Machine learning approach for anxiety and sleep disorders analysis during COVID-19 lockdown

L. Jani Anbarasi[1] · Malathy Jawahar[2] · Vinayakumar Ravi[3] · Sherin Miriam Cherian[1] · S. Shreenidhi[1] · H. Sharen[1]

Abstract

The Severe Acute Respiratory Syndrome (SARS)-CoV-2 virus caused COVID-19 pandemic has led to various kinds of anxiety and stress in different strata and sections of the society. The aim of this study is to analyse the sleeping and anxiety disorder for a wide distribution of people of different ages and from different strata of life. The study also seeks to investigate the different symptoms and grievances that people suffer from in connection with their sleep patterns and predict the possible relationships and factors in association with outcomes related to COVID-19 pandemic induced stress and issues. A total of 740 participants (51.3% male and 48.7% female) structured with 2 sections, first with general demographic information and second with more targeted questions for each demographic were surveyed. Pittsburgh Sleep Quality Index (PSQI) and General Anxiety Disorder assessment (GAD-7) standard scales were utilized to measure the stress, sleep disorders and anxiety. Experimental results showed positive correlation between PSQI and GAD-7 scores for the participants. After adjusting for age and gender, occupation does not have an effect on sleep quality (PSQI), but it does have an effect on anxiety (GAD-7). Student community in spite of less susceptible to COVID-19 infection found to be highly prone to psychopathy mental health disturbances during the COVID-19 pandemic. The study also highlights the connectivity between lower social status and mental health issues. Random Forest model for college students indicates clearly the stress induced factors as anxiety score, worry about inability to understand concepts taught online, involvement of parents, college hours, worrying about other work load and deadlines for the young students studying in Universities.

Keywords Cross-sectional study · PSQI · GAD-7 · COVID-19 · Insomnia · Anxiety · ANOVA

1 Introduction

The COVID-19 pandemic has taken the world by surprise and brought with it an acute amount of stress and anxiety in all sections of the society and amongst all ages. Social distancing, school and college closures, working from home, isolation, massive layoffs at the workplace, healthcare workers at constant risk and anxiety about testing positive has rendered many people with disrupted daily schedules, a lot of uncertainty about the future and skyrocketing levels of stress. All these can contribute to insomnia and other sleep problems. Clear communication, shift hours, limited rest areas, as well as broad access and detailed rules on the use and management of protective equipment and specialised

✉ Vinayakumar Ravi
vravi@pmu.edu.sa

L. Jani Anbarasi
Janianbarasi.l@vit.ac.in

Malathy Jawahar
malathyj@clri.res.in

Sherin Miriam Cherian
sherin.miriam2017@vitstudent.ac.in

S. Shreenidhi
s.shreenidhi2017@vitstudent.ac.in

H. Sharen
sharen.h2020@vitstudent.ac.in

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai 600 127, India

[2] Leather Process Technology Division, CSIR-Central Leather Research Institute, Adyar, Chennai 600020, India

[3] Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar, Saudi Arabia

training on handling COVID-19 patients, could alleviate anxiety caused by the perceived unfamiliarity and unpredictability of the hazards involved [1]. Previous studies have looked into the prevalence of psychological consequences in Health care workers during previous infectious disease outbreaks, as well as the factors that influence them [2, 3]. The Indian population was restricted in social isolation for nearly six months, with the exception of the ability to leave the house to obtain food or medicine. These restrictions have a substantial impact on social relationships and way of life. As a result of their fear of being infected, many people are experiencing increased anxiety. In the general population, there was a considerable increase in psychological discomfort and mental disorder symptoms [4]. Even if the restriction of freedom was necessary to reduce viral transmission, the COVID-19 epidemic may have had a significant psychological impact. Psychological discomfort and anxiety symptoms can be generated by traumatic events like the COVID-19 epidemic, which can have a poor impact on sleep quality [5].

