

Predictive machine learning model for mental health issues in higher education students due to COVID-19 using HADS assessment

ML model for
mental health
in college
students

Received 22 February 2023
Revised 10 July 2023
Accepted 16 October 2023

Reshma Krishnan

Muscat College, Muscat, Oman

Shantha Kumari

SRM Institute of Science and Technology, Chennai, India

Ali Al Badi

Gulf College, Maabelah, Oman, and

Shermina Jeba and Menila James

Muscat College, Muscat, Oman

Abstract

Purpose – Students pursuing different professional courses at the higher education level during 2021–2022 saw the first-time occurrence of a pandemic in the form of coronavirus disease 2019 (COVID-19), and their mental health was affected. Many works are available in the literature to assess mental health severity. However, it is necessary to identify the affected students early for effective treatment.

Design/methodology/approach – Predictive analytics, a part of machine learning (ML), helps with early identification based on mental health severity levels to aid clinical psychologists. As a case study, engineering and medical course students were comparatively analysed in this work as they have rich course content and a stricter evaluation process than other streams. The methodology includes an online survey that obtains demographic details, academic qualifications, family details, etc. and anxiety and depression questions using the Hospital Anxiety and Depression Scale (HADS). The responses acquired through social media networks are analysed using ML algorithms – support vector machines (SVMs) (robust handling of health information) and J48 decision tree (DT) (interpretability/comprehensibility). Also, random forest is used to identify the predictors for anxiety and depression.

Findings – The results show that the support vector classifier produces outperforming results with classification accuracy of 100%, 1.0 precision and 1.0 recall, followed by the J48 DT classifier with 96%. It was found that medical students are affected by anxiety and depression marginally more when compared with engineering students.

Research limitations/implications – The entire work is dependent on the social media-displayed online questionnaire, and the participants were not met in person. This indicates that the response rate could not be evaluated appropriately. Due to the medical restrictions imposed by COVID-19, which remain in effect in 2022, this is the only method found to collect primary data from college students. Additionally, students self-selected themselves to participate in this survey, which raises the possibility of selection bias.

Practical implications – The responses acquired through social media networks are analysed using ML algorithms. This will be a big support for understanding the mental issues of the students due to COVID-19 and

© Reshma Krishnan, Shantha Kumari, Ali Al Badi, Shermina Jeba and Menila James. Published in *Arab Gulf Journal of Scientific Research*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

This publication is part of the Research project funded by the Sultanate of Oman Ministry of Higher Education, Research and Innovation (MoHERI), Oman, with reference code: MoHERI/BFP/MC/01/2020.



can taking appropriate actions to rectify them. This will improve the quality of the learning process in higher education in Oman.

Social implications – Furthermore, this study aims to provide recommendations for mental health screening as a regular practice in educational institutions to identify undetected students.

Originality/value – Comparing the mental health issues of two professional course students is the novelty of this work. This is needed because both studies require practical learning, long hours of work, etc.

Keywords Predictive analysis, Early identification, Anxiety, Depression, HADS

Paper type Case study

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused the widespread communicable disease known as coronavirus disease 2019 (COVID-19), causing the pandemic outbreak. As of August 2022, the World health Organisation officially recorded 589,680,368 confirmed COVID-19 cases, which include 6,436,519 deaths ([World Health Organization, 2022](#)). COVID-19 is an infectious disease that can lead to death in some cases, and strict lockdown was imposed all over the world. As a result, everyday survival of layperson became extremely uncertain ([Atchison et al., 2021](#)).

The medical community has been working hard since the pandemic began to contain the impact of SARS-CoV-2 virus on the health of COVID-positive patients. In parallel, clinical psychologists issued a warning regarding the intensity of problems corresponding to the mental well-being that arose as a result of the post-traumatic stress and depression ([Mazza et al., 2020](#); [Shader, 2020](#); [Pramukti et al., 2020](#)). Recent findings demonstrated that demographic information like age, residence location, economic status, etc., habits like how a person sleeps and what kind of family a person have and education, such as subject, duration of study and so on, are the pressures that contributed to an increase in mental health problems ([Alimoradi et al., 2022](#)).

The educational sector suffered the most from the pandemic's unfavourable and uncertain conditions. Due to the severe lockdown, the teaching and learning process in the educational sector underwent a radical transformation. The move to online or remote education aimed to make learning easier. However, this type of knowledge transfer's success rate could not achieve the educational system's objectives ([Jena, 2020](#)).

Numerous COVID-19 waves have occurred since December 2019. Young people between the ages of 6 and 21 were unable to comprehend the shift in their educational patterns, and those between the ages of 18 and 22 were particularly affected ([Chu & Li, 2022](#)). During COVID-19, this group's psychological and physiological patterns were severely disrupted, as students began to stay at their houses fulltime ([Cahyadi, 2021](#)). Major studies conducted during this time pointed to the following psychological issues such as stress with depression ([Houben-Wilke et al., 2022](#)), anxiety ([Malta, Julian Bond, Smith, & Naomi, 2022](#); [Bourmistrova, Solomon, Braude, Strawbridge, & Carter, 2022](#)), apprehensions ([Trógolo, Moretti & Medrano, 2022](#)), reduced physical activity ([López-Valenciano, Alejandr, Suárez-Iglesias, Sanchez-Lastra & Ayán, 2021](#)) etc. that students experienced during COVID-19 waves. Schools are opening again after strict lockdowns, and the normal mode of education is catching up. Nevertheless, post-traumatic COVID-19 effects and re-entry anxiety are common amongst college students pursuing professional degrees in fields like engineering, medicine and other fields.

This is study presents a machine learning (ML) methodology towards the early finding of students who suffer from various kinds of mental illness due to COVID-19 even in 2022. The reason for choosing the medical and engineering students as main interest because (1) both are professional courses that requires practical learning, (2) the knowledge transfer through online classes is found to be inadequate with respect to the classroom teaching ([Shyadligeri, Vaz, & Lokapure, 2022](#)) and (3) the course content, curriculum and evaluation patterns are very cumbersome, etc. ([Kecojevic, Basch, Sullivan, & Davi, 2020](#)).

This paper proposes ML models to classify “early” the mental health-affected medical- and engineering-pursuing students based on the severity levels. This study tries to bring out the impact of mental health with respect to the education course that would help to provide personalised attention to different students. Initially, the responses from various groups of students are collected using Hospital Anxiety and Depression Scale (HADS) Questionnaire (Snaith, 2003).

The global SARS-CoV-2 outbreak and subsequent lockdown had a significant impact on people’s daily lives, with strong implications for stress levels due to the threat of contagion and restrictions to freedom. Given the link between high stress levels and adverse physical and mental consequences, the COVID-19 pandemic is certainly a global public health issue (Flesia *et al.*, 2020).

