combined contributions of these elements

Dependence Plot

It is an interaction plot displaying the effect of two features collaborating to appraise the model. Figure 9 shows dependence plots of the top four features and two lower features. Features “medicines_shortage”, “trouble_concentrating”, “touching_something_publicspace”, and “disinfectant_supplies_shortage” were found to be top contributing features as they produced higher SHAP values, this is also evident from SHAP Feature Importance Summary Plot. Such dependency charts show how one feature influences the model’s performance. Every point represents a distinct participant. In figure 9(a)

Figure 5. SHAP Feature Importance Summary Plot

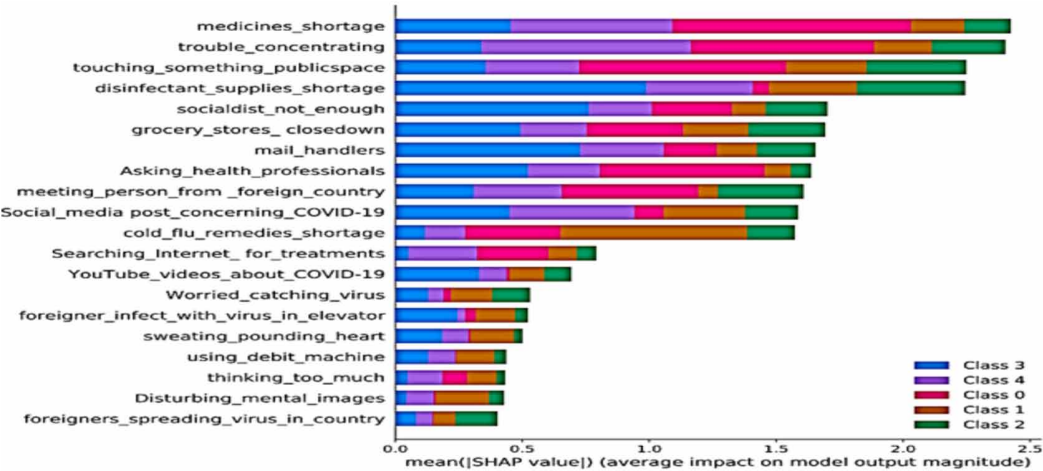


Figure 6. XGBoost Evaluating Feature Importance

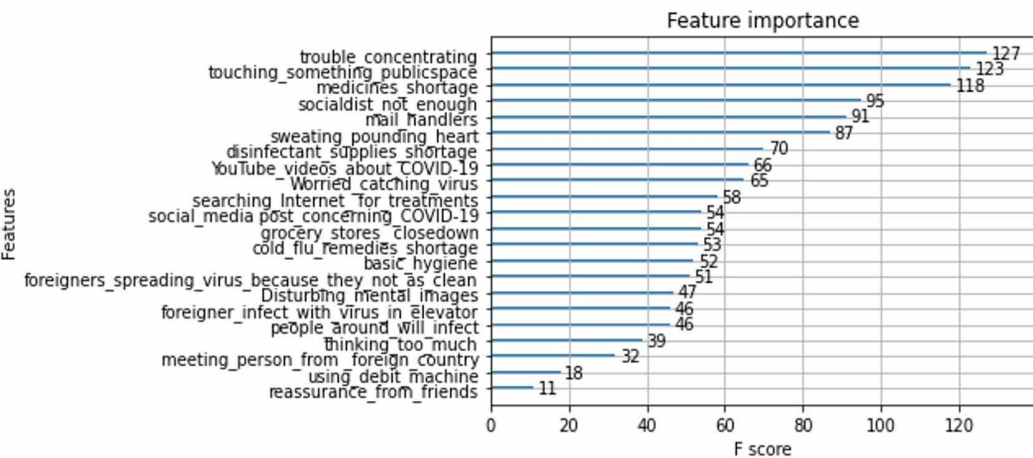


Figure 7. SHAP Feature Importance Plot (severe)