Previous research has discovered that acute viral diseases, such as SARS, can produce anxiety, sadness, stress, and posttraumatic stress disorder in both survivors and non-infected individuals [6–8]. Previous research has shown that persons who experience unexpected occurrences have a negative influence on their mental health, which could lead to symptomatology similar to those of post-traumatic stress disorder [9]. As a result, during the COVID-19 emergency, researchers concentrated on physical and mental health, as well as sleep. Female gender, being a student, having suggestive COVID-19 symptoms, and a perceived low quality of life were all linked to a greater risk of anxiety and depression in recent Chinese studies on the general population [10–13]. Another study found that after the COVID-19 epidemic, 7% of Wuhan inhabitants, particularly women, experienced PTSD symptoms. Anxiety levels were higher in people younger than 35 years old who spent more than 3 h watching COVID-19 news [14].

Sleep and insomnia were highlighted in a recent study by the European CBT-I Academy on sleep problems during home confinement due to the COVID-19 epidemic [15]. Cellini et al. [16] investigated changes in sleep patterns, time, and use of electronic devices among 1310 young adults' workers and university students and discovered that digital media use is higher in the evening before bedtime. They also discovered that people went to bed later and awoke later, spent more time in bed with poor quality sleep. Sleep problems were more likely in people who were depressed, anxious, or stressed. According to Li et al. [17], the prevalence of insomnia increased significantly during the COVID-19 outbreak (including new onsets of insomnia in certain instances), time in bed (TIB) and total sleep time (TST) increased, and sleep efficiency fell significantly.

It is also critical to provide timely and suitably customised mental health help via hotline teams, the media, or multidisciplinary teams that include mental health professionals. International Labor Organization [14] reported that education sector has been strongly affected by the COVID-19 pandemic. COVID-19 pandemic has caused the deepest global economic recession after second world war [10]. From our analysis and its conclusions, we hope to find correlations between different kinds of people who suffer from the same sleep related symptoms or disorders, with the hope that such insight will lead to finding ways to mitigate the stress and anxiety so we can get through this phase unscathed. Individuals are made to fill questionnaires based on standard indexes like ISI, PSQI, SAS etc. The data has been pre-processed by removing null values, grouping people based on certain factors for further analysis. Statistical tests were like ANOVA, t-test, chi-square are performed to analyse causes of poor sleep. Correlation tests such as Pearson's and Spearman's co-efficient are computed to study the relationship among multiple factors. Logistic regression is used to the study causes when the dependent variable is categorical. Most of the existing work focuses on a subset of people like students or healthcare professionals and analyses their sleep as a whole. We want to ask people belonging to different professions questions that are relevant to them and study how sleep quality for a given profession can be improved. There are also not many Indian studies that focus on the effect of COVID-19 on sleep patterns and cover enough of the demographic.

COVID-19 pandemic had a huge impact globally in many ways [18]. Mental well-being of the people is highly affected due to the outbreak of coronavirus and the post symptoms after the attack. Managing good physical and psychological health are essential for proper functioning of human body and immune system to prevent from infections. COVID-19 quarantine practices has led to many psychological disturbances like stress, anxiety, sleep disturbance and mental depression. It is a well-known fact that the traumatic events damage the human psyche and its outcome affects the individual body and brain. However, investigating the factors that are responsible for anxiety and sleep disorder during the COVID-19 pandemic would lead to understand the cognitive mechanisms and to enhance the healthcare system as a whole. The impact of the current extraordinary crisis on the psychological well-being of students and on other persons with COVID-19 have not yet been determined.

During the COVID-19 epidemic, sleep disorders seem to have been frequent. Furthermore, sleep issues are consistently linked to higher levels of psychological distress. Individuals with COVID-19 infection seemed to have a higher incidence of sleep disorders, whereas healthcare workers had the second highest frequency. Sleep disorders were

found to have moderate relationships with psychological distress (particularly depression and anxiety). The goal of this study is to fill in the gaps in knowledge about the association between COVID-19 and mental health symptoms and sleep disorders, as well as to identify potential risk factors. As a result, nationwide data is collected with the voluntary and willing participation of healthy and COVID-19 affected individuals in order to conduct a study relating to sleep and anxiety issues. The results will be helpful in identifying distressed patients who are at risk of developing psychiatric symptoms and formulating prevention efforts to help these people avoid poor mental health. The main goal of this study is to find the significant association between the symptoms such as sleep quality, anxiety, depression, and insomnia with the COVID-19 pandemic lockdown.

