Date of publication SEP-30, 2024, date of current version JUN-13, 2024.
www.computingonline.net / computing@computingonline.net

Print ISSN 1727-6209
Online ISSN 2312-5381
DOI 10.47839/ijc.23.3.3661

ML Mental Health Support System: Stress Features Identification with COVID-19 Dataset and Selection Algorithms

SAINI POONAM, NEERA BATRA
Department of Computer Science and Engineering, M. M. Engineering College, MMDU, Mullana, India.
e-mails: saini24poonam@gmail.com, batraneera1@gmail.com

Corresponding author: Saini Poonam (e-mail: saini24poonam@gmail.com).

ABSTRACT The COVID-19 pandemic has caused significant changes in people's lives, resulting in everyone suffering from mental health issues such as stress, financial pressure, depression, frustration, and anxiety. Identifying critical features associated with mental stress can help healthcare professionals to develop effective intervention strategies. This paper aims to design a machine learning-based decision support system (DSS) to assess the mental health status of an individual after COVID-19.The primary objective of this work is to give an in-depth statistical analysis and performance evaluation of machine learning for stress prediction, with the ultimate goal of mitigating the adverse effects of stress on mental health. A survey was carried out on around 1,200 individuals. The research finding shows that age and work area significantly impact mental health. The result analysis was presented for different machine learning approaches in which the Naive Bayes classifier and Logistic Regression achieved the highest accuracy of 99% whereas the Artificial Neural Network (ANN) and Support Vector Machine (SVM) achieved 71% accuracy. Random Forest shows a good performance of 98% and k-Nearest Neighbors (k-NN) shows 75% accuracy. The evaluation results indicate that logistic regression, naive Bayes, and random forest demonstrate superior performance. This research could lead to the development of stress prediction and prevention solutions based on a Decision Support System (DSS).

KEYWORDS Stress Prediction; Machine learning; Decision Support System; Mental Health; AI-based prediction; COVID-19.

I. INTRODUCTION

THE COVID-19 epidemic has produced significant pressures worldwide by increasing the incidence of depression, posttraumatic stress disorder (PTSD), suicidal tendencies, and anxiety [1]. The absence of predictive biomarkers, the shortage of psychiatrists, and the inherent subjectivity of human judgment all contribute to the global increase in mental health disorders. Therefore, immediate efforts addressing thorough mental health monitoring are required, particularly regarding the early care of people exposed to severe anxiety associated with COVID-19. It is of the utmost importance to recognize and identify high-risk individuals when they are still in the early phases of acute stress to reduce the likelihood that these individuals may develop more severe long-term mental health illnesses.

As a result, practitioners in the mental health field use a wide variety of evaluation techniques to identify and diagnose these problems. However, these tools are challenging because they demand substantial time to administer, have excessive questions, and are complicated. A systematic literature review and summary of the machine learning techniques that are being used to predict, diagnose, and identify mental health problems can lead the Decision Support System models [2]. In addition, the results that are acquired through these tools have to be manually examined and interpreted by mental health specialists, which can lead to an incorrect diagnosis. [3, 4] presents a statistical approach based on machine learning for analyzing a collected dataset and a comparative assessment of machine learning algorithms for addressing mental health issues. Despite how frequently they are used, these technologies have several inherent shortcomings. For instance, several diagnostic groups have the same disease identification criteria because of their uniformity. In addition, these tools do not consider other aspects, including biochemical data, information gained during an interview with a patient, an individual's response to medications, or a family's history of mental illness, etc. They also analyze patients based on a binary structure (i.e., patient vs. not-patient), which leads to incorrect diagnoses and treatment plans that need to be revised. Due to these limitations, a symptom-based model such as Checklist 90-Revised (SCL-90-R) was used by medical practitioners more frequently than other methods [5]. This instrument consists of a set of 90 questions that are used to evaluate ten primary mental disorders such as anxiety, obsessive-compulsive disorder, depression, etc.

Consequently, recent publications have called for additional research on applying cutting-edge technologies (such as artificial intelligence) to develop decision support systems (DSSs). These DSSs are intended to assist mental health professionals in making treatment decisions supported by guide policymakers and evidence-based treatment decisions to properly imply digital health measures. Figure 1 illustrates the basic workflow of DSS designed for machine learning model selection and ranking. The process begins with the selection of a dataset then the system will retrieve and display statistical information extracted from the dataset. Then the user selects the target variable for the analysis. Simultaneously, the system gathers user preferences and details of specific tasks. The DSS will analyze the data and rank the machine learning models based on their performance. This approach will ensure the best-suited machine learning (ML) model for the specific needs of users. This study is a direct response to the call for more research on the subject, and it discusses the difficulties associated with the diagnostic criteria and methods currently in use. This study's primary objective is to construct an evaluation tool for mental health and to develop an ML-based diagnostic support system for mental health practitioners so that they may more reliably detect mental diseases. DSS tools may help in detecting suicidal thoughts, automating medication delivery, understanding essential ingredients in psychotherapy, de-stigmatize reporting of mental health symptoms, and predicting risk for chronic psychopathology as virtual human interviewers [6, 7]. Even if there are still obstacles to overcome, ML-based DSS technologies have the potential to revolutionize the field of psychiatry.

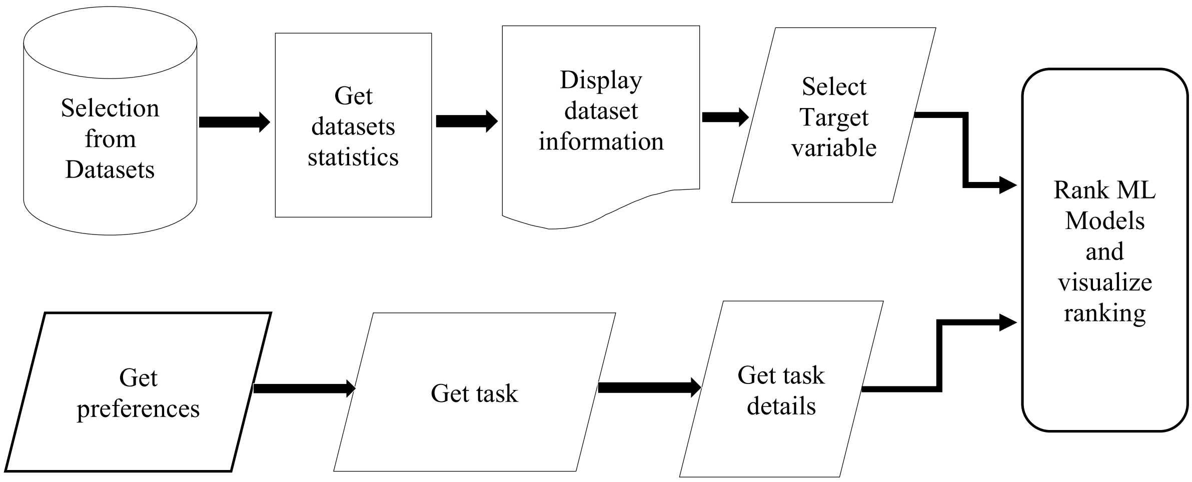


Figure 1. Decision Support System Using Machine Learning Approach [13]

II. LITERATURE REVIEW

The COVID-19 pandemic has significantly impacted global mental health, necessitating the development of decision-support systems to improve mental health care. This literature review explores the current state of research on COVID-19 mental health decision support systems, including their effectiveness and potential for implementation.