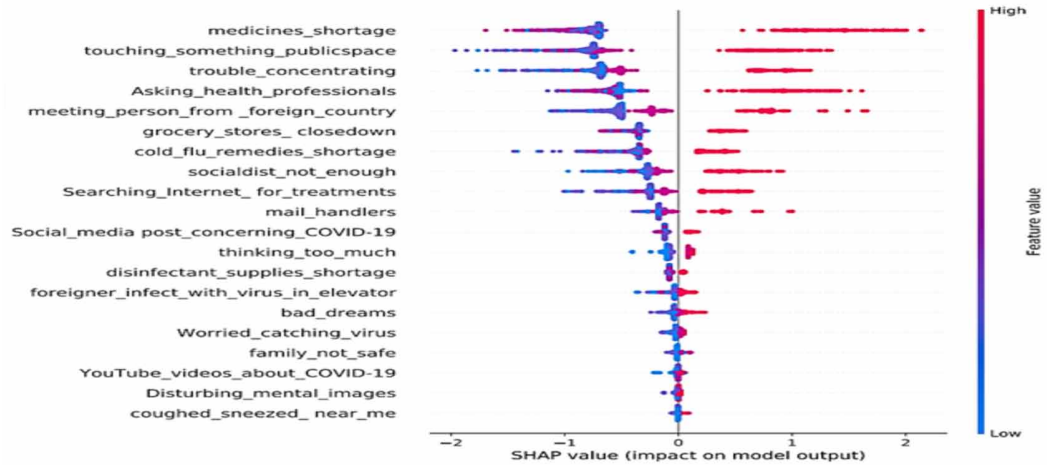
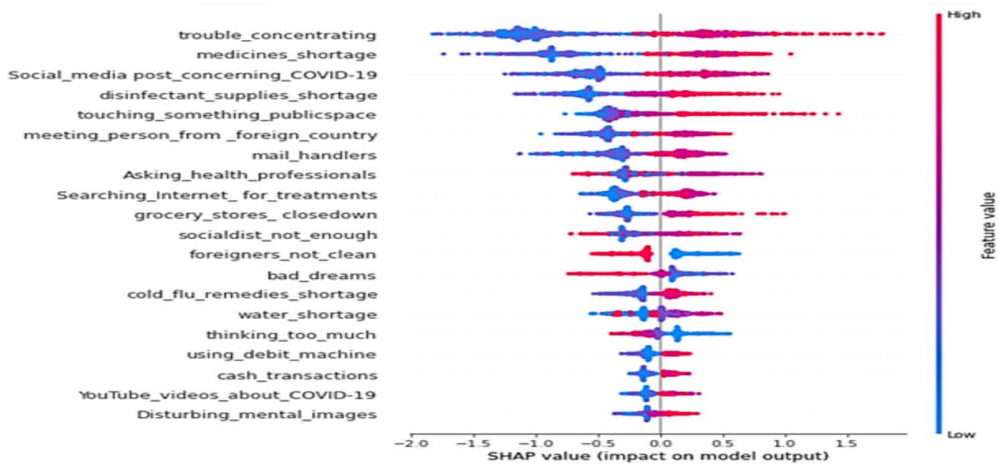


Figure 8. SHAP Feature Importance Plot (extremely severe)



dependence plot of the feature “medicines shortage” is shown, SHAP module involves automatically another feature that “medicines shortage” interacts most frequently with. It demonstrates that there is an almost linear and positive relation between “medicines shortage” and the target variable, and “medicines shortage” collaborates with “trouble_concentrating” frequently. Greater values of “medicines shortage” are more predictive of stress. It can be observed from this plot, that the gradient color of each dot corresponds to the original value of the “trouble_concentrating” from low (blue) to high (red). It was observed that “trouble_concentrating” was more predictive of stress when there is a higher shortage of medicines as shown in the figure 9 (b). There is a linear relationship between “trouble_concentrating” and stress. This visualization can be interpreted as higher values for “trouble_concentrating” leading to more stress. Figure 9(c) shows that “touching_something_publicspace” interacts more with “medicines shortage” to predict the individual’s stress. Similarly, figure 9(d) shows the effect of the “disinfectant_supplies_shortage” on predicted stress concerning “medicines shortage”. It was observed that the SHAP values for “disinfectant_supplies_shortage” were negative when there was less shortage of disinfectant supplies.