2 Related work

With outbreak of COVID-19 pandemic, a lots of surveys have been conducted [10, 11, 14, 19–22] to analyse how it has affected people's mental health, sleep quality etc. The preventive measures practised during pandemic has led to worsening of mental well-being [11, 19–22]. These studies make use of standard questionnaires such as PSQI, GAD7, SSRS, SARS etc. and their variants. Statistical tests such as chi-square tests, t-tests, ANOVA was used to study the effect of age, gender etc. on sleep quality, distress [14, 20]. Machine learning models such as logistic regression, multivariate analysis [10, 11, 14, 19] are used to study the effects of various parameters on sleep quality, distress etc. ANOVA tests are done to find out whether a parameter (categorical) has any significant effect on a dependent variable. A study has been conducted using ANOVA to reveal the effect of a student's position in his classroom on his performance [23].

ANOVA assumes that data points are drawn from normal distribution [23, 24]. To check this Shapiro-Wilks test were performed [24]. A non-parametric alternative to ANOVA is Kruskal–Wallis test [25]. Levene's test can be used to check for homogeneity of variance [26, 27]. Random forest was used to study the effects of various parameters (categorical and numerical) on a dependent variable [28, 29]. During training phase based the cuts made (reduction in entropy) each parameter gets an importance score. Random forest feature importance can be biased towards high-cardinality categorical attributes or numerical attributes, hence permutation importance that is based on random forest can be calculated for more reliable results [30]. K-means requires k to be known beforehand, elbow method can be used to estimate this [31, 32].

An anonymous online survey of 1,200 adolescents in the Chinese regions of Hunan and Guangxi from February to April 2020 were performed where a total of 150 middle school students with anxiety levels greater than 50 was chosen to engage in the intervention experiment with 75 students in each group. Peer education based on the Model 328 is provided to these students to help adolescents feel less anxious and depressed while also improving their sleep quality [33]. An online questionnaire poll was done utilising several scales to investigate Wuhan citizens' psychological reactions to the COVID-19 epidemic. 1242 residents were involved in this study. The relationship between demographic characteristics and anxiety, depression, sleep disturbance, and passive coping style was studied using multivariate logistic regression models. Wuhan inhabitants' psychological well-being and sleep quality were both lower than before the COVID-19 outbreak, although the rate of passive coping with stress was higher [34]. A web-based survey that includes sleep, anxiety, and depression symptoms surveys was performed with 400 people where 307 students and 93 were university.

Between before and during the COVID-19 [35] emergency, an increase in Bed Time hour, Sleep Latency, and Wake-Up time was observed to be associated with a worsening of sleep quality and insomnia symptoms. The impact of the delay in Bed Time and Wake-Up was most noticeable in students during the lockdown. Employees with insomnia had a prevalence of 24% prior to COVID-19, which grew to 40% during COVID-19, but workers with sleep initiation issues had a prevalence of only 15% prior to COVID-19, which increased to 42% during COVID-19. Depressive symptoms were found in 27.8% of the sample, whereas anxious symptoms were present in 34.3 percent, especially among students and administrative staff. Students were more affected by lockdown than workers, and females were more affected than males. In terms of psycho-emotional factors, nearly a third of our participants displayed depressive or anxious symptoms. These findings could help to encourage the adoption of various well-being interventions in pandemic situations.

To analyse the COVID-19 outbreak, an online survey was employed to do a fast assessment, with 11,835 adolescents and young adults [36]. Sleep disorders in adolescents and young adults were found to be negatively related to COVID-19 trend estimates. Social support and sleeplessness symptoms were both mediated by anxiety and depression symptoms. During the COVID-19 epidemic, this study discovered a high frequency of sleep disorders among adolescents and young people, particularly senior high school and college students, which were negatively connected with students' projection of trends. Two questions in four repeating cross-sectional surveys were performed to investigate sleep disorders in an epidemiological survey on Covid and Confinement (COCONEL). The prevalence of sleep disorders dropped dramatically in the final weeks of confinement, and this pattern was confirmed one month after confinement ended. The likelihood of regaining a decent night's sleep is highly

dependent on the type of sleep disturbance. During confinement, persons with modest sleep disorders saw a reduction in their sleep problems [37].