These responses are analysed both statistically as well as scientifically using various ML mechanisms like support vector machines (SVMs), decision trees (DTs) and random forest (Priya, Garg, & Tigga, 2020; Sau & Bhakta, 2019). By constructing models for prediction and classification that complement the findings of statistical analysis, ML has the advantage of being able to reveal previously unknown associations and trends (Alghamdi *et al.*, 2017).

This work has the following objectives: (1) to evaluate the mental well-being after COVID-19 in terms of anxiety and depression through ML algorithms amongst students pursuing medical and engineering and rank them; (2) to model the collective inputs received through HADS from the student’s community scientifically and (3) to have an early identification of weak students who may need urgent medical support using ML classifiers.

2. Related works

Earlier works provided enough evidence for educational and employment details, family type, gender, marital status, age (Shyadligeri *et al.*, 2022; Kecojevic *et al.*, 2020; Ciro *et al.*, 2020; Xiao *et al.*, 2020), etc. are the factors that contribute to mental health issues prevalent across students pursuing higher education. Plenty of research works based on statistical measures (SonSudeep, SmithXiaomei, & Sasangohar, 2020; Wang, 2020; Chenyang Lin, 2022) are available that comprehended the concerns of mental health impact amongst college students during the pandemic. The findings from the above listed works conclude that poor mental health prevailed amongst college students due to COVID-19.

The authors in Alharthi (2020) used ML models – AdaBoost and neural networks to identify the danger level of anxiety amongst students. This research identified that gender, support system and income of a family are the significant driving factors for the anxiety levels. The works presented in Donna Wang (2021), Zeng Xiaofang and Tingzeng (2021) supported our methodology, which also used a similar methodology for collecting the responses through a standard survey and analysing through ML algorithms. These works used hierarchical multiple regression analysis and found that age of the student, friends and family support, support from school, usage of alcohol and drugs, measures taken by government, exposure to COVID-19-related news and life imbalance contribute in a distributive manner towards the ill mental health.

Like the above listed work, the researchers in Herbert Cornelia (2021) used the outputs from the extensive survey that was conducted to investigate the mental health, behaviour and experiences during COVID-19 amongst Egyptian college students. Subsequently, this work showcased the usage of gradient boosting regression and support vector regression to understand the life pattern of students. The results indicated the stress and anxiety due to issues in online and self-regulated learning.

Works including Ren (2021) and Khattar (2020) used various ML models to study COVID-19 effects on the mental health of students in China and India, respectively. The outcome of Ren (2021) listed that isolation and families’ income are the major pointers

towards anxiety and depression, whilst [Khattar \(2020\)](#)'s outcome states that online learning is not inclusive and could not provide the required knowledge transfer.

Studies were also available to understand the effectiveness of online learning process, which was prevalent during the COVID-19 lockdown period. The works in [Al-Mawee Wassnaa \(2021\)](#), [Shivangi \(2020\)](#) and [Gonzalez-Ramirez \(2021\)](#) brought out the results that give the overall view of both advantages and limitations of online mode of learning. Furthermore, this work listed that the year of study of a student (fresher, second year or final year) has a relationship with the effectiveness of the online learning classes. Mostly, the students felt that there is no adequate connection or interaction with the faculty and friends. Also, they could not have the self-assurance of learning the course content.

Several works are available to assess the mental health issues prevalent amongst various streams of people. However, there is a void in measuring the health status of students undergoing two different professional courses – medical and engineering sciences specifically. These professional course faculties struggled a lot to conduct the classes in online mode. The students pursuing these studies have already taken up several admission tests, and when they could not be served properly, there enters the reasons for stress.

As a case study, students from southern India are chosen here because medical and engineering courses are popular amongst the students as well as their parents for several years. Hence, it is very important to study individually as well as comparatively how these students cope up the pressure during COVID-19. This paper's objective aligns with that purpose.

The outcome of this work contributes to clinical psychology screening process through the ML models as well as provides a platform for analysing the mental health parameters of students pursuing professional courses.

3. Methods

In this research paper, participants were enlisted based on their individual interests. The survey questions were disseminated through various social media platforms, ensuring a broad reach beyond any particular college. The recruitment process involved leveraging existing connections to create a network effect, resembling the snowball sampling technique. Snowball sampling, a non-probability sampling method, involves the recruitment of new participants by existing participants, allowing for the inclusion of individuals with specific characteristics that may be challenging to identify through other means.

3.1 Participants and survey design

A Google Forms-based online questionnaire is used to conduct this cross-sectional survey. This survey took place in June and July of 2022 and was purely self-administered. Google Forms is distributed to undergraduate, graduate and PhD students studying engineering, medicine and allied health fields.

The purpose of the survey is stated in the first section, which also states that responses can be submitted anonymously. The second section collected the personal as well as educational profile of the students. As already stated, the HADS questions to evaluate anxiety and depression form the third and fourth section, respectively.

Social media is used to distribute this questionnaire to college students, who are asked to fill it out with helpful responses. The respondents were identified using a snowball sampling strategy and simple random sampling methods. The limitations of each of these sampling methods are overcome by combining them. As a result, this study's combined sampling method has advantages such as reducing selection bias and reaching unknown respondents through referrals.

The sample collected includes responses from 272 participants (152 engineering students, 88 medical students and 32 allied medicine students). The allied medicine and medical students are considered together.

3.2 Measures

This work is concerned primarily to evaluate the anxiety and depression severity prevalent in medical and engineering students. For this purpose, the questionnaire is prepared to collect the demographic details as well as the inputs for measuring the anxiety and depression previously published and validated HADS ([Snaith, 2003](#)).

3.2.1 Demographic information. To establish a connection between the sample population's attributes and the measuring variable, demographic data are required. The following categories are used to collect this information:

- (1) Individual's age, gender and lifestyle

The term "lifestyle" refers to a set of questionnaire items or questions that aim to capture various aspects of an individual's daily routine, habits and behaviour. These questions may explore areas such as sleep patterns, exercise habits, diet, social interactions, leisure activities and other relevant factors that contribute to a person's overall lifestyle. By including lifestyle-related questionnaires in the study, the researchers aim to assess the potential impact of these lifestyle factors on the mental health of higher education students during the COVID-19 pandemic.

- (2) Branch of study, year of study, current location and level of study (UG/PG/PhD);
- (3) Family type and economic information.

As independent variables, they either directly or indirectly contribute to symptoms of anxiety and depression; thus, the aforementioned variables are used. They are collectively referred to as covariates.