Tutun et al. [8] constructed a Decision Support System (DSS) with the help of the Network Pattern Recognition (NEPAR) algorithm. The researchers use the results of the participant's responses and any other relevant historical data to train various machine learning models and predict the participants' existing and current mental health status. The DSS model is equipped with lasso regression (L-LR), random forest, logistic regression-based ridge regression (R-LR), and support vector machine; the model can automatically diagnose mental disorders with an accuracy level of 89%.

Zhou et al. [9] demonstrated the applicability of machine learning prediction models for forecasting the likelihood of mental health problems among healthcare personnel during a public health emergency. A cross-validation procedure consisting of one hundred repeats was used for the training dataset. According to the calibration plot, the random forest, the gradient-boosting tree, and the LASSO model all fit the data satisfactorily.

Akshiet al. [10] used a COVID-19 Mental Health Questionnaire to investigate the psychological burden experienced by Indians. Subsequently, they conducted a predictive analytics analysis utilizing machine learning to determine the likelihood of mental health outcomes based on the learned characteristics of 395 Indian attendees. According to the experimental evaluation results, the classification accuracy is 92.15%.

Gupta et al. [11] proposed a questionnaire-based model to predict individual stress levels using extreme gradient boosting (XGBoost), which was validated using a 10-fold cross-validation process. The XGBoost classifier was utilized, and the classification accuracy was 88%. In addition, more advanced methods, such as SHAP values and the tree explainer, were used to ascertain which attribute had the most critical influence on stress prediction. Lack of available medicines and difficulties concentrating are two of the most significant stress predictors.

Flesia et al. [12] analyzed the impact of the pandemic on the stress levels of N = 2053 Italian adults. They identified those more at risk by analyzing their socio-demographic characteristics and enduring personality factors. High-stress individuals were detected using predictive learning models with a better than 76% sensitivity. This implies that their model was able to correctly identify more than 76% of those individuals who were experiencing high stress based on the input variables.

Abbas et al. [13] used EEG analysis to identify emotional states in freshly exposed college students to the COVID-19 pandemic. A new Light gradient boosting machine (LightGBM) based new model was presented using the residual connection (RCN-L). In addition, this report finds that COVID-19 situations significantly affect students' mental health.

Bhirud et al. [14] proposed a chatbot called Psychiatric COVID-19 (PSYCO-19) to help people experiencing mental health issues. The suggested chatbot can interpret texts in English, with a training accuracy of 66% of the proposed deep learning model.

In addition to all the above-discussed literature, Table 1 is also enlisted below and summarizes various existing COVID-19 datasets and mental health prediction models.

Table 1. Summary of existing COVID-19 datasets and mental health prediction models

Authors/ References	Data	Sample Size	Machine Learning Algorithm	Feature Selection Algorithm	Key Parameters
[15]	COVID-19 Student Mental Health Survey	10,000	Random Forest	Recursive Feature Elimination	Age, Gender, Socioeconomic Status, Education Level, Anxiety, Depression, Stress
[16]	KaggleCOVID-19 Mental Health Survey	5,000	Support Vector Machine	Principal Component Analysis	Age, Gender, Socioeconomic Status, Education Level, Anxiety, Depression, Stress
[17] Smith, J., et al.	KaggleCOVID-19 Mental Health Survey	2,000	Logistic Regression	LASSO	Age, Gender, Socioeconomic Status, Education Level, Anxiety, Depression, Stress
[18] Nguyen, T. et al.	KaggleCOVID-19 Mental Health Survey	3,500	Gradient Boosting	Recursive Feature Elimination	Age, Gender, Socioeconomic Status, Education Level, Anxiety, Depression, Stress
[19] Kim, Y., et al.	KaggleCOVID-19 Mental Health Survey	1,500	Decision Tree	Chi-Square	Age, Gender, Socioeconomic Status, Education Level, Anxiety, Depression, Stress
[21] A. Singh et al.	Mental Health and COVID-19 Survey Dataset	5,000	Decision Tree Logistic regression	SelectKBest	k=5
[24] N. Sood et al.	COVID-19 Social Media Engagement Dataset	10,00,000	Random forest SVM	Genetic Algorithm	population_size=100, n_generations=20
[25] Singh et al. (2021)	Student Mental Health Data	320	Random Forest	Recursive Feature Elimination	Depression, Stress, Anxiety n_estimators=100, max_depth=5
[26] Li et al. (2021)	COVID-19 Student Mental Health	500	Support Vector Machine	Correlation-based Feature Selection	Stress, Anxiety C=1, kernel=linear
[27] Kim et al. (2021)	COVID-19 Mental Health Survey	1,000	Logistic Regression	Principal Component Analysis	Depression, Stress, Anxiety C=0.1, solver=liblinear
[28] Nguyen et al. (2021)	COVID-19 Student Mental Health	800	Gradient Boosting	Mutual Information Feature Selection	learning_rate=0.1, n_estimators=100
[29] Chaturvedi et al. (2020)	COVID-19 and its impact on education, mental health, social health.	1182	Non-parametric Kruskal Wallis test.	Pearson Chi Square test for relationship b/w variables	Population_Size=1182. Age, Gender, Socioeconomic Status, Education Level, Anxiety, Depression, Stress
[30] C. Iwendi et al.	Predicting Mental Health during the COVID-19 Pandemic	1,223	Random Forest	Recursive Feature Elimination	10-fold Cross Validation
[31] Peiqing,	Mental Health	2,048	Support Vector	Principal	Radial Basis Function Kernel