Global Joint Feature Contribution

To automatically record the features and other features' joint contributions, the SHAP interaction values method was applied. Figure 10 displays the combined feature contribution ranking results together with the amount of feature contribution to the overall contribution. Join contribution measured from the dataset occurred among “medicines shortage”, “sweating_pounding_heart” and “trouble_concentrating”; “Disturbing_mental_images” and “touching_something_publicspace”; “socialdist_not_enough” and “sweating_pounding_heart”; “disinfectant_supplies_shortage” and “trouble_concentrating”. Therefore, it is concluded that health professionals could experience high stress if they are disturbed because of a shortage of medicines during the COVID-19 second wave, suffer from excessive sweating and pounding of the heart due to extreme pressure, and experience poor concentration at work. Similarly, unwanted intrusive thoughts/images and fear of touching something in a public space could equally increase the stress levels.

Figure 9. SHAP Dependence Plots

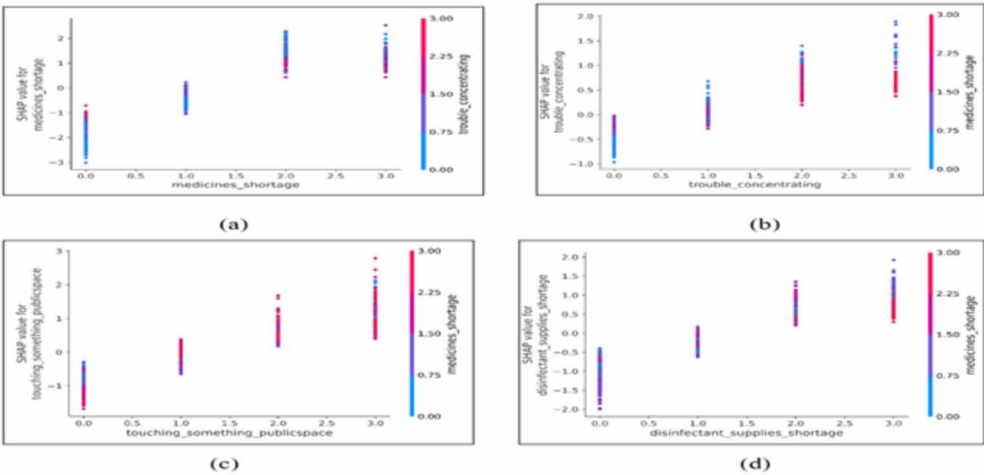
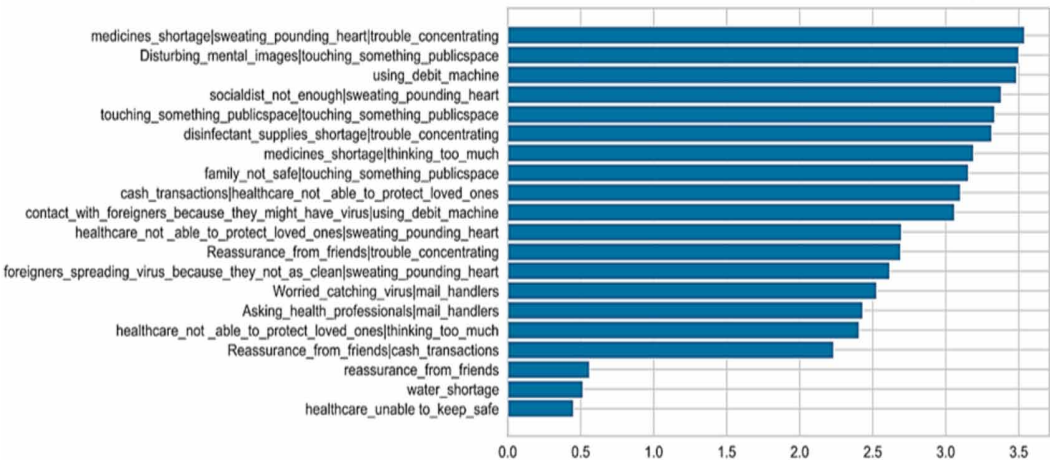