3.2.2 Evaluating mental health using Hospital Anxiety and Depression Scale (HADS). [Zigmond and Snaith \(1983\)](#) created the HADS to assess patients' anxiety and depression. It is now used as a standard for both clinical research and practice. The HADS has totally 14 questions in which each of 7 contribute to anxiety and depression, respectively. Anxiety-related questions were focussed to capture the fear, tense, worrying thoughts, restless, panic feelings, etc. prevalent in students thoughts. Likewise, the questions related to depression aim to elucidate the factors contributing to abrupt changes in students' behaviour resulting from the impact of COVID-19. However, it is important to note that the questionnaire does not specifically target the measurement of anxiety and depression solely related to COVID-19. In the HADS survey, each response is scored on a four-point scale, with a maximum score of 21. The severity levels of anxiety and depression are divided into normal (0–7), borderline abnormal (8–10) and abnormal (11–21) cases based on the scores ([Snaith, 2003](#)).

3.2.3 Machine learning methodologies. There is the need to model the multiple inputs and parameters received as responses through survey altogether to understand the psychological effect on the students due to COVID-19. Hence, there is the need for such ML algorithms that are capable of capturing the multi-variant distribution available in the survey responses are used. Based on this observation, SVM and DT are used in this paper. SVM serves as both classifier and regressor. It has been found to be effective in medical applications. Similarly, DT is a simple classifier that gives clarity outputs.

In order to comprehend their impact on scores for anxiety and depression, it is vital to understand the highly impacting pointers from the personal profile of the students. A supervised random forest algorithm creates individual DTs that work as an ensemble. For training, each DT utilises distinct samples. The random forest makes use of the majority of votes generated by

individual trees as a classifier. The statistical reflection of each significant factor on anxiety and depression can be calculated using this type of ensemble technique. Thus, overall architecture of the ML methodology applied in this paper is presented as follows in [Figure 1](#).

4. Statistical data analysis

The study had different survey questions focussing on demographic details, HADS questions on anxiety and depression and online vs offline mode of learning. The online vs offline mode of learning section is removed for this paper, since it is considered to be irrelevant for the chosen objective. The subsequent section lists the numerical attributes of the survey responses categorically.

4.1 Demographic characteristics and personal perspectives

The detailed demographic data about the sample population are given in [Table 1](#). At least 80% of the sample population are from the age group between 18 and 24; 55.88% of engineering stream-based students have responded to the survey and they study from second to final year. Most of the students stay with their family (62%). Nearly half respondents' household are middle-income families (52%), and many people reside in urban and metro cities. At least, 42% students' sleep schedule got increased during COVID-19.

4.2 Responses to the HADS questionnaires-statistical assessment of anxiety and depression

This section summarises the responses to anxiety and depression questions framed according to HADS. The assessment based on HADS scale is shown in [Table 2](#).

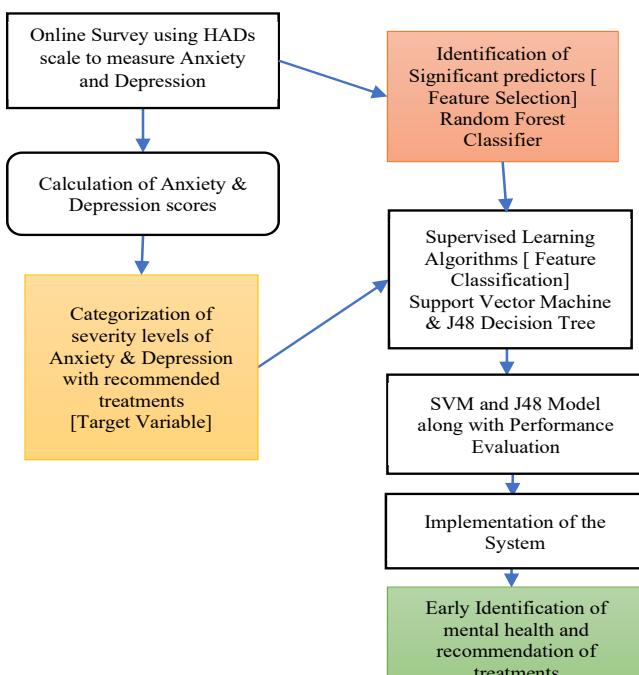


Figure 1.
ML architecture used in this work

Factors	Values	Engineering students (%)	Medical students (%)	Total students (%)	ML model for mental health in college students
Age	18 - 20	152	120	272	
	21 - 24	55.26	45.83	51.10	
	25 - 30	30.26	28.33	29.41	
	31 - 39	5.26	4.16	4.78	
	40 - 50	7.23	5	6.25	
Gender	Male	1.3155	0	0.74	
	Female	67.11	60.00	63.97	
	Prefer not to say	32.24	38.33	34.93	
Study field	Medicine	0.66	1.67	1.10	
	Allied medicine	11.76		11.76	
	Engineering	32.35		32.35	
Level of study	UG	55.88		55.88	
	PG	76.97	74.17	75.74	
	PhD	13.16	25.83	18.75	
Year of study	First year	9.87	0.00	5.51	
	Second year	17.76	23.33	20.22	
	Third year	36.84	26.67	32.35	
	Fourth year	28.95	18.33	24.26	
	Fifth year	16.45	15.00	15.81	
Staying in hostel	Yes	16.67		7.35	
	Yes, only on weekdays	19.08	55.83	35.29	
	No	0.66	4.17	2.21	
Number of members in the family	<4	80.26	40.00	62.50	
	>4	17.76	18.33	18.01	
Type of family	Nuclear	82.24	81.67	81.99	
	Joint	80.26	72.50	76.84	
Family income level	Economically weak	19.74	27.50	23.16	
	Lower middle class	5.26	4.17	4.78	
	Middle class	16.45	10.83	13.97	
	Upper middle class	64.47	37.50	52.57	
	Upper class	13.16	31.67	21.32	
Living place	Upper class	0.66	15.83	7.35	
	Rural (village)	13.82	5.83	10.29	
	Semi-urban (Tier-3 cities)	18.42	27.50	22.43	
	Urban (Tier-2 cities)	38.16	31.67	35.29	
Sleeping hours (During COVID)	Metro cities (Tier-1 cities)	28.95	35.00	31.62	
	<4 hours	3.95	24.17	12.87	
	4 - 6 hours	30.92	18.33	25.37	
	6 - 8 hours	48.68	35.83	43.01	
Sleeping pattern (During COVID)	>8 hours	17.11	21.67	19.12	
	Mostly during day time	72.37	65.00	69.12	
	Mostly during night-time	27.63	35.00	30.88	
Sleep schedule (During COVID)	Increased	40.13	45.83	42.65	
	Decreased	30.26	28.33	29.41	
	Normal	30.26	25.83	28.31	

Table 1.
Respondents' demographic characteristics

In discussion with the clinical psychologists, it is recommended for any student in abnormal anxiety or depression to be treated as in-house patient, students diagnosed with borderline anxiety or depression are to be treated as out-patient regularly with counselling and the students with low score in anxiety and depression are said to be normal. Based on these recommendations, the following [Table 3](#) projects the type of treatments needed by subset population of the sample considered.