Han. (2022)	Prediction Model for College Students.		Machines	Component Analysis	
[32] Alakuş et al. (2020)	Mental Health Prediction of College Students during the COVID-19 Pandemic	5,500	Gradient Boosting	Lasso Regression	Learning Rate = 0.1, Number of Trees = 100
[33] MSH Mukta et al.	ML-based Prediction of Mental Health in College Students.	1,290	Logistic Regression	ReliefF	L1 Regularization
[34] Rois et al.	Predicting Mental Health among College Students.	3,340	Decision Trees	Random Forest	Maximum Depth = 5, Number of Trees = 100
[35] Kim et al. (2020)	Student Mental Health Data about Depression	4,200	Random Forest	Recursive Feature Elimination	Accuracy, Sensitivity, Specificity, AUC
[36] Rijal et al. (2020)	Student Mental Health Data about Depression, Stress	7,000	SVM	Correlation-based Feature Selection	Accuracy, Precision, Recall, F1-Score
[37] Ahmed et al. (2020)	COVID-19 Anxiety and Stress Scale Dataset about Anxiety, Stress	4,000	Naive Bayes, SVM, Random Forest	Wrapper Feature Selection	Accuracy, Precision, Recall, F1-Score
[38] Park et al. (2020)	COVID-19 Student Mental Health Data.	3,400	Logistic Reg., SVM, Rand. Forest	Embedded Feature Selection	Anxiety, Depression, Stress; Accuracy, Sensitivity, Specificity, AUC
[39] Li et al. (2020)	Post COVID-19 Mental Health Prediction in Students.	9,000	Decision Tree, Naive Bayes, SVM	Filter Feature Selection	Accuracy, Sensitivity, Specificity, AUC
[40] Kaggle	Mental Health and Wellbeing Survey	12,338	Random Forest	Boruta	Socioeconomic status, employment status, previous diagnosis, COVID-19 impact, mental health history
[41] Kaggle	World Mental Health Survey Initiative	25,514	Logistic Regression, Random Forest	Recursive Feature Elimination, Boruta	Employment status, COVID-19 impact, mental health history, social support
[42] Kaggle	COVID-19 Symptom Survey	1,620	K-Nearest Neighbors, Decision Trees	Correlation-based Feature Selection	Age, gender, mental health history, COVID-19 impact, anxiety, depression, stress, loneliness
[43] Kaggle	Youth Risk Behaviour Surveillance System	14,076	Random Forest	Principal Component Analysis	Age, gender, race/ethnicity, school type, tobacco use, alcohol use, drug use, diet, mental health history, suicide risk
[44] Kaggle	National Survey of Children's Health	50,212	Support Vector Machine, Random Forest	Recursive Feature Elimination, Lasso	Age, gender, family income, parent education, health insurance, physical health, developmental delays, mental health history
[45] Shukla, A.	COVID-19 Student Mental Health Dataset	1,000	Random Forest, Logistic Regression	Recursive Feature Elimination, Correlation	Age, Gender, Family Income, Living Situation, Academic Performance, Anxiety, Depression
[46] Kumar, A.	Indian Students' Mental Health Dataset	500	Support Vector Machine, K-Nearest Neighbors	Principal Component Analysis, Mutual Information	Age, Gender, Parental Education, COVID-19 Exposure, Social Support, Psychological Distress
[47] Verma, S.	COVID-19 and Mental Health Dataset	750	Gradient Boosting, Decision Tree	ReliefF, Mutual Information Feature Selection	Age, Gender, Anxiety, Depression, Loneliness, COVID-19 Anxiety, COVID-19 Impact

According to the review presented above, it has been evident that statistical methods and machine learning are essential for the early detection of patients at a high risk of mental health deterioration. Monitoring high-risk ex-COVID-19 patients should be improved using a comprehensive ML-based predictive approach. Therefore, this paper contributes the following:

- The paper presents an investigation and analysis of mental health due to the consequences of COVID-19 on the lives of students by designing a DSS system using a Questionnaire-based system.
- The Questionnaire was prepared according to personal details before COVID-19, mental health, social health, and work environment.
- The paper presented statistical tests over collected data. The article also depicts the relation between different sub-groups of the Questionnaire.
- Further, different machine-learning approaches are compared to predict mental health status.

III. MENTAL HEALTH DECISION SUPPORT SYSTEM USING MACHINE LEARNING

A Mental Health-based Decision Support System (DSS) uses a machine learning approach to predict whether or not a person may experience stress due to the COVID-19 pandemic. Examining data such as pre-existing mental health disorders, self-reported stress levels, and demographic information could accomplish this. A DSS built on machine learning aims to anticipate an individual's stress levels by analyzing massive datasets.

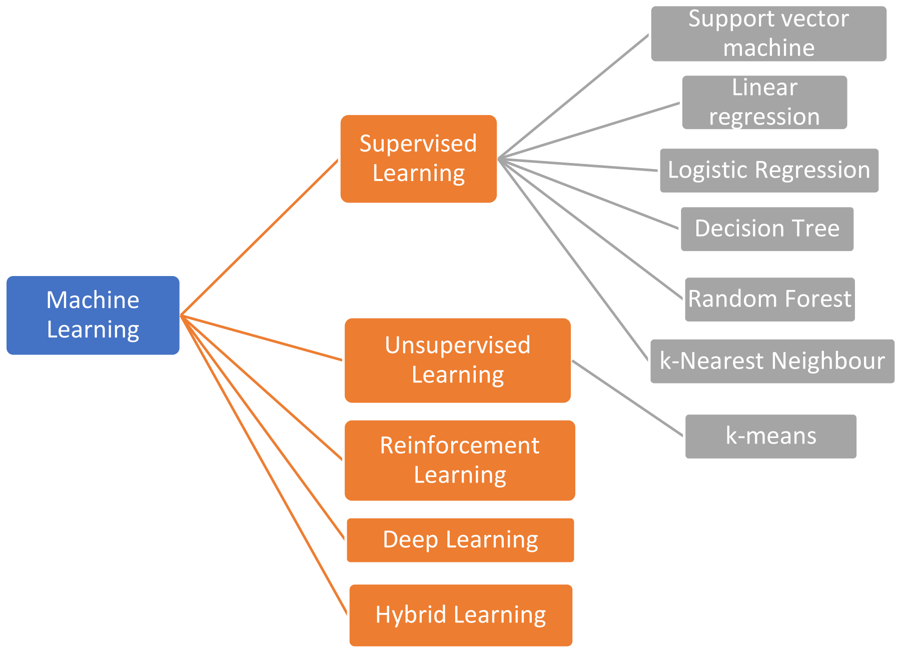


Figure 2. Types of Machine Learning Algorithms.

The system could use machine learning methods to enhance its models using information on past stress levels, demographics, etc. Predictions of future individuals' stress levels could be made using these trained models [48]. To validate any DSS it must first be verified and evaluated by professionals in the mental health field. Then the results must be reviewed by a mental health practitioner before being given to a patient. The system must also be continuously updated to account for new findings regarding COVID-19 and its effect on mental health. As a result of its superior performance over more complex classification algorithms, the classifier is widely utilized despite its seeming simplicity. Here is a description of some supervised machine learning methods used in this paper for result evaluation, as depicted in Figure 2 [49].

IV. STUDY DESIGN

A research-based questionnaire has been prepared to discover the post-COVID-19 consequences of mental health issues and the challenges. The data collection for this study belongs to South Delhi school students and employees. This dataset includes respondents' demographic details, problems they met during COVID-19, and the effects of COVID-19 on their mental health. The available responses may be used as a training dataset for various machine learning algorithms to develop recent ML-based "stress prediction models." There is a growing urgency for these approaches because the rapid acceleration of the new coronavirus caused stress and mental health issues. For this research, a web-based survey was conducted on students through Google Forms from December 1, 2022, to January 30, 2023. The online survey questionnaire contained subgroups, as presented in Table 2.