Figure 10. Joint Contribution Plot



DISCUSSION

The primary goal of the research was to measure Indian healthcare professionals' mental health during COVID-19. The authors used the XGBoost algorithm to make predictions, further SHAP values were also calculated to clarify and clinically validate the findings. More than 400 people information were used in the training of the model. As shown by the results of this research, the approximate percentage of healthcare professionals in the dataset that exceeds the severe psychiatric morbidity criteria is 52.6%. Individuals perceived that their current psychological health is worse than it was before the pandemic, and this is particularly true among females. Another goal was to provide more insight into why the model performed as it did. On the one hand, features "medicines_shortage", "trouble_concentrating", "touching_something_publicspace" and, "disinfectant_supplies_shortage" were the most important for the model in general, these features pushed the prediction higher. On the other hand, features "thinking_too_much", "disturbing_mental_images" and "foreigners_spreading_virus_in_country" were ranked at the bottom as these features pushed the prediction lower. The remaining features were somewhere in between. Relative stress is denoted by the SHAP values: the higher the SHAP value, the more it is likely to contribute to stress prediction. Understandings of global patterns of the model, as well as into single participant's variability are provided by SHAP feature dependence plots (Figure 9). The plots disclose a noticeable observation worth conversing about. It is evident why features "medicines shortage", and "trouble_concentrating" were the most significant as they generated a large range of SHAP values, which dominated the models' performance. On the other hand, it is revealed why features "Disturbing_mental_images" and "disinfectant_supplies_shortage" were the least significant for the model's outcome as any changes to their scores had only a modest (close to zero) impact on their SHAP values. Another interesting observation can be drawn from the feature dependence plot (Figure 11), each dot represents a row of data in this example. The x-axis represents the dataset's actual value, while the y-axis depicts the impact of that value on the prediction. The fact this slope upward suggests that there is a positive linear relation between trouble in concentrating and higher prediction of stress. The distribution suggests that other features must interact with "trouble_concentrating". Two points are highlighted with a similar level of "trouble_concentrating", however, that value made one prediction increase while the other prediction decreased. This recommends that the model used the interaction between two features for making predictions.

This interaction between features and their importance can be discussed in the following Figure 12 for concreteness. Points stand out spatially as being far away from the upward trend. It can be interpreted that in general, having high trouble concentrating at work increases a person's chance of having high stress. But if they feel less stressed about the shortage of medicines, then that trend reverses and the final predicted value for stress scores little.

The degree and trend of feature interactions are inspected thoroughly by examining the contributions of global joint characteristics.

CONCLUSION

This study made use of the XGBoost model to make a stress prediction for health professionals amidst the COVID-19 second wave in India. Furthermore, by assessing the feature contribution with SHAP values and SHAP interaction values features were identified that influence individual's stress, their contributions were quantified, and the interaction effects between features were also captured, which was lacking in the previous studies. The study displays the stress prediction's major feature contribution and joint feature contribution. From the experiment, some important implications can be taken like (1) shortage of medicines and trouble in concentrating were found to be the most contributing features. These two features were also ranked higher in the ablation study (2) There are indeed interactions between features and it was concluded that significant features are more likely to interact with other features. (3) It has been discovered that the more prominent a feature's main contribution is, the more

Figure 11. Similar values but different impact on stress

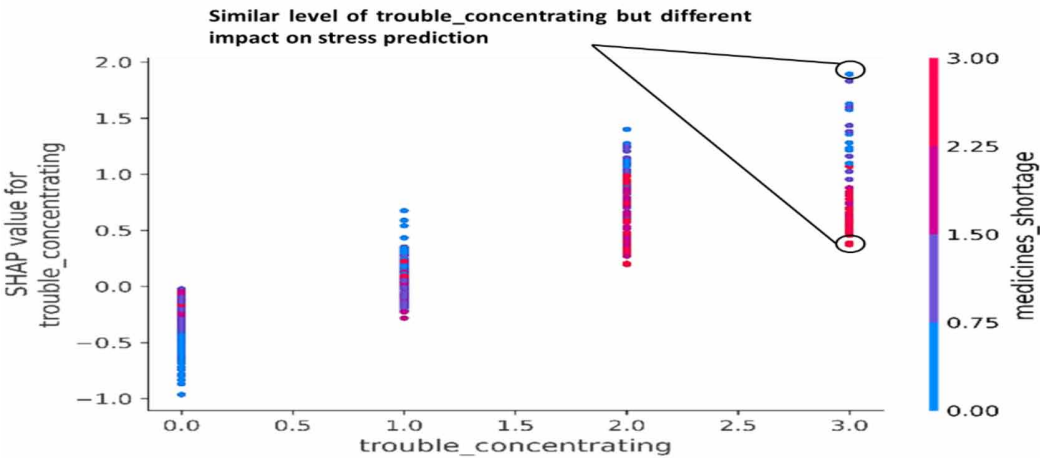
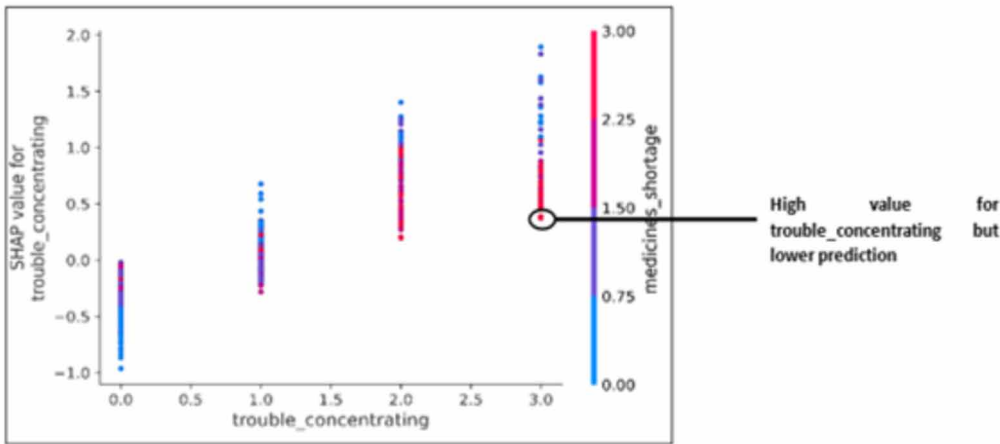