This table shows that only 30.88% of students are fit mentally and medical students are having higher abnormal mental health issues when compared to engineering students as per the following [Figure 2](#).

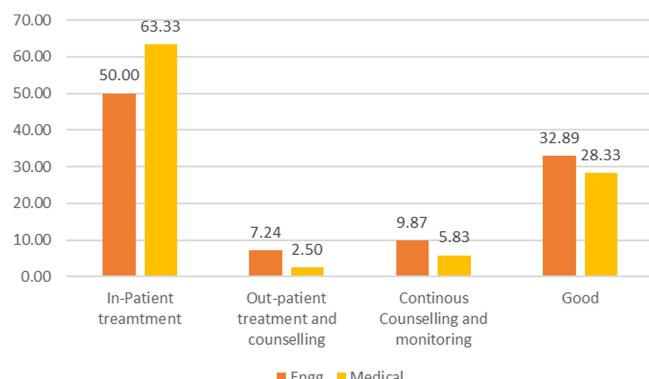
Table 2.
Severity of mental
health parameters

Factors	Values	Engineering (%)	Medical (%)	Total (%)
Anxiety	Normal	152	120	272
	Borderline	50	30	41.17
	Abnormal	17.76	35.83	25.73
Depression	Normal	32.23	34.16	33.08
	Borderline	42.11	35.00	38.97
	Abnormal	25.66	38.33	31.25
Anxiety and depression	Normal	32.24	43.33	37.13
	Borderline	32.89	30.00	31.62
	Abnormal	7.24	12.50	9.56
		14.47	23.33	18.38

Table 3.
Treatment
recommendations

Factors	Values	Engg (%)	Medical (%)	Total (%)
Abnormal anxiety and depression	In-patient treatment	152	120	272
Borderline anxiety and depression	Out-patient treatment and counselling	50.00	63.33	55.88
Normal or borderline anxiety and depression	Continuous counselling and monitoring	7.24	2.50	5.15
Normal anxiety and depression	Good	9.87	5.83	8.09
		32.89	28.33	30.88

Figure 2.
Comparative analysis
of medical and
engineering students'
mental health issues



5. Data analysis using machine learning models

This work has the main aim to design a ML model that can classify professional-course-pursuing students based on the severity levels of mental health in terms of anxiety and depression due to continuing effect of COVID-19 in 2022. As seen from the statistical analysis, it is understood that medical students are affected marginally higher than engineering college students. Now it is important to find out the factors that influence the anxiety and depression in these set of students.

5.1 Support vector machines

Cortes, Vapnik and Boser developed supervised learning algorithm to support linear as well as non-linear data known as SVMs (Brereton & Lloyd, 2010). SVM utilises generalisation control technique to rule out the curse of dimensionality issues. As stated, earlier SVM is famously used as classifier and regressor. It uses the concept of hyperplane to classify any two instances. The optimal hyperplane resulted by maximising the distance between the support vectors is explained in Figure 3.

5.2 J48 decision tree

J48 DT other called as C4.5 is developed by Quinlan (1992). It is a supervised ML tool extended from ID3 algorithm. J48 typically follows the tree structure, where the interior nodes detail the attributes, and the branching nodes provide information about the possible values a node can have. The leave nodes detail the final step of classification (Bienvenido-Huertas, Nieto-Julian, Moyano, Macías-Bernal, & Castro, 2020). A sample DT is shown below in Figure 4.

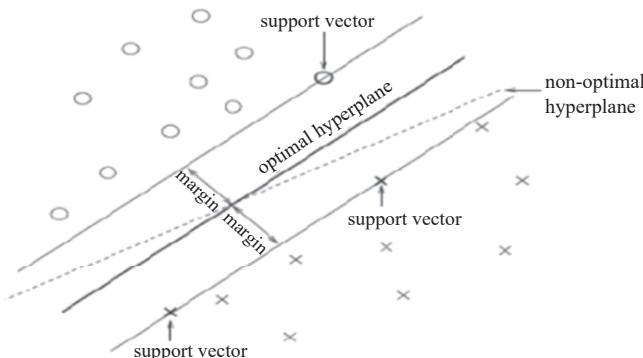


Figure 3.
Optimal hyperplane

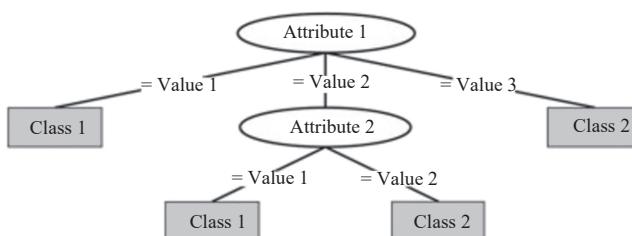


Figure 4.
Sample DT

5.3 Description of data set

Initially, the collected responses from the students were thoroughly checked for any errors that could be expected in any dataset such as missing values, out of range values, duplicate records etc. The data set is cleansed and processed. As per the standard rule, 70% of the dataset is used as training set and remaining is kept as 30% testing set. The ML models are trained using this 70% data. The accuracy of the model is then evaluated using the test set.

The model is refined further based on how well it performs compared to the test set. The performance parameters determine the model's effectiveness in terms of predictability (predicting the levels of anxiety and depression) and interpretability (interpreting the significance of the significant factors that cause the severity levels in a student's mental health).

5.4 Evaluation measures

The main objective of applying ML models is to classify "Early" the engineering and medical students with respect to the severity levels of anxiety and depression, so that corrective measures can be taken using SVM and DT and, furthermore, to identify the best model out of this analysis.

For this purpose, confusion matrixes are used to compare the evaluation metrics such as precision, recall, F-score and accuracy. The confusion matrix for analysing the severity levels in anxiety and depression is standardly defined in [Table 4](#).

Similarly, to predict the corrective measures, the following confusion matrix given in [Table 5](#) can be used.

The definitions for the metrics are tabulated in [Table 6](#) as follows.

5.5 Experimental setup

The ML algorithms were coded in python in collaboration with Google (from sklearn.svm import SVC). Using the kernel trick that separates non-linear separable problems to linear separable problems, SVM is implemented to be more accurate classifier. The kernel function is given as follows:

	In-patient treatment	Out-patient treatment and counselling	Continuous counselling and monitoring	Good
In-Patient treatment	True	False	False	False
Out-patient treatment and counselling	False	True	False	False
Continuous Counselling and monitoring	False	False	True	False
Good	False	False	False	True

Table 4.
Confusion matrix for anxiety and depression

	Normal	Borderline	Abnormal
Normal	True	False	False
Borderline	False	True	False
Abnormal	False	False	True

Table 5.
Confusion matrix for treatment recommendation

$$K(\bar{x}) = \begin{cases} 1 & \text{if } \|\bar{x}\| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

It is important to pass the values of kernel and cost[c] along with other parameters when SVM is constructed using Python. By default, kernel parameter uses radial basis function as its value. However, other kernel functions such as Linear, Poly Kernel and Sigmoid are also used in this work. Other parameters are assumed to be default. The value of cost can also be modified for tuning purposes. To gain optimal accuracy, the kernel function that achieved highest accuracy was then tried for cost range from 1 to 15. However, the implement of J48 DT (from `sklearn.tree import DecisionTreeClassifier`) is done with only the confidence factor hyperparameter being altered within a range from 0.15 to 0.90. The confidence factor in DT talks about the confidence intervals to estimate the gains associated with each split.