The effects of the pandemic on shift workers, health care employees, students, and COVID-19 patients are studied by Salehinejad et al. [38]. The impact of sleep problems and poor sleep quality during the pandemic, as well as their contributory variables, rate and prevalence, on both the general population and COVID-19 patients, are examined next. Circadian and sleep issues have a significant impact on home confinement and its physiological, circadian, and psychological consequences. Behavioural therapies that attempt to normalise the variables that cause circadian and sleep problems are beneficial. Sleep disorder is also a rsisk factor for critical COVID-19 patients [39]. Long-term results of mental health therapies in community and hospital settings need to be studied further to see how we might reduce the detrimental consequences of virus epidemics in the general public, particularly among vulnerable patients with mental health issues [40].

3 Proposed work

The proposed work as shown in Fig. 1 is a fast systematic review and meta-analysis to find the significant association between the symptoms such as sleep quality, anxiety, depression, and insomnia with the COVID-19 pandemic lockdown.

3.1 Data source

The present study conducted the survey via Google forms. Authors sent the invitation to participate in the survey through varied means of platforms such as email, student offices, Moodle e-learning platform, and other social media. The current cross-sectional study was carried out for different strata and sections of the society during the COVID-19 pandemic.

The survey was circulated for a period of roughly 1 month between 30th August to 4th October, 2020, through emails, WhatsApp groups, Instagram, Facebook with a general push to increase the number of people it reached. A total of 704 participants sent their responses from all the states of India, a majority being from Tamil Nadu and Maharashtra (Fig. 2).

The google form survey was constructed with two sections:

a) Demographic information, PSQI and GAD-7 questionnaires.
b) Occupation-targeted questions.

Put together, that amounted to a total of 156 columns/attributes for each data point, with different rows having different sections filled up, based on their occupation. Section 1 was mandatory for everyone to fill. We conduct a cross-sectional survey and attained as much data as possible from different demographics of the society. For this purpose, we divide our survey into sections based on the occupation/current status of the person. This can vary from students in schools, colleges, working professionals, healthcare workers etc. For each of the categories, we ask targeted and specific
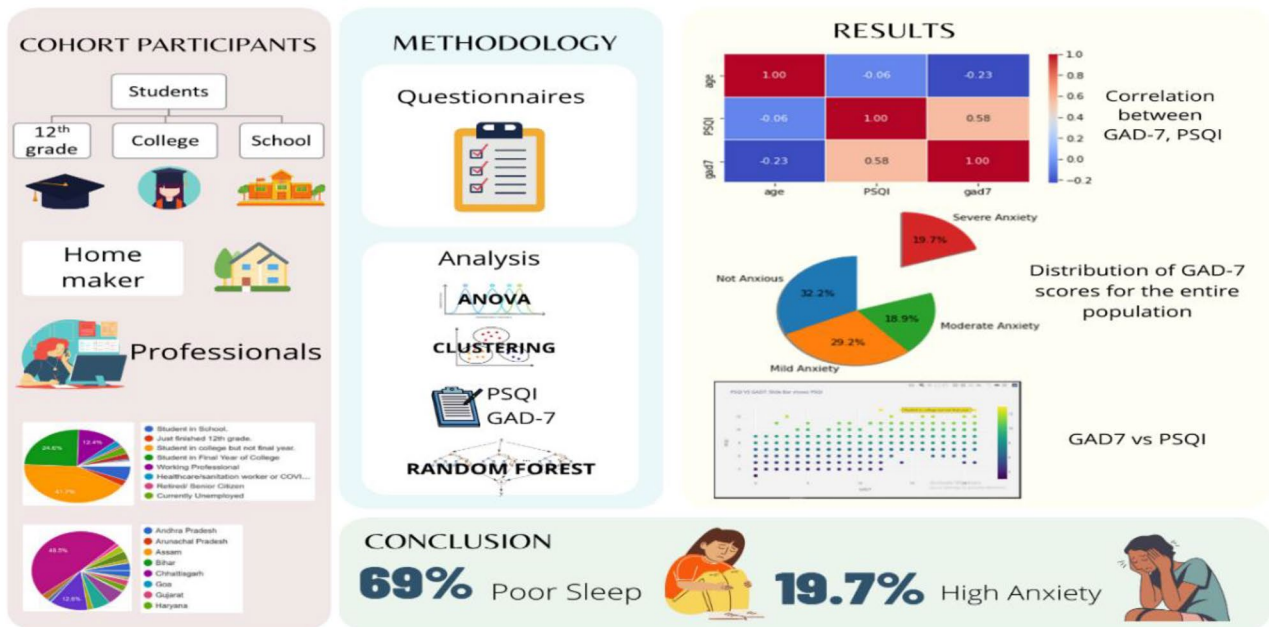