Figure 12. High value but lower prediction



likely it is to collaborate on other prominent features. Other recent research has looked into intensely emotional weariness and negative signs in Indian healthcare personnel at the peak of the COVID-19 pandemic, as well as the mental health implications of working in the COVID-19 clinical setting. (Suryavanshi *et al.*, 2021; Parthasarathy *et al.*, 2021; Selvaraj *et al.*, 2021), Situational psychological well-being is adversely correlated with COVID-19, according to this study, if healthcare professionals do not receive proper support, might have serious long-term effects.

IMPLICATION & FUTURE SCOPE

The results of this study identify that healthcare professionals in India are completely burned out, they were suffering from mental trauma during COVID-19. Since most healthcare professionals had direct contact with patients as a result, interventions must ensure to protect these professionals working in COVID-19 treatment settings that hurt one's mental health. The Major strength of this study lies with the interpretation of the model where most contributing features have been highlighted. Stress

generated due to shortage of medicines seemed to dominate other features hence the government has to ensure an uninterrupted supply of important medicines and Personal protection equipment (PPE) such as disposable gowns, masks, gloves, and face shields, as well as ventilators. Medical Association claims 734 Doctors have lost their lives from COVID-19 and they had very little time to grieve the death of their loved ones (Mohanti, 2021). There is already a shortage of healthcare workers in India. This figure is far below the WHO's requirement (Karan *et al.*, 2021). This suggests a need to provide mental health support programs for vulnerable groups to improve their mental health. Strategic planning and coordination for psychological first aid is required during major disasters. Overall, our findings suggest that regional and national organizations must focus on mental health care for healthcare professionals throughout the epidemic. There are certain limitations to this study that might be looked into for future research. This study did not include features like whether the participants had any firsthand experience of loss or disease in their family or friends as a result of COVID-19, as a result, the data cannot be used to make valid generalizations regarding COVID-19, also this study did not take into account the pre-pandemic stress level of healthcare professionals, so it not very convincing to state how much stress has been increased due to pandemic.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

FUNDING AGENCY

This research received no specific grant from any funding agency.