However, it is experimentally found that the kernel function, poly kernel, with cost function above 2 in SVM produced optimal results in multi-class classification as shown in [Figures 5 and 6](#).

Now, the DT is trained with the given data set with various confidence factors ranging from 0.5 to 0.90. It is observed that the confidence interval when set at 0.15 DT can perform well as shown in [Figure 7](#).

6. Results

The study concentrates on 120 medical students and 152 engineering college students residing in southern part of India. There were no missing data, as every question in the survey was marked as required. The sampling population's age is 21.74 with 4.2 as standard deviation value. Male students responded in comparison with the female students, and they were 63.97% in the entire sample size.

Precision	Fraction of the correct decisions to the total number of the given decisions in a particular class	$\frac{TP}{TP+FP}$
Recall	Fraction of the correct decisions that are given by the machine learning method to the total number of cases in a particular subset	$\frac{TP}{TP+FN}$
Accuracy	Measures how close the obtained decisions are to the actual classification	$\frac{TP+TN}{TP+TN+FP+FN}$
F1-score	Harmonic mean of precision and recall	$\frac{2 * Precision * Recall}{Precision + Recall}$

Note(s): TP – True Positive, FP- False Positive, TN – True Negative and
FN – False Negative

Table 6.
Evaluation metrics

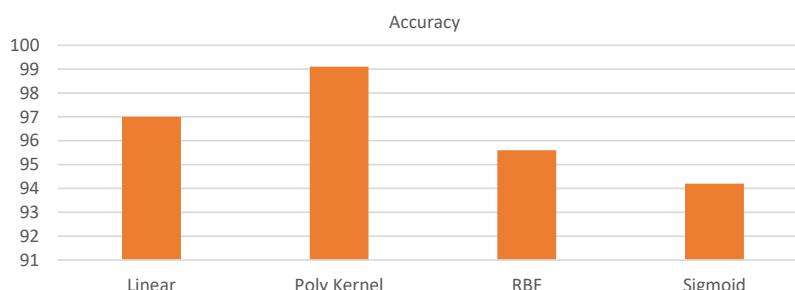


Figure 5.
Kernel tuning
parameters for SVM

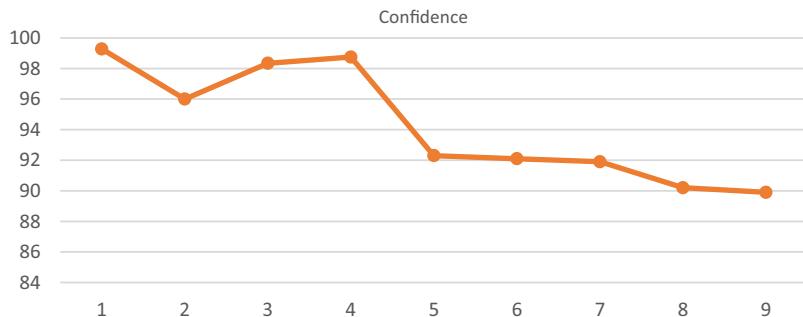


Figure 6.
Cost Tuning
parameters for SVM

6.1 Feature selection using random forest

The collected responses were split with respect to engineering- and medical-course-pursuing students. These data are further processed to find out the insights and cross-validated by applying stratified cross-validation techniques. The target variables – severity levels of anxiety and depression among the students – were calculated as per HADS score. Then the target parameters are set as “Abnormal, Borderline and Normal” according the HADS scale given in [Table 7](#).

From these target variables, it is possible to recommend the correction actions as given in the following [Table 8](#).

Thereafter, the most significant features are selected and the classification techniques as discussed above, SVM and DT, are applied to train the data set and derive into the suitable best model.

6.1.1 Anxiety and depression significant factors. Random forest algorithm as classifier is used to identify the feature set from the entire data set and it performs with 97.86% accuracy. [Figures 8 and 9](#) show the significant factors that impact the mental health of the students.

Similarly, the most significant factors that contribute to depression according to their impact is ranked and given in [Figures 10 and 11](#).

Similar to the anxiety assessment, the accuracy of the random forest classifier used to measure depression in the testing set is 97.86%. The analysis reveals that age is the most important factor, followed by the number of family members and the year of study.

6.1.2 Finding the students with highest risk factors. Certain students may be at high danger risk of having both anxiety and depression in abnormal levels. Those students need to have much more attention and immediate medical intervention. Hence it is needed to identify the significant contributors for identifying the students with highest risk factors. Following the

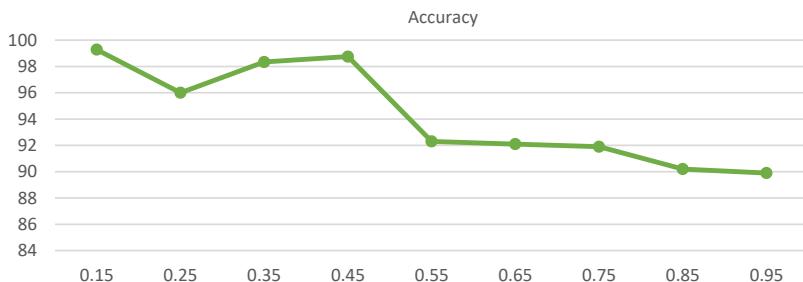


Figure 7.
Confidence level for DT

same approach used in assessing anxiety and depression, the combined effect of anxiety and depression is also performed. The top 3 significant factors are given in [Figure 12](#).

In summary, the random forest classification takes various independent variables into account and tries to draw relationships between independent variables and a dependent variable with 99.87 % accuracy.

6.2 Construction of the classification model

The features selected through random forest classifiers are fed to SVM and DT, which are tuned as mentioned earlier for further classification with respect to the course of study. This is done to classify the students early based on the severity levels. The following table showcase the performance of these two classifiers for the given dataset.

As given in [Table 9](#), SVM is showing a promising performance with accuracy for classification as 100% compared to J48 DT.

7. Discussion

This study aimed to build a classification model that could analyse the stress pointers amongst medical and engineering students in southern India. This study architected a ML model for early identification of mental health-affected students. The random forest technique is used to identify the significant features that contribute majorly to anxiety and depression. The SVM and J48 DT take the inputs based on the output from random forest to predict the severity levels of mental health. The analysis was done separately for engineering and medical students. The comparative analysis between two streams of students is not majorly addressed till now. Therefore, this study is conducted to identify the impact by considering the students from engineering and medical domain as case study.