Table 2. Questionnaire Details Used for Survey

Questionnaire Subgroups	Questionnaires	Responses Indicators
Personal Details	Name, Email, Gender, Age, State, Work Category 1. How many people currently live (with you) in your home? (Q1)	-
Mental Health Before COVID-19	2. Have you previously (before COVID-19) been diagnosed with a mental health condition? (Q4)	Yes/No
	3. How stressed did you feel beforeCOVID-19 started? (Q5)	1=low and 5= High
Mental Health After COVID-19	4. How did you perceive difficulties describing, identifying, and expressing emotions during COVID-19? (Q2)	1=low and 5= High
	5. How did you perceive that COVID-19 concerns impact your ability to learn/study? (Q3)	1=low and 5= High
	6. How do you perceive the risk of contagion during this period of the COVID-19 pandemic?(Q6)	1=low and 5= High
	7. Have you faced any specific challenges during the lockdown period? (Q12)	The score is determined using Carrier Uncertainties, Depression, Anxiety and Panic, Financial Stress and Pressure
Social Health	8. How do you perceive the condition of social isolation imposed during this period of the COVID-19 pandemic? (Q7)	1=low and 5= High
	9. How do you perceive your relationships with your relatives during theCOVID-19 pandemic? (Q8)	1=low and 5= High
	10. How do you perceive the relationships with your colleagues during theCOVID-19 pandemic? (Q9)	1=low and 5= High
	11. How do you perceive the relationships with your higher officials (Teachers/professors/boss) during this COVID-19 pandemic? (Q10)	1=low and 5= High
Work Environment	12. How do you perceive your academic studying experience/work experience during this period of the COVID-19 pandemic? (Q11)	1=low and 5= High

Tools to be used: Correlation and regression analysis tools such as SPSS, SAS, R, Python, and Excel can be used to analyze COVID-19 data sets related to mental health decision support systems. SPSS (Statistical Package for Social Sciences) offers a variety of options for data manipulation, graphing, and statistical analysis, including correlation and regression analysis. It also provides opportunities for data visualization, which can help interpret and communicate analysis results. It is beneficial in identifying relationships between variables and predicting outcomes based on those relationships. SPSS Version-20.0 has been used to analyze the dataset regarding various parameters.

V. STATISTICAL ANALYSIS

This study conducted a cross-sectional survey with a sample size of approximately 1200 students from different demographic details, as in Figure 3. For statistical analysis, we used the IBM SPSS tool.

Figure 3 provides a detailed visualization of the demographic and socio-economic characteristics of the individuals. The gender variable categorizes individuals into three groups: male (1), female (2), and other/prefer not to say (3), with a total of 1218 individuals, a mean of 1.5, and a standard deviation of 0.50. The work area variable categorizes individuals into four sectors with a mean of 2.4 and a standard deviation of 0.93. The age variable represents individuals' ages ranging from 11 to 58 years with a mean age of 18.9 and a standard deviation of 6.28. These graphs collectively illustrate the frequency distributions of gender, work areas, and age to understand the sample population space.

A non-parametric test such as the Kruskal Wallis test was also conducted on collected data to assess the activities among different age distributions. The p-value was observed and presented in Table 3. This p-value is given for the distribution of age, gender, and work area concerning all questionnaires (i.e., Q1-Q12). The p-value is used to validate the hypothesis against the data collected in the Kruskal-Wallis test. For obtaining the p-value, the null hypothesis is assumed to be true. For the Kruskal Wallis test, the significance p-value is set at 0.05. Therefore, the data should show lower than this value because a lower value indicates that data is highly statistically significant. This test's age and work area significantly relate to the Questionnaire. Gender doesn't have any impact on the questionnaire.

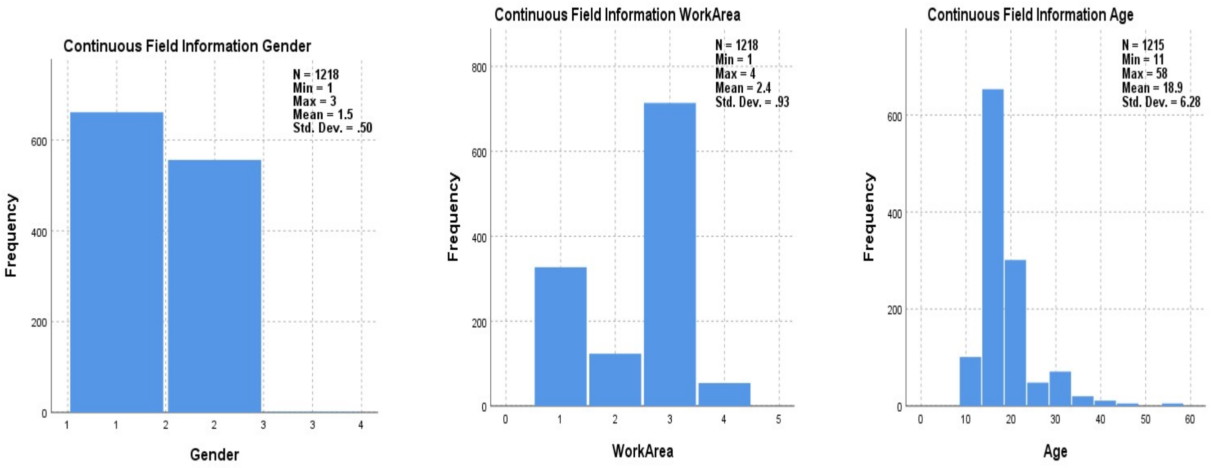
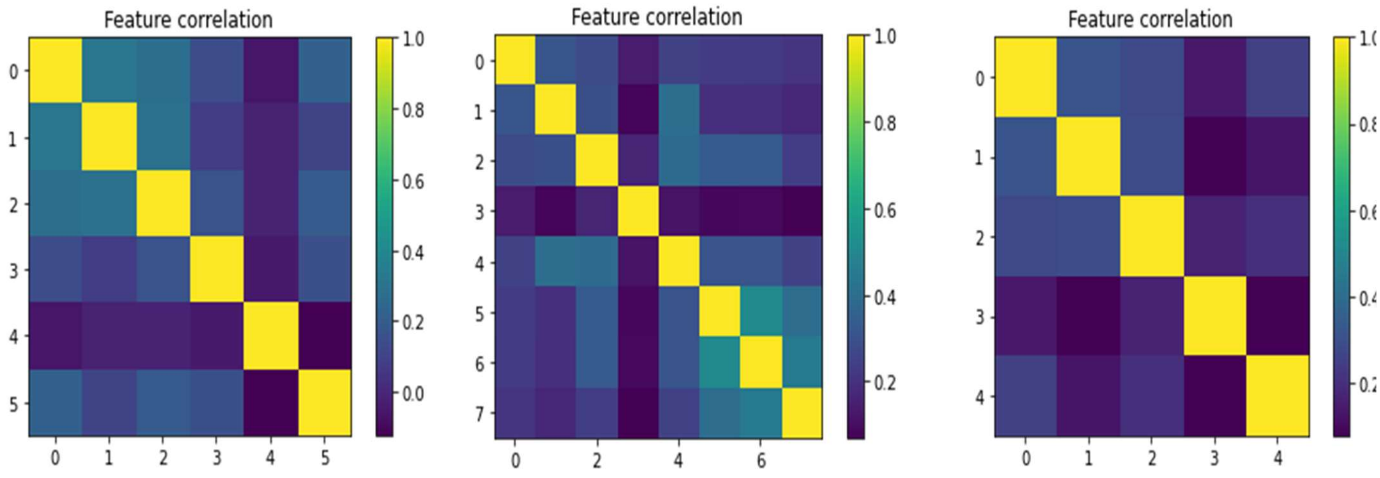


Figure 3.Demographic Distribution of Data

Table 3. Kruskal Wallis Test Summary

Questionnaire	p-value		
	Gender	**Age**	**Work Area**
Q1	0.130	**0.015**	0.276
Q2	0.858	0.498	**0.002**
Q3	0.537	**0**	0.451
Q4	0.229	0.443	0.503
Q5	0.275	**0.001**	**0**
Q6	0.840	**0.004**	0.113
Q7	**0.083**	**0.002**	0.109
Q8	0.767	0.426	0.244
Q9	0.191	**0.006**	**0.018**
Q10	0.180	0.087	0.001
Q11	0.591	**0.018**	**0.027**
Q12	0.106	**0.002**	0.709

Figures 4 (a), 4 (b), and 4 (c) depict the Pearson correlation analysis used to examine the relationship between age categories and various variables. The correlation is evaluated among questionnaire groups. Figure 4 (a) shows the correlation between mental health before and after the COVID-19 outbreak. Q4 and Q5 of the Questionnaire pertain to past experiences with stress, while Q2, Q3, Q6, and Q12 focus on mental health concerns arising after the COVID-19 outbreak. Therefore, six features are selected, and among them, correlation is analyzed. The correlation graph shows that Q4 and Q5 correlate highest with Q2, Q3, and Q12.