REFERENCES

- Almutairi, A. F., Adlan, A. A., Balkhy, H. H., Abbas, O. A., & Clark, A. M. (2018). "It feels like I'm the dirtiest person in the world.": Exploring the experiences of healthcare providers who survived MERS-CoV in Saudi Arabia. *Journal of Infection and Public Health*, 11(2), 187–191. doi:10.1016/j.jiph.2017.06.011 PMID:28676285
- Chatterjee, S. S., Chakrabarty, M., Banerjee, D., Grover, S., Chatterjee, S. S., & Dan, U. (2021). Stress, sleep and psychological impact in healthcare workers during the early phase of COVID-19 in India: A factor analysis. *Frontiers in Psychology*, 12, 473. doi:10.3389/fpsyg.2021.611314 PMID:33716874
- Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). doi:10.1145/2939672.2939785
- Chong, M. Y., Wang, W. C., Hsieh, W. C., Lee, C. Y., Chiu, N. M., Yeh, W. C., Huang, T.-L., Wen, J.-K., & Chen, C. L. (2004). Psychological impact of severe acute respiratory syndrome on health workers in a tertiary hospital. *The British Journal of Psychiatry*, 185(2), 127–133. doi:10.1192/bjp.185.2.127 PMID:15286063
- Delgado-Gallegos, J. L., Montemayor-Garza, R. D. J., Padilla-Rivas, G. R., Franco-Villareal, H., & Islas, J. F. (2020). Prevalence of stress in healthcare professionals during the covid-19 pandemic in Northeast Mexico: A remote, fast survey evaluation, using an adapted covid-19 stress scales. *International Journal of Environmental Research and Public Health*, 17(20), 7624. doi:10.3390/ijerph17207624 PMID:33086744
- Eder, S. J., Steyrl, D., Stefanczyk, M. M., Pieniak, M., Martínez Molina, J., Pešout, O., Binter, J., Smela, P., Scharnowski, F., & Nicholson, A. A. (2021). Predicting fear and perceived health during the COVID-19 pandemic using machine learning: A cross-national longitudinal study. *PLoS One*, 16(3), e0247997. doi:10.1371/journal.pone.0247997 PMID:33705439
- Flesia, L., Monaro, M., Mazza, C., Fietta, V., Colicino, E., Segatto, B., & Roma, P. (2020). Predicting perceived stress related to the Covid-19 outbreak through stable psychological traits and machine learning models. *Journal of Clinical Medicine*, 9(10), 3350. doi:10.3390/jcm9103350 PMID:33086558
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. doi:10.1214/aos/1013203451
- Janzing, D., Minorics, L., & Blöbaum, P. (2020, June). Feature relevance quantification in explainable AI: A causal problem. In *International Conference on Artificial Intelligence and Statistics* (pp. 2907-2916). PMLR.
- Jha, I. P., Awasthi, R., Kumar, A., Kumar, V., & Sethi, T. (2021). Learning the mental health impact of COVID-19 in the United States with explainable artificial intelligence: Observational study. *JMIR Mental Health*, 8(4), e25097. doi:10.2196/25097 PMID:33877051
- Karan, A., Negandhi, H., Hussain, S., Zapata, T., Mairembam, D., De Graeve, H., Buchan, J., & Zodpey, S. (2021). Size, composition and distribution of health workforce in India: Why, and where to invest? *Human Resources for Health*, 19(1), 1–14. doi:10.1186/s12960-021-00575-2 PMID:33752675
- Fujimoto, K., Kojadinovic, I., & Marichal, J. L. (2006). Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55(1), 72–99. doi:10.1016/j.geb.2005.03.002
- Khalid, I., Khalid, T. J., Qabajah, M. R., Barnard, A. G., & Qushmaq, I. A. (2016). Healthcare workers' emotions, perceived stressors and coping strategies during a MERS-CoV outbreak. *Clinical medicine & research*, 14(1), 7–14. doi:10.3121/cmr.2016.1303 PMID:26847480
- Lai, J., Ma, S., Wang, Y., Cai, Z., Hu, J., Wei, N., Wu, J., Du, H., Chen, T., Li, R., Tan, H., Kang, L., Yao, L., Huang, M., Wang, H., Wang, G., Liu, Z., & Hu, S. (2020). Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. *JAMA Network Open*, 3(3), e203976–e203976. doi:10.1001/jamanetworkopen.2020.3976 PMID:32202646
- Lee, S. M., Kang, W. S., Cho, A. R., Kim, T., & Park, J. K. (2018). Psychological impact of the 2015 MERS outbreak on hospital workers and quarantined hemodialysis patients. *Comprehensive Psychiatry*, 87, 123–127. doi:10.1016/j.comppsy.2018.10.003 PMID:30343247
- Lundberg, S. M., & Lee, S. I. (2017, December). A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems* (pp. 4768-4777). Academic Press.