Fig.1 The schematic illustration of the factors associated with sleep, anxiety, and depression during COVID-19 lockdown
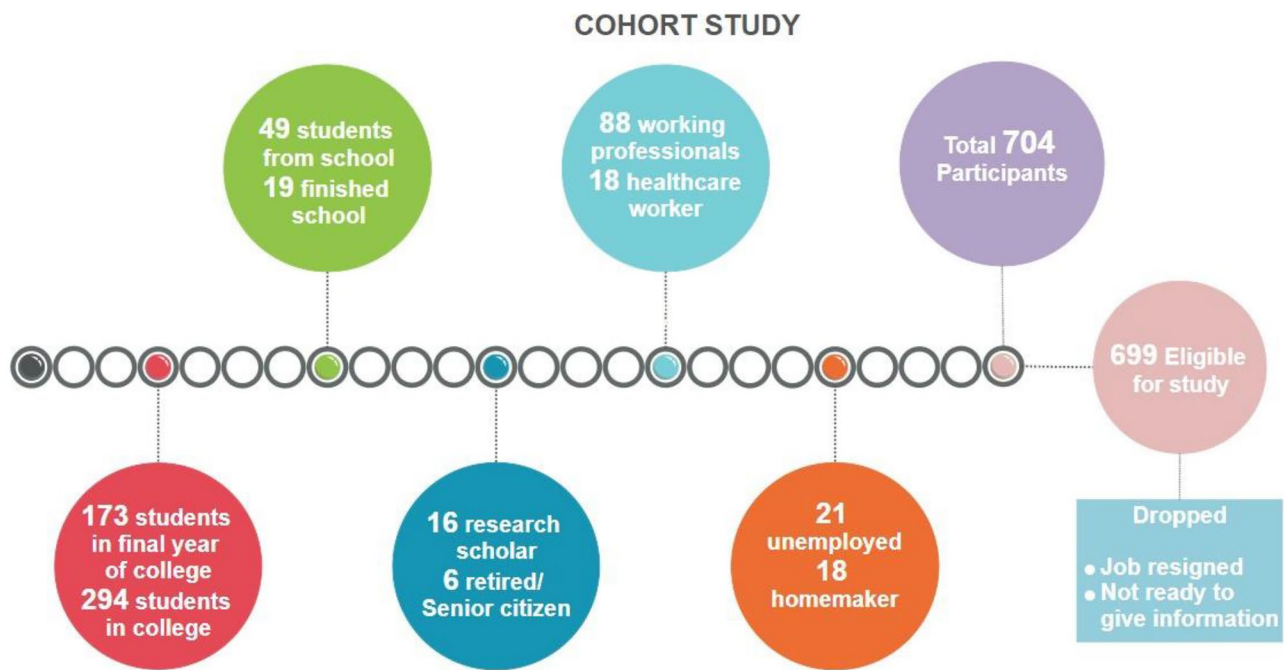
Fig. 2 Study Methodology

questions related to issues they might be facing because of the COVID-19 pandemic. Most of these issues have a direct impact on their stress/ general anxiety levels and can hence influence their sleep patterns. The link to the survey (on Google forms): http://bit.do/covidsleep.