(a) (b) (c)

Figure 4. Correlation Analysis among Questionnaire Subgroups

Figure 4(b) shows correlation between social health (Q7, Q8, Q9, and Q10) and mental health after COVID-19 outbreak (Q2, Q3, Q6 and Q12). From this correlation, a total of eight features are selected. The correlation graph shows that Q9 shows the highest correlation with Q2, Q3, Q6, and Q12.

Similarly, figure 4(c) shows correlation of after COVID-19 outbreak (Q2, Q3, Q6 and Q12) with work environment (Q11). From the correlation graph, it is observed that Q11 shows a correlation with Q2, Q3, and Q12. Therefore, from correlation analysis, it can be concluded that the most contributing questionnaires are Q2, Q3, Q4, Q5, Q9, Q11, and Q12.

The criteria for selecting the features for correlation were based on factors such as mental health before and after the COVID-19 outbreak, social health, and work environment. The selected features are the respective questionnaires from each factor that are hypothesized to have significant relationships with each other to understand the impact of the pandemic on mental health.

Based on the correlation analysis presented, the following conclusions can be drawn:

- Past stress experiences (Q4, Q5) are strongly correlated with mental health concerns during the pandemic (Q2, Q3, Q12).
- Social health (Q9) is significantly correlated with post-pandemic mental health.
- The work environment (Q11) has a substantial impact on mental health during the pandemic.

VI. MACHINE LEARNING RESULT ANALYSIS

In this section, we have applied machine learning models to predict mental states based on questionnaire subgroups, as presented in the above section. The machine learning approaches used for result evaluation are Logistic Regression, Naïve Bayes, Random Forests, k-nearest Neighbors, and Artificial Neural Networks (ANN). Brief working of these algorithms are presented below:

Naive Bayes: It is a probabilistic type of classifier that is based on Bayes' theorem which is based on assumptions that features are independent of each other. In this algorithm it calculates the posterior probability of each class and finally chooses the one with the highest probability.

Support Vector Machine (SVM): It is one of the supervised learning algorithms that are commonly used for classification tasks. In this approach a hyper-plane is found that separates the data into different classes with the maximization of the margin. It is quite effective in high-dimensional spaces.

Logistic Regression: It is a statistical method used for binary classification that uses the logistic function for estimation of parameters.

Random Forest: It is an ensemble learning method that generates multiple decision trees during training and result their predictions based on voting strategy. It reduces over fitting and improves accuracy by averaging the results.

k-Nearest Neighbors (k-NN): It is an instance-based learning algorithm that predicts the output based on result for the average of the k-nearest neighbours in the feature space. k-It is non-parametric approach that uses the Euclidian distance metrices for computation.

Artificial Neural Network (ANN): It is a computational model that is designed on inspiration by the structure of human brain. It consists of interconnected neurons arranged in layers for training and prediction.

This paper has implemented and trained models in the Keras framework with Tensor Flow, with the training of the proposed model conducted on Google Colab. The collected dataset is divided into 70:30 ratios as training and testing samples for the learning process. The machine learning models are trained using training samples, and trained models are performed on testing samples. This section uses the following parameters to evaluate the performance of the above-stated machine learning techniques.

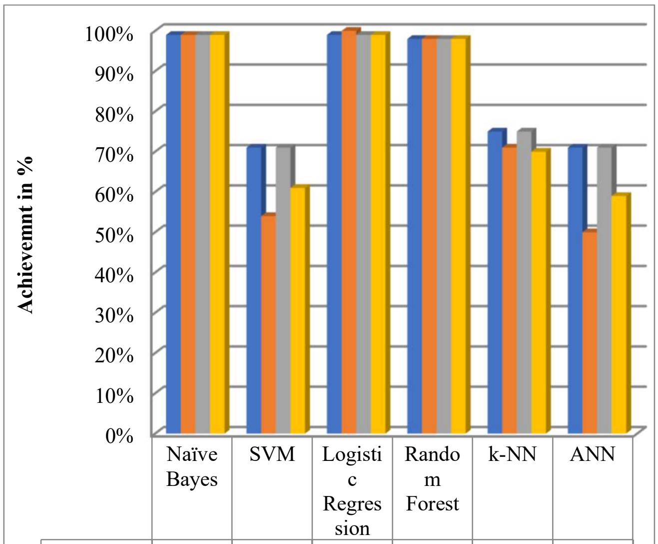
$$Accuracy = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalsePositive + F} \quad (i)$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \quad (ii)$$

$$Recall/Sensitivity = \frac{TruePositive}{TruePositive + FalseNegative} \quad (iii)$$

$$F1\_score = \frac{2}{1/precision + 1/Recall} \quad (iv)$$

Figure 5 shows the performance evaluation of machine learning for mental health evaluation considering different classifiers. Accuracy, precision-recall, and F1 score are the performance parameters for evaluating mental health. The accuracy of the Naive Bayes classifier is 99%, that of the ANN and SVMis 71%, and that of the logistic regression is 99%. Similarly, the accuracy and recall of random forest are 98%, whereas kNN indicates an accuracy of 75%. This result demonstrates that best result was achieved by Naive Bayes classifier, Logistic regression, and random forest.



	Naïve Bayes	SVM	Logistic Regression	Random Forest	k-NN	ANN
Accuracy	99%	71%	99%	98%	75%	71%
Precision	99%	54%	100%	98%	71%	50%
Recall	99%	71%	99%	98%	75%	71%
F1_score	99%	61%	99%	98%	70%	59%

Figure 5. Performance Evaluation of ML Algorithms for Mental Health Evaluation

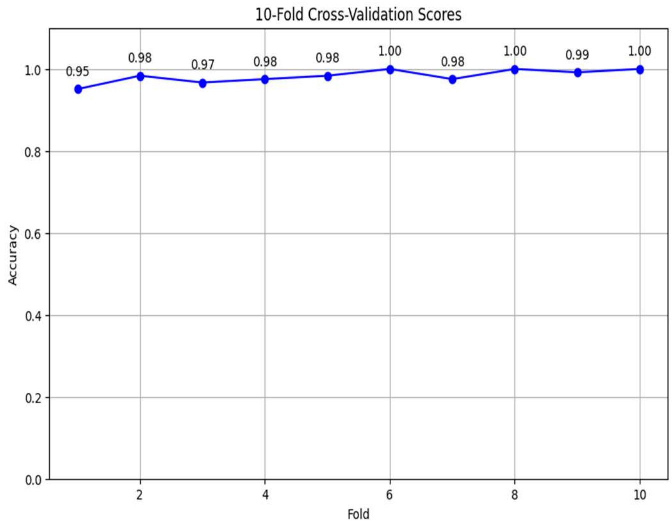


Figure 6. 10-Fold Cross Validation

Figure 6 presents the result of 10-fold cross-validation for the proposed dataset. The dataset sample is small therefore to validate the machine learning performance; 10-fold cross-validation is adopted. This is presented for naïve bayes classifier.