- Maunder, R. G., Lancee, W. J., Balderson, K. E., Bennett, J. P., Borgundvaag, B., Evans, S., Fernandes, C., Goldbloom, D., Gupta, M., Hunter, J., McGillis Hall, L., Nagle, L., Pain, C., Peczeniuk, S., Raymond, G., Read, N., Rourke, S., Steinberg, R., Stewart, T., & Wasylenki, D. A. et al. (2006). Long-term psychological and occupational effects of providing hospital healthcare during the SARS outbreak. *Emerging Infectious Diseases*, 12(12), 1924–1932. doi:10.3201/eid1212.060584 PMID:17326946
- Meng, Y., Yang, N., Qian, Z., & Zhang, G. (2021). What makes an online review more helpful: An interpretation framework using XGBoost and SHAP values. *Journal of Theoretical and Applied Electronic Commerce Research*, 16(3), 466–490. doi:10.3390/jtaer16030029
- Mohanti, M. (2021, Apr 15). *IMA Claims 734 Doctors Dead from Covid-19, Families Need Help*. Indiatime. Available at <https://www.indiatimes.com/technology/science-and-future/indian-medical-association-doctor-deaths-india-covid-19-538272.html>
- Nemesure, M. D., Heinz, M. V., Huang, R., & Jacobson, N. C. (2021). Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. *Scientific Reports*, 11(1), 1–9. doi:10.1038/s41598-021-81368-4 PMID:33479383
- Nooripour, R., Hosseinian, S., Hussain, A. J., Annabestani, M., Maadal, A., Radwin, L. E., Hassani-Abharian, P., Pirkashani, N. G., & Khoshkonesh, A. (2021). How resiliency and hope can predict stress of Covid-19 by mediating role of spiritual well-being based on machine learning. *Journal of Religion and Health*, 60(4), 2306–2321. doi:10.1007/s10943-020-01151-z PMID:33398655
- Parthasarathy, R., Jaisoorya, T. S., Thennarasu, K., & Murthy, P. (2021). Mental health issues among health care workers during the COVID-19 pandemic—a study from India. *Asian Journal of Psychiatry*, 58, 102626. doi:10.1016/j.ajp.2021.102626 PMID:33721830
- Praveen, S. V., Ittamalla, R., & Deepak, G. (2021). Analyzing the attitude of Indian citizens towards COVID-19 vaccine—A text analytics study. *Diabetes & Metabolic Syndrome*, 15(2), 595–599. doi:10.1016/j.dsx.2021.02.031 PMID:33714134
- RanjanR.SharmaA.VermaM. K. (2021). Characterization of the Second Wave of COVID-19 in India. medRxiv. 10.1101/2021.04.17.21255665
- Rodríguez-Hidalgo, A. J., Pantaleón, Y., Dios, I., & Falla, D. (2020). Fear of COVID-19, stress, and anxiety in university undergraduate students: A predictive model for depression. *Frontiers in Psychology*, 11, 3041. doi:10.3389/fpsyg.2020.591797 PMID:33224080
- Roth, A. E. (Ed.). (1988). *The Shapley value: essays in honor of Lloyd S. Shapley*. Cambridge University Press. doi:10.1017/CBO9780511528446
- Roy, D., Tripathy, S., Kar, S. K., Sharma, N., Verma, S. K., & Kaushal, V. (2020). Study of knowledge, attitude, anxiety & perceived mental healthcare needs in the Indian population during COVID-19 pandemic. *Asian Journal of Psychiatry*, 51, 102083. doi:10.1016/j.ajp.2020.102083 PMID:32283510
- Salari, N., Hosseinian-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), 1–11. doi:10.1186/s12992-020-00589-w PMID:32631403
- Selvaraj, P., Muthukanagaraj, P., Saluja, B., Jeyaraman, M., Anudeep, T. C., Gulati, A., Es, S., M, D., Jain, R., Kadhir, I., & Rao, S. P. (2020). Psychological impact of COVID-19 pandemic on health-care professionals in India—A multicentric cross-sectional study. *Indian Journal of Medical Sciences*, 72(3), 141–147. doi:10.25259/IJMS_193_2020
- Sim, K., Chong, P. N., Chan, Y. H., & Soon, W. S. W. (2004). Severe acute respiratory syndrome-related psychiatric and posttraumatic morbidities and coping responses in medical staff within a primary health care setting in Singapore. *The Journal of Clinical Psychiatry*, 65(8).
- Sundararajan, M., & Najmi, A. (2020, November). The many Shapley values for model explanation. In *International Conference on Machine Learning* (pp. 9269–9278). PMLR.
- Suryavanshi, N., Kadam, A., Dhumal, G., Nimkar, S., Mave, V., Gupta, A., Cox, S. R., & Gupte, N. (2020). Mental health and quality of life among healthcare professionals during the COVID-19 pandemic in India. *Brain and Behavior*, 10(11), e01837. doi:10.1002/brb3.1837 PMID:32918403