The medical students are found to be much more abnormal compared to engineering students due to their workplace, study nature, etc. These students who are pursuing final year are put to treat the COVID-19 patients, and this created much panic amongst the students as well as their parents. This is found to be similar in the results found in [Collantoni](#)

Independent variables	Scale	Target variables	
		Anxiety	Depression
Demographic details	0-7	Normal	Normal
Educational details	8-10	Borderline	Borderline
Personal health details	11-21	Abnormal	Abnormal

Table 7.
Target variables

Anxiety	Depression	Corrective treatment
Normal	Normal	Good – no treatment
	Borderline	Continuous counselling and monitoring
	Abnormal	In-patient treatment
Borderline	Normal	Continuous counselling and monitoring
	Borderline	Out-patient treatment and counselling
	Abnormal	In-patient treatment
Abnormal	Normal	In-patient treatment
	Borderline	In-patient treatment
	Abnormal	In-patient treatment

Table 8.
Corrective treatments

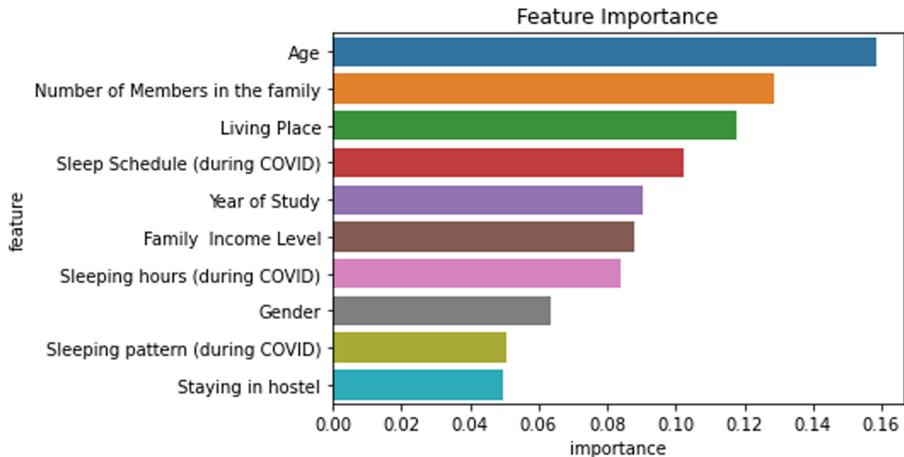


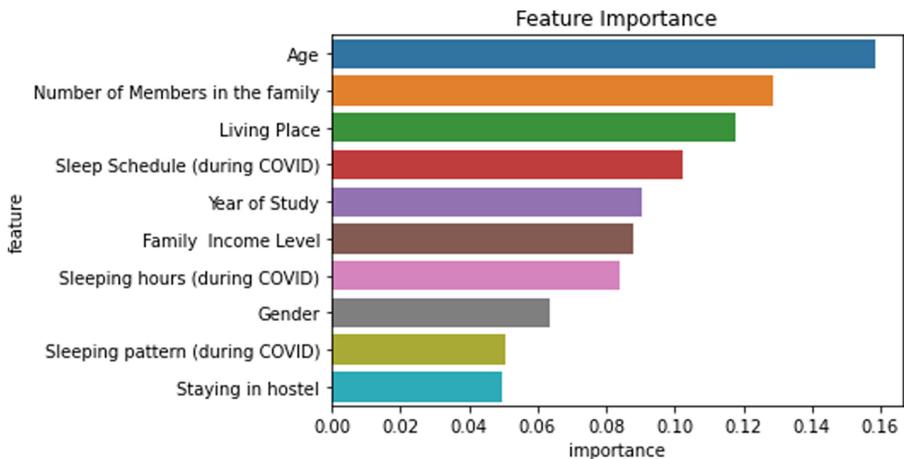
Figure 8.
Significant features for anxiety

	feature	importance
0	Age	0.134763
6	Number of Members in the family	0.128902
4	Year of Study	0.106169
11	Sleep Schedule (during COVID)	0.097454
7	Family Income Level	0.096405
9	Sleeping hours (during COVID)	0.091038
8	Living Place	0.086678
10	Sleeping pattern (during COVID)	0.075196
1	Gender	0.053458
5	Staying in hostel	0.049328
3	Level of Study	0.043573
2	Study Field	0.037036

Figure 9.
Importance of pivotal features

et al. (2021) and *Luca et al.* (2023). Meanwhile, engineering students fail to cope up the practical learning through mobile and their family, location, etc. played a huge impact in their studies pattern.

In this study, the effect of other factors such as demographic, family and sleeping patterns are considered to assess the mental health level of the students. HADS brought the outcome that any person scoring above 7 is having anxiety or depression. However, ML tools helped to study the reasons for anxiety and depression, so that corrective measures can be proactively managed for the well-being of the students. Thus, this work can complement the work of clinical psychologists to identify and treat students in effective manner.



ML model for mental health in college students

Figure 10. Significant features for depression

	feature	importance
0	Age	0.158560
6	Number of Members in the family	0.128585
8	Living Place	0.117692
11	Sleep Schedule (during COVID)	0.102165
4	Year of Study	0.090576
7	Family Income Level	0.087825
9	Sleeping hours (during COVID)	0.083776
1	Gender	0.063759
10	Sleeping pattern (during COVID)	0.050757
5	Staying in hostel	0.049650
2	Study Field	0.042023
3	Level of Study	0.024631

Figure 11. Importance of pivotal features

8. Limitations

The entire work is dependent on the social media-displayed online questionnaire, and the participants were not met in person. This indicates that the response rate could not be evaluated appropriately. Due to the medical restrictions imposed by COVID-19, which remain in effect in 2022, this is the only method found to collect primary data from college students. Additionally, students self-selected themselves to participate in this survey, which raises the possibility of selection bias. Also, to be noted is that the HADS survey is designed to assess anxiety and depression in general. This study gives a try to use it for assessing anxiety and depression due to COVID-19 in particular. Though the results are encouraging, it should be validated properly.