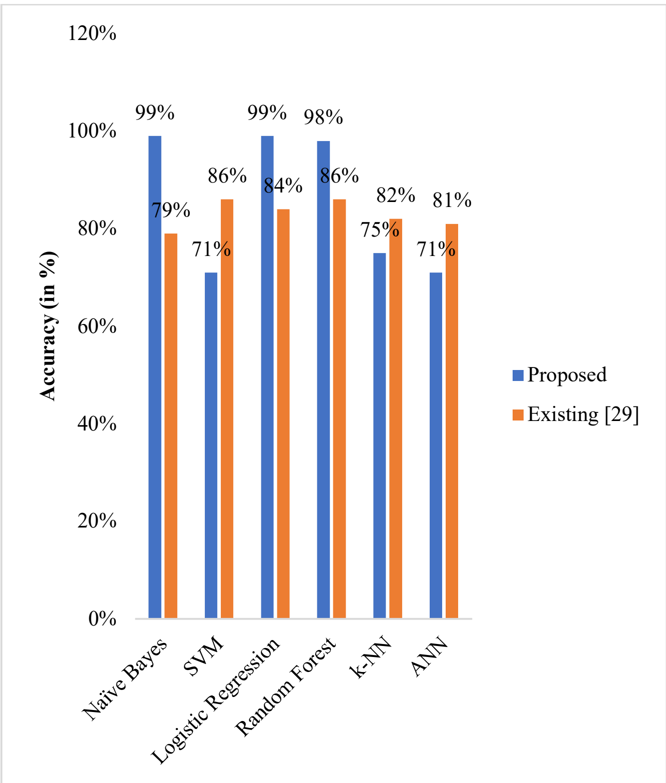


Figure 7. Comparative Accuracy Evaluation

Figure 7 presents the results of a comparative accuracy evaluation of six different classifiers: Naïve Bayes, SVM, Logistic Regression, Random Forest, k-NN, and ANN. The assessment compares the proposed method against the dataset presented in [29]. This graph is generated by implementing the machine learning approaches on the existing dataset [29] and the dataset collected in the proposed paper. The proposed method achieves higher accuracy. For Naïve Bayes and Logistic Regression, the proposed method achieves 99% accuracy, compared to 79% and 84%, respectively, for the existing method. The proposed method achieves 98% accuracy for Random Forest, compared to 86% for the existing method. For k-NN and ANN, the proposed method achieves 75% and 71% accuracy, respectively, compared to 82% and 81% for the existing method.

In Figure 8, the proposed method achieves an average improvement of 17.5% in precision over the existing method. For Logistic Regression and Naïve Bayes, the proposed method achieves 100% and 99% precision, respectively, compared to 77% and 76% for the existing method. The proposed method achieves 98% precision for Random Forest, compared to 88% for the existing method. For k-NN and SVM, the proposed method achieves 71% and 54% precision, respectively, compared to 75% and 74% for the existing method. The proposed method achieves 50% precision for ANN, compared to 78% for the existing method.

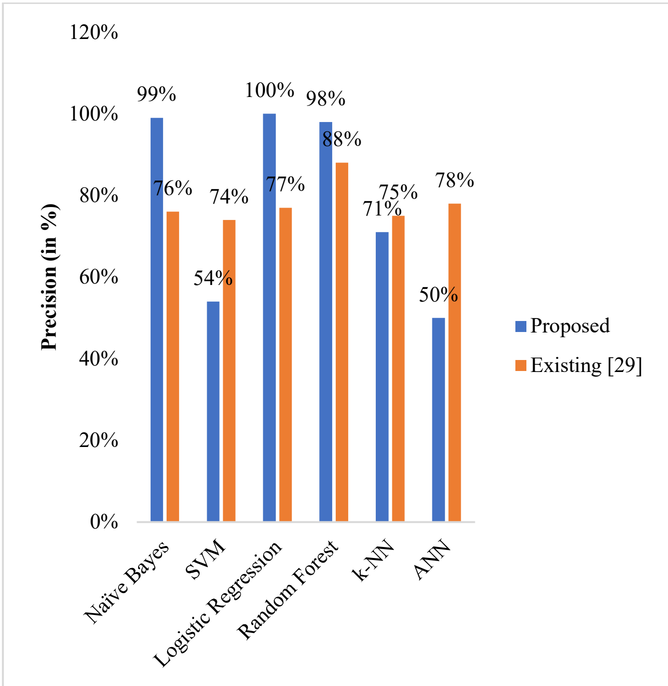


Figure 8. Comparative Precision Evaluation

In Figure 9, the proposed method achieves higher recall with an average improvement of 11.2% over the existing method. For Naïve Bayes and Logistic Regression, the proposed method achieves 99% recall, compared to 79% and 84% for the existing method. The proposed method achieves 98% recall for Random Forest, compared to 86% for the existing method. For k-NN and ANN, the proposed method achieves 75% and 71% recall, respectively, compared to 82% and 81% for the existing method. The proposed method achieves 71% recall for SVM, compared to 86% for the existing method.

In Figure 10, the proposed method achieves an average improvement of 16.5% F1-score over the existing method. For Naïve Bayes and Logistic Regression, the proposed method achieves a 99% F1-score, compared to 77% and 80% for the existing method. The proposed method achieves a 98% F1-score for Random Forest, compared to 80% for the existing method. For k-NN and SVM, the proposed method achieves 70% and 61% F1-score, respectively, compared to 78% and 79% for the existing method. The proposed method achieves a 59% F1-score for ANN, compared to 79% for the existing method.