The demographic groups in the survey included Age, Gender, Residence, Occupation/ Type: (a) Student in School (9–12) (b) Working Professional (c) Student in Final Year of College (d) Healthcare Worker (e) Currently Unemployed (f) Homemaker (g) Retired (h) Just Finished 12th Grade and (i) Student in College (Not Final Year). The state-wise distribution of the form data is given in Fig. 3(a) Distribution of collected data by occupations/ categories of individuals are given in Fig. 3(b) :

The data includes 49 from Student in School, 88 Working Professional, 173 Student in Final Year of College, 18 Healthcare Worker, 21 Currently Unemployed, 18 Homemaker, 6 Retired / Senior citizens, 19 Just Finished 12th Grade and 294 Student in College (Not Final Year) and 16 research scholar.

A majority of the people who filled the form belonged to students in college, either in their final year or otherwise. It can also be observed that 48% of the data is from Tamil Nadu. This convenience bias was inevitable because of the nature by which the form was circulated, largely amongst the student population and a starting point. The data still covers nearly all states as well as all the intended categories of occupation, in different numbers. It is also quite evenly distributed as far as gender is concerned (Fig. 4), although

we can begin to notice a few anomalies in the collected data, highlighting the importance of pre-processing.

3.2 Data pre-processing

The following steps were undertaken for the pre-processing of the raw form data:

(a) Changing column names

Since all the questions were long and detailed in nature, the names of the columns were replaced with shorter versions for better code readability and ease of analysis. Timestamp was dropped. GAD-7 and PSQI scores were calculated for each row of data, and appended to the data as two new columns in the end. Occupations were grouped together wherever there were overlaps, especially in the "Other" option of the form. Dubious sounding occupations were dropped.

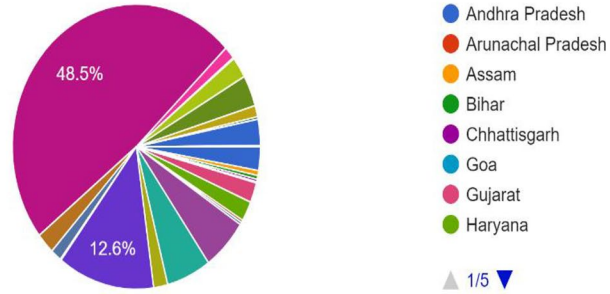(b) Occupation types were changed to shorter names

Unusual data/ outliers can greatly impact the statistical results of a survey-based study, hence we attempted to identify these and correct them. Most of these occur and can be identified by examining the questions that are typed in by the survey-takers rather than choosing an option on the form.

(c) Removal of rows that contain no data

Checking the minimum and max range of each occupation type to look for unusual data/ outliers was per-

Fig. 3 **a** State-wise data distribution. **b** Distribution of data by occupations/ categories of individuals

Current State Of Residence:
707 responses

- Andhra Pradesh
- Arunachal Pradesh
- Assam
- Bihar
- Chhattisgarh
- Goa
- Gujarat
- Haryana

▲ 1/5 ▼

(a) State-wise data distribution

Pick your occupation:
707 responses

- Student in School.
- Just finished 12th grade.
- Student in college but not final year.
- Student in Final Year of College
- Working Professional
- Healthcare/sanitation worker or COVI…
- Retired/ Senior Citizen
- Currently Unemployed

▲ 1/4 ▼

(b) Distribution of data by occupations/ categories of individuals

formed. Outliers were either dropped or imputed (mean imputation for age in some cases) based on how legitimate the rest of their answers were. On observing the "hours of online classes" question for school students, final year students and other college students, it was observed that the range of hours did not always fit practical/ logical limits. In the case of school students, it can be assumed that the students mistook hours/ week for hours/ day while filling out the form, when their hours are < 10/ week. These rows were identified and multiplied by 5 for a rough estimate of the correct value of hours/ week. It was observed that a single student had reported 65 h/ week in school amounting to roughly 11 h of school a day, even if the school worked on Saturdays. This student also reported to be attending online support (Byjus, Vendata), making it possible that the student counted

Fig. 4 Distribution of Gender wise data

Gender
707 responses

- Female
- Male
- Prefer not to say
- Apache helicopter
- Asexual