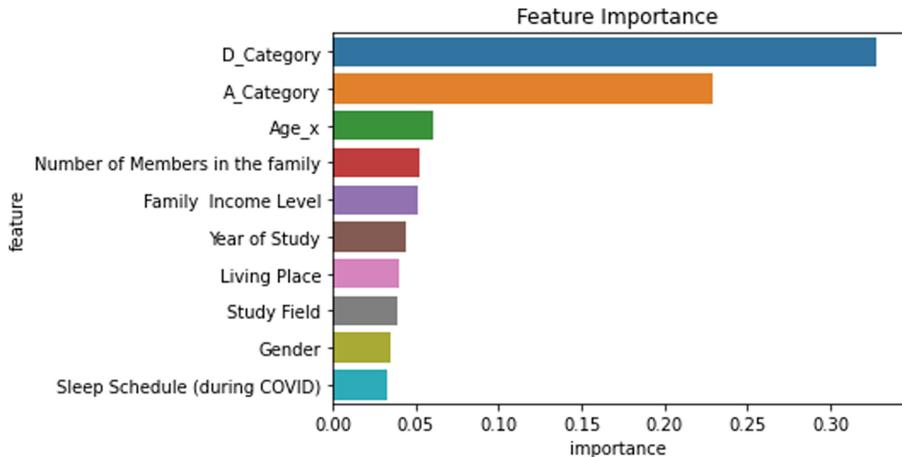


Figure 12.
Significant features for anxiety and depression

Table 9.
Comparison of results

	Engineering college students		Medical college students	
Evaluation metrics	SVM	J48 DT	SVM	J48 DT
Accuracy (%)	100	95.69	100	96.82
Precision	1	0.963	1	0.972
Recall	1	0.975	1	0.965
f-measure	1	0.974	1	0.964

9. Practical implications

This research work that assesses anxiety and depression amongst students due to COVID-19 in 2022 has several practical implications. These include:

Early identification and intervention: The research findings can help to identify students who are at risk for anxiety and depression, so that they can receive early intervention. This could involve providing counselling, stress management techniques or other mental health supports.

Development of new mental health resources: New mental health resources for students that include online resources, self-help guides or training for faculty on how to identify and support students with anxiety and depression can be developed.

Changes to school policies and practices: This could include providing more flexibility for students who are struggling with anxiety or depression or creating a more supportive environment for students with mental health needs.

Public awareness: The research findings can also be used to raise public awareness about the issue of anxiety and depression in students. This could help to reduce stigma and encourage students to seek help if they are struggling.

By identifying and intervening early, developing new resources and making changes to school policies and practices, we can help to ensure that all students have the support they need to thrive. In summary, the need for more mental health support for students, the importance of social connection and the need for flexibility in learning places helps to improve the mental health of their students and create a more supportive learning environment for all.

10. Conclusion

In summary, this work is the continuous work to distinguish the critical elements adding to mental health issues in higher education institutions even in 2022 because of COVID-19. Results show that students in the 20- to 24-year-old engineering and medicine fields experience abnormal levels of anxiety and depression. Anxiety and depression are significantly influenced by factors such as age, family size and location. When evaluating the combined effects of anxiety and depression on a student's mental health, the individual anxiety and depression scores have a significant impact. Students' mental health can be improved through policies, intervention procedures and awareness exercises based on this study's findings.

References

- Al-Mawee, W., Kwayu, K.M. and Gharaibeh, T. (2021). Student's perspective on distance learning during COVID-19 pandemic: A case study of western Michigan university, united. *International Journal of Educational Research Open*, 2, 100080 doi: [10.1016/j.ijedro.2021.100080](https://doi.org/10.1016/j.ijedro.2021.100080).
- Alghamdi, M., Al-Mallah, M., Ketyian, S., Brawner, C., Ehrman, J., & Sakr, S. (2017). Predicting diabetes mellitus using SMOTE and ensemble machine learning approach: The Henry Ford Exercise Testing (FIT) project. *PLoS ONE*, 12(7). doi: [10.1371/journal.pone.0179805](https://doi.org/10.1371/journal.pone.0179805).
- Alharthi, H. (2020). Predicting the level of generalized anxiety disorder of the coronavirus pandemic among college age students using artificial intelligence technology. In *19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES)*. IEEE.
- Alimoradi, Z., Ohayon, M. M., Griffiths, M. D., Lin, C. -Y., & Pakpour, A. H. (2022). Fear of COVID-19 and its association with mental health-related factors: Systematic review and meta-analysis. *BJPsych Open*, 8(2). doi: [10.1192/bjo.2022.26](https://doi.org/10.1192/bjo.2022.26).
- Atchison, C., Bowman, L. R., Vrinten, C., Redd, R., Pristerà, P., Eaton, J., & Ward, H. (2021). Early perceptions and behavioural responses during the COVID-19 pandemic: A cross-sectional survey of UK adults. *BMJ Open*, 11(1). doi: [10.1136/bmjopen-2020-043577](https://doi.org/10.1136/bmjopen-2020-043577).
- Bienvenido-Huertas, D., Nieto-Julián, J. E., Moyano, J. J., Macías-Bernal, J. M., & Castro, J. (2020). Implementing artificial intelligence in H-BIM using the J48 algorithm to manage historic buildings. *International Journal of Architectural Heritage*, 20(8), 1148–1160. doi: [10.1080/15583058.2019.1589602](https://doi.org/10.1080/15583058.2019.1589602).
- Bourmistrova, N. W., Solomon, T., Braude, P., Strawbridge, R., & Carter, B. (2022). Long-term effects of COVID-19 on mental health: A systematic review. *Elsevier Public Health Emergency Collection*, 299, 118–125. doi: [10.1016/j.jad.2021.11.031](https://doi.org/10.1016/j.jad.2021.11.031).
- Brereton, R. G., & Lloyd, G. L. (2010). Support vector machines for classification and regression. *Analyst*, 135(1), 230–267. doi: [10.1039/B918972F](https://doi.org/10.1039/B918972F).
- Cahyadi, A. (2021). Anxiety barriers in joining digital online learning during COVID19 pandemic outbreaks. *El-Buhuth: Borneo Journal of Islamic Studies*, 1–12.
- Chenyang Lin, Y. T. (2022). Prevalence and correlates of depression and anxiety among Chinese international students in us colleges during the COVID-19 pandemic: A cross-sectional study. *PLOS One*, 17(4), e0267081. doi: [10.1371/journal.pone.0267081](https://doi.org/10.1371/journal.pone.0267081).
- Chu, Y.-H., & Li, Y.-C. (2022). The impact of online learning on physical and mental health in university students during the COVID-19 pandemic. *International Journal of Environmental Research and Public Health*, 19(5).
- Ciro, C., Mariagrazia, D. G., Mario, M., Rebecca, C., Angelo, G., & Graziella, O. (2020). Mindfulness, age and gender as protective factors against psychological distress during COVID-19 pandemic. *Frontiers in Psychology*, 11. doi: [10.3389/fpsyg.2020.01900](https://doi.org/10.3389/fpsyg.2020.01900).
- Collantoni, E., Saieva, A. M., Meregalli, V., Girotto, C., Carretta, G., Boemo, D. G., ... Favaro, A. (2021). Psychological distress, fear of COVID-19, and resilient coping abilities among healthcare workers in a tertiary first-line hospital during the coronavirus pandemic. *Journal of Clinical Medicine*, 10(7). doi: [10.3390/jcm10071465](https://doi.org/10.3390/jcm10071465).