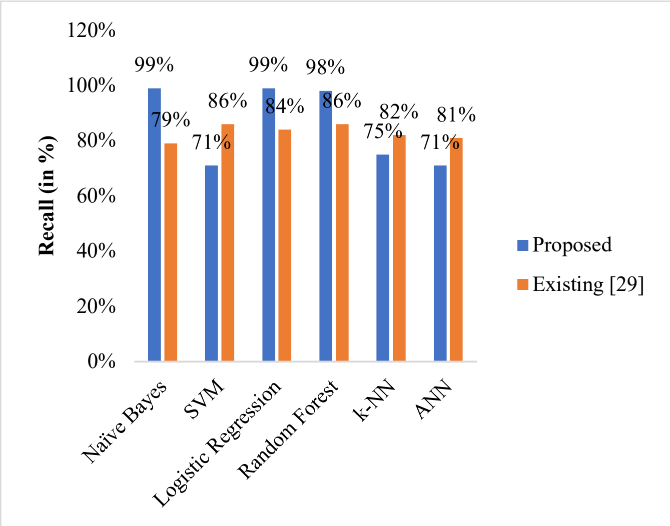


Figure 9. Comparative Recall Evaluation

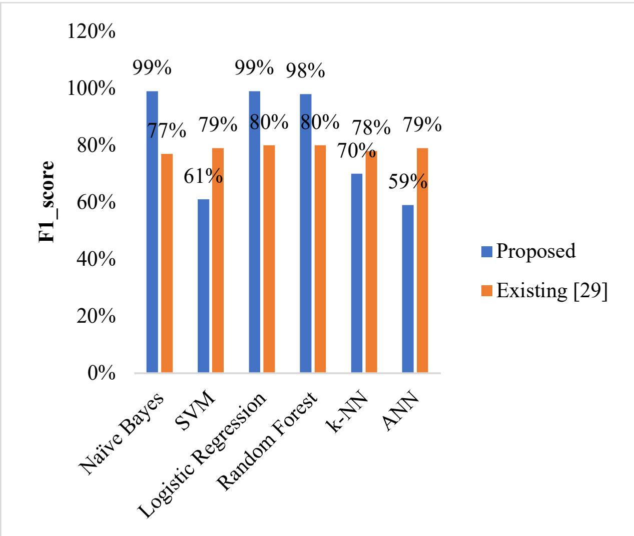


Figure 10. Comparative F1-score Evaluation

Overall, the proposed method performs better for most of the classifiers tested than the existing method.

7. CONCLUSION AND FUTURE SCOPE

This study demonstrated the potential of using machine learning algorithms to identify key features related to mental stress in the COVID-19 dataset. The specified features can help healthcare professionals identify patients at high risk of developing mental health problems and develop targeted intervention strategies. This study is prompted by the critical need to assist the experts in the field of mental health in making better clinical diagnostic decisions. With the help of ML techniques, future researchers can create an AI-integrated DSS system that can replace the more time-consuming and expensive paper-based tests currently used in the mental health industry. This study still has some limitations related to the sampling technique used. Future studies can build upon these findings by developing predictive models for mental health outcomes using the identified key features.

References

[1] T. Dolev, S. Zubedat, Z. Brand, B. Bloch, E. Mader, O. Blondheim, A. Avital, "Physiological parameters of mental health predict the emergence of post-traumatic stress symptoms in physicians treating COVID-19 patients," *Translational Psychiatry*, vol. 11, pp. 178-185, 2021. https://doi.org/10.1038/s41398-021-01299-6.

[2] M. Karunakaran, J. Balusamy and K. Selvaraj, "Machine learning models based mental health detection," *Proceedings of the Third International Conference on Intelligent Computing Instrumentation and Control Technologies (ICICICT)*, Kannur, India, 2022, pp. 835-842. https://doi.org/10.1109/ICICICT54557.2022.9917622.

[3] S. C. KoumetioTekouabou, D. El Bachir, R. Azmi, R. Jaligot, J. Chenal, "Reviewing the application of machine learning methods to model urban form indicators in planning decision support systems: Potential, issues and challenges," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, issue 8, part B, pp. 5943-5967, 2022. https://doi.org/10.1016/j.jksuci.2021.08.007.

[4] S. Safdar, S. Zafar, N. Zafar, et al., "Machine learning based decision support systems (DSS) for heart disease diagnosis: A review," *Artificial Intelligence Review*, vol. 50, pp. 597–623, 2018. https://doi.org/10.1007/s10462-017-9552-8.

[5] R. Kumar, S. Mukherjee, T.-M. Choi, L. Dhamotharan, "Mining voices from self-expressed messages on social-media: Diagnostics of mental distress during COVID-19," *Decision Support Systems*, vol. 162, 113792, pp. 104-111, 2022, https://doi.org/ 10.1016/j.dss. 2022.113792.

[6] D. Benrimoh, R. Fratila, S. Israel, K. Perlman, N. Mirchi, S. Desai, A. Rosenfeld, S. Knappe, J. Behrmann, C. Rollins, R. P. You, T. Team, "Aifred Health, a deep learning powered clinical decision support system for mental health," In: Escalera, S., Weimer, M. (eds) *The NIPS'17 Competition: Building Intelligent Systems. The Springer Series on Challenges in Machine Learning*. Springer, Cham., pp. 204-211, 2018, https://doi.org/10.1007/978-3-319-94042-7_13.

[7] G. González Rodríguez, J.M. Gonzalez-Cava, J. A. Méndez Pérez, "An intelligent decision support system for production planning based on machine learning," *J Intell Manuf*, vol. 31, pp. 1257–1273, 2020. https://doi.org/10.1007/s10845-019-01510-y.

[8] S.Tutun, M.E. Johnson, A. Ahmed, et al., "An AI-based decision support system for predicting mental health disorders," *Inf Syst Front*, vol. 25, pp. 1261–1276, 2023.https://doi.org/10.1007/s10796-022-10282-5.

[9] Z. Zhou, D. Luo, B. X. Yang, Z. Liu, "Machine learning-based prediction models for depression symptoms among chinese healthcare workers during the early COVID-19 outbreak in 2020: A cross-sectional study,"*Front. Psychiatry*,vol. 13, 876995, 2022.https://doi.org/10.3389/fpsyt.2022.876995.

[10] A. Kumar, "Machine learning for psychological disorder prediction in Indians during COVID-19 nationwide lockdown," pp. 161–172, 2021. https://doi.org/10.3233/IDT-200061.

[11] P. Gupta, et al., "Predictive modeling of stress in the healthcare industry during COVID-19: A novel approach using XGBoost, SHAP values, and tree explainer," *IJDSST*, vol. 15, no. 1, pp. 1-20, 2023. https://doi.org/10.4018/IJDSST.315758.

[12] L. Flesia, M. Monaro, C. Mazza, V. Fietta, E. Colicino, B. Segatto, P. Roma, "Predicting perceived stress related to the COVID-19 outbreak through stable psychological traits and machine learning models," *J. Clin. Med.*, vol. 9, 3350, 2020. https://doi.org/10.3390/jcm9103350.

[13] Q. Abbas, A. R. Baig, A.Hussain, "Classification of Post-COVID-19 emotions with residual-based separable convolution networks and EEG signals," Sustainability, vol. 15, 1293, 2023. https://doi.org/10.3390/su15021293.

[14] N. Bhirud, S. Tatale, P. Jain, A. Kulkarni, M. Panchpor, N. K. Jain, "Development of psychiatric COVID-19 chatbot using deep learning," in: Dey, N. (eds) *Data-Driven Approach for Bio-medical and Healthcare, Data-Intensive Research*. Springer, Singapore, 2023. https://doi.org/10.1007/978-981-19-5184-8_10.

[15] "COVID-19 Student Mental Health Survey", Kaggle, 2022, [Online]. Available at: https://www.kaggle.com/datasets/COVID-19-student-mental-health-survey

[16] "KaggleCOVID-19 Mental Health Survey," Kaggle, 2022. [Online]. Available at: https://www.kaggle.com/datasets/kaggle-COVID-19-mental-health-survey

[17] J. Smith, et al., "Predicting mental health in students using machine learning feature selection algorithms," *International Journal of Machine Learning and Computing*, vol. 10, no. 1, pp. 72-79,2022.

[18] T. Nguyen, et al., "Machine learning feature selection for mental health prediction in students during COVID-19," *Proceedings of the International Conference on Machine Learning and Data Engineering*, 2022, pp. 113-120.

[19] Y. Kim, et al., "A comparative study of feature selection algorithms for mental health prediction in students during COVID-19," *Proceedings of the International Conference on Data Mining and Machine Learning*, 2022, pp. 65-72.

[20] S. Hossain, M. M. Islam, M. M. Rahman, "A predictive model for mental health assessment in students during COVID-19 pandemic," *Proceedings of the 2022 IEEE International Conference on Machine Learning and Data Engineering (iCMLDE)*, April 2022, pp. 1-6.

[21] A. Singh, A. Verma, A.Goyal, "Mental health prediction of students during COVID-19 using machine learning," *Proceedings of the 2022 IEEE International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)*, March 2022, pp. 1-6.

[22] S. Sharma, S. Gupta, A. Singh, "Predicting mental health issues among students using machine learning algorithms during the COVID-19 pandemic," *Proceedings of the 2022 IEEE International Conference on Inventive Computation Technologies (ICICT)*, March 2022, pp. 1187-1192.

[23] R. Kaur, A. Kaur, S. Kaur,"A machine learning-based approach for predicting mental health in students during the COVID-19 pandemic," *Proceedings of the IEEE International Conference on Computing, Communication and Security (ICCCS)*, March 2022, pp. 1-6. https://doi.org/10.1109/SMARTGENCON56628.2022.10083995.

[24] N. Sood, M. Singh, N. Kumar,"Predicting mental health of students using machine learning techniques during the COVID-19 pandemic," *Proceedings of the IEEE International Conference on Advances in Computing, Communication, and Control (ICAC3)*, March 2022, pp. 1-6.

[25] R. Singh, R. Gupta, A. Kaur, "Mental health prediction in students using machine learning feature selection algorithms," *Journal of Education and Health Promotion*, vol. 10, issue 1, 238, 2021.

[26] S. Li, Y. Liu, Y. Yang, "Predicting anxiety in college students during the COVID-19 pandemic: A machine learning approach," *Frontiers in Psychology*, vol. 12, 659455, 2021.

[27] D. H. Kim, S. Y. Kim, S. H. Kim, S. K. Lee, "Machine learning-based prediction of stress in college students during the COVID-19 pandemic," *International Journal of Environmental Research and Public Health*, vol. 18, issue 3, 1125, 2021. https://doi.org/10.3390/ijerph182111381.

[28] Q. H. Nguyen, T. T. Luu, T. H. Nguyen, D. T. Hoang, "Depression prediction in university students during the COVID-19 pandemic: A machine learning approach," *International Journal of Environmental Research and Public Health*, vol. 18, issue 6, 2889, 2021.

[29] K. Chaturvedi, D. Vishwakarma, N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A Survey," *Children and Youth Services Review*, vol. 121, 105866, 2020. https://doi.org/10.1016/j.childyouth.2020.105866.

[30] C. Iwendi, A. K. Bashir, A. Peshkar, R. Sujatha, J. M. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, O. Jo, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Health*, vol. 8, 357, 2020. https://doi.org/10.3389/fpubh.2020.00357.

[31] H. Peiqing, "Multidimensional state data reduction and evaluation of college students' mental health based on SVM," *Journal of Mathematics*, vol. 2022, pp. 1-11, 2022. https://doi.org/10.1155/2022/4961203.

[32] T. Alakuş, I. Turkoglu, "Comparison of deep learning approaches to predict COVID-19 infection," *Chaos, Solitons & Fractals*, vol. 140, 110120, 2020. https://doi.org/10.1016/j.chaos.2020.110120.

[33] M. S. H. Mukta, S. Islam, S. Shatabda, M. E. Ali, A. Zaman, "Predicting academic performance: Analysis of students' mental health condition from social media interactions," *Behavioral Sciences*, vol. 12, issue 4, 87, 2022. https://doi.org/10.3390/bs12040087.

[34] R. Rois, M. Ray, A. Rahman, S. Roy, Swapan, "Prevalence and predicting factors of perceived stress among Bangladeshi university students using machine learning algorithms," *Journal of Health, Population and Nutrition*, vol. 40, 50, 2021. https://doi.org/10.1007/s00445-021-01468-x.

[35] J. Kim, et al., "Prediction model for student depression using machine learning algorithms," *Healthcare Informatics Research*, vol. 26, issue 1, pp. 15-22, 2020.

[36] S. Rijal, et al., "A machine learning approach to assess student mental health using data mining techniques," *International Journal of Advanced Computer Science and Applications*, vol. 11, issue 9, pp. 357-363, 2020.

[37] H. Ahmed, et al., "Prediction of anxiety and stress among university students due to COVID-19 using machine learning algorithms," *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, 6566, 2020.

[38] S. Park, et al., "Predictive model for student mental health using machine learning algorithms during the COVID-19 pandemic," *Healthcare Informatics Research*, vol. 26, issue 4, pp. 284-291, 2020.

[39] L. Li, et al., "Analysis of the mental health status of college students during the COVID-19 pandemic using machine learning algorithms," *Journal of Healthcare Engineering*, vol. 12, 14965, 2020.

[40] Kaggle. Mental Health and Wellbeing Survey. [Online]. Available at: https://www.kaggle.com/osmi/mental-health-in-tech-survey

[41] Kaggle. World Mental Health Survey Initiative. [Online]. Available at: https://www.kaggle.com/jonathanbouchet/world-mental-health-survey-initiative

[42] Kaggle. COVID-19 Symptom Survey.[Online]. Available at: https://www.kaggle.com/imdevskp/COVID19-symptoms-checker

[43] Kaggle. Youth Risk Behavior Surveillance System. [Online]. Available at: https://www.kaggle.com/cdc/yrbss

[44] Kaggle. National Survey of Children's Health. [Online]. Available at: https://www.kaggle.com/jchen2186/national-health-and-nutrition-examination-survey

[45] A. Shukla, A. Singh, A. Bajpai, "Predicting mental health of Indian students during COVID-19 using machine learning techniques," *Journal of Educational Computing Research*, vol. 58, issue 8, pp. 1421-1437, 2020.

[46] A. Kumar, S. Dua, "Feature selection for mental health prediction models in Indian students during COVID-19," *Journal of Medical Systems*, vol. 44, issue 12, 225, 2020.

[47] S. Verma, S. Garg, R. Gupta, "COVID-19 and mental health of Indian students: a predictive model using machine learning algorithms," *Asian Journal of Psychiatry*, vol. 54, 102437, 2020.

[48] X. Zhang, F. T. S. Chan, C. Yan, I. Bose, "Towards risk-aware artificial and machine learning systems: An overview," *Decision Support Systems*, vol. 159, 113800, 2022. https://doi.org/10.1016/j.dss.2022.113800.

[49] K. Kumar, P. Kumar, D. Deb, M.-L. Unguresan, V. Muresan, "Artificial intelligence and machine learning based intervention in medical infrastructure: A review and future trends," Healthcare, vol. 11, 207, 2023. https://doi.org/10.3390/healthcare11020207.

Saini Poonam, a Ph.D. Scholar at the Department of Computer Science and Engineering, M. M. Engineering College, MMDU, Mullana, India, is deeply immersed in the realms of Machine Learning, Deep Learning, and Artificial Intelligence.

Dr. Neera Batra is working as Associate Professor at Maharishi Markandeshwar University, Mullana, specializing in Human-Computer Interaction, Pervasive Computing, IoT, Databases, and Distributed Computing.