Tam, C. W., Pang, E. P., Lam, L. C., & Chiu, H. F. (2004). Severe acute respiratory syndrome (SARS) in Hong Kong in 2003: Stress and psychological impact among frontline healthcare workers. *Psychological Medicine*, 34(7), 1197–1204. doi:10.1017/S0033291704002247 PMID:15697046

Taylor, S. (2019). *The psychology of pandemics: Preparing for the next global outbreak of infectious disease*. Cambridge Scholars Publishing.

Taylor, S., Landry, C. A., Paluszec, M. M., Fergus, T. A., McKay, D., & Asmundson, G. J. (2020). Development and initial validation of the COVID Stress Scales. *Journal of Anxiety Disorders*, 72, 102232. doi:10.1016/j.janxdis.2020.102232 PMID:32408047

Wang, L., Wang, X., Chen, A., Jin, X., & Che, H. (2020, September). Prediction of type 2 diabetes risk and its effect evaluation based on the XGBoost model. [J. Multidisciplinary Digital Publishing Institute.]. *Health Care*, 8(3), 247.

Rajkumar, R. P. (2020). COVID-19 and mental health: A review of the existing literature. *Asian Journal of Psychiatry*, 52, 102066.

Pfefferbaum, B., & North, C. S. (2020). Mental health and the Covid-19 pandemic. *The New England Journal of Medicine*, 383(6), 510–512. doi:10.1056/NEJMp2008017 PMID:32283003

Cullen, W., Gulati, G., & Kelly, B. D. (2020). Mental health in the COVID-19 pandemic. *QJM: An International Journal of Medicine*, 113(5), 311–312. doi:10.1093/qjmed/hcaa110 PMID:32227218

Usher, K., Durkin, J., & Bhullar, N. (2020). The COVID-19 pandemic and mental health impacts. *International Journal of Mental Health Nursing*, 29(3), 315–318. doi:10.1111/inm.12726 PMID:32277578

Zhou, T., Xu, C., Wang, C., Sha, S., Wang, Z., Zhou, Y., Zhang, X., Hu, D., Liu, Y., Tian, T., Liang, S., Zhou, L., & Wang, Q. (2022). Burnout and well-being of healthcare workers in the post-pandemic period of COVID-19: A perspective from the job demands-resources model. *BMC Health Services Research*, 22(1), 1–15. doi:10.1186/s12913-022-07608-z PMID:35236354

Pooja Gupta is working as Assistant Professor at the School of Computing, DIT University, Dehradun, India. Ms. Gupta did her Master's in Computer Science and Engineering from Uttarakhand Technical University. She has Twelve years of experience including in academics and research. Ms. Gupta has publications in various journals of repute. Her research interest includes Machine learning and Data Sciences. She is an avid learner, hence keeps upgrading her skills. She has completed various short-term courses from reputed institutes like three months SFRF from IIT Delhi.

Srabanti Maji is working as Assistant Professor in the Department of CSE in DIT University, Dehradun. Previously she worked at SGHCMT Patiala under Punjabi University in Punjab. She has done Ph.D(CSE) in the year 2013 from Thapar University, Patiala. She has more than 5 year's research experience and 13 years teaching experience. She has 4 SCI Journal publications in the field of Machine Learning, many International Journal publications and 5 conferences, 2 book chapter publications. She is having ACM membership.

Ritika Mehra is working as a Professor and Dean of School of Computer Science & Engineering at Dev Bhoomi Group of Institutions, Dehradun, Uttarakhand. She received her Ph.D. degree in Computer Science from Gurukul Kangri University, Haridwar in the year 2010, M.Tech degree in Computer Science and Engineering from Uttarakhand Technical University, Dehradun and MCA from Gurukul Kangri University Dehradun. She specializes in core areas of computer science and holds experience of more than 19 years. She is an innovative person with deep knowledge of Computer Networking, Mobile Computing, Data Mining, Bigdata Analytics etc. She has also published many research papers in various National and International Journals.