- Donna Wang, K. K. -S. (2021). Micro, mezzo, and macro factors associated with coping in the early phase of COVID-19. *Journal of Human Behavior in the Social Environment*, 31, 60–69. doi: [10.1080/10911359.2020.1838985](https://doi.org/10.1080/10911359.2020.1838985).
- Flesia, L., Monaro, M., Mazza, C., Fietta, V., Colicino, E. . . ., & 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). doi: [10.3390/jcm9103350](https://doi.org/10.3390/jcm9103350).
- Gonzalez-Ramirez, J. M. (2021). Emergency online learning: College students' perceptions during the COVID-19 pandemic. *College Student Journal*, 55(1), 29–46.
- Herbert Cornelia, E. B. (2021). How do you feel during the COVID-19 pandemic? A survey using psychological and linguistic self-report measures, and machine learning to investigate mental health, subjective experience, personality, behaviour during the COVID-19 pandemic among univers stu. *BMC Psychology*, 9(1), 1–23.
- Houben-Wilke, S., Goertz, Y. M., Delbressine, J. M., Vaes, A. W., Meys, R., Machado, F. V., . . . & Janssen, D. J. (2022). The impact of long COVID-19 on mental health: Observational 6-month follow-up study. *JMIR Mental Health*.
- Jena, P. K. (2020). Impact of Covid-19 on higher education in India. *International Journal of Advanced Education and Research*, 5(3), 77–81.
- Kekojevic, A., Basch, C. H., Sullivan, M., & Davi, N. K. (2020). The impact of the COVID-19 epidemic on mental health of undergraduate students in New Jersey, cross-sectional study. *PLoS ONE*, 15(9). doi: [10.1371/journal.pone.0239696](https://doi.org/10.1371/journal.pone.0239696).
- Khattar, A. P. (2020). Effects of the disastrous pandemic COVID 19 on learning styles, activities and mental health of young Indian students-a machine learning approach. In *4th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE.
- López-Valenciano, Alejandr, Suárez-Iglesias, D., Sanchez-Lastra, M. A., & Ayán, C. (2021). Impact of COVID-19 pandemic on university students' physical activity levels: An early systematic review. *Frontiers in Psychology*, 11, 624567. doi: [10.3389/fpsyg.2020.624567](https://doi.org/10.3389/fpsyg.2020.624567).
- Luca, F., Adeeb, M., Waseem, A., Helmy, M., & Monaro, M. (2023). Psychological distress related to the COVID-19 pandemic: The protective role of hope. *European Journal of Investigation in Health, Psychology and Education*, 13(1), 67–80. doi: [10.3390/ejihpe1301005](https://doi.org/10.3390/ejihpe1301005).
- Malta, G. D., Julian Bond, D. C., Smith, K., & Naomi (2022). Distance education students' mental health, connectedness and academic performance during COVID-19: A mixed-methods study. *Distance Education*, 43(1), 97–118.
- Mazza, M. G., Lorenzo, R. D., Conte, C., Poletti, S., Vai, B., Bollettini, I., . . . Ben, F. (2020). Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors. *Brain, Behavior, and Immunity*, 89, 594–600. doi: [10.1016/j.bbi.2020.07.037](https://doi.org/10.1016/j.bbi.2020.07.037).
- Pramukti, I., Strong, C., Sitthimongkol, Y., Setiawan, A., Pandin, M. G., Yen, C.-F., . . . Ko, N.-Y. (2020). Anxiety and suicidal thoughts during the COVID-19 pandemic: Cross-country comparative study among Indonesian, Taiwanese, and Thai University Students. *Journal of Medical Internet Research*, 22(12). doi: [10.2196/24487](https://doi.org/10.2196/24487).
- Priya, A., Garg, S., & Tiggia, N. P. (2020). Predicting anxiety, depression and stress in modern life using machine learning algorithms. *Procedia Computer Science*, 167, 1258–1267. doi: [10.1016/j.procs.2020.03.442](https://doi.org/10.1016/j.procs.2020.03.442).
- Quinlan, J. R. (1992). C4.5: Programs for machine learning.
- Ren, Z. X.Y. (2021). Psychological impact of COVID-19 on college students after school reopening: A cross-sectional study based on machine learning. *Frontiers in Psychology*. doi: [10.3389/fpsyg.2021.641806](https://doi.org/10.3389/fpsyg.2021.641806).
- Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among the seafarers using machine learning technology. *Informatics in Medicine Unlocked*, 19. doi: [10.1016/j.imu.2018.12.004](https://doi.org/10.1016/j.imu.2018.12.004).
- Shader, R. I. (2020). COVID-19 and depression. *Clinical Therapeutics*, 42, 962–963.

- Shivangi, D. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, 49(1), 5–22. doi: [10.1177/0047239520934018](https://doi.org/10.1177/0047239520934018).
- Shyadligeri, A. A., Vaz, F. S., & Lokapure, S. R. (2022). Psychological distress among first-year medical students amidst COVID-19-related uncertainty at a medical college in western India: A cross-sectional study. *Education Health*, 35, 20–25.
- Snaith, R. P. (2003). The hospital anxiety and depression scale. Health qual life outcomes. *Health and Quality of Life Outcomes*, 1. doi: [10.1186/1477-7525-1-29](https://doi.org/10.1186/1477-7525-1-29).
- SonSudeep, C., SmithXiaomei, H., & Sasangohar, W. (2020). Effects of COVID-19 on college students' mental health in the United States: Interview survey study. *Journal of Medical Internet Research*, 22(9), e21279. doi: [10.2196/21279](https://doi.org/10.2196/21279).
- Tróglolo, M., Moretti, L., & Medrano, L. (2022). A nationwide cross-sectional study of workers' mental health during the COVID-19 pandemic: Impact of changes in working conditions, financial hardships, psychological detachment from work and work-family interface. *BMC Psychology Volume*, 73(10).
- Wang, X. H.S. (2020). Investigating mental health of US college students during the COVID-19 pandemic: Cross-sectional survey study. *Journal of Medical Internet Research*, 22(9), e22817. doi: [10.2196/22817](https://doi.org/10.2196/22817).
- World Health Organization (2022). WHO Coronavirus (COVID-19) Dashboard, Available from: <https://covid19.who.int/>
- Xiao, H., Zhang, Y., Kong, D., Li, S., & Yang, N. (2020). The efects of social support on sleep quality of medical staf treating patients with coronavirus disease 2019 (COVID-19) in january and february 2020 in China. *Medical Science Monitor*.
- Zeng, X., & Tingzeng, W. (2021). *International Journal of Multidisciplinary Perspectives in Higher Education*, 6(1), 182–195.
- Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67, 361–370.

Corresponding author

Reshmy Krishnan can be contacted at: reshmy@muscattcollege.edu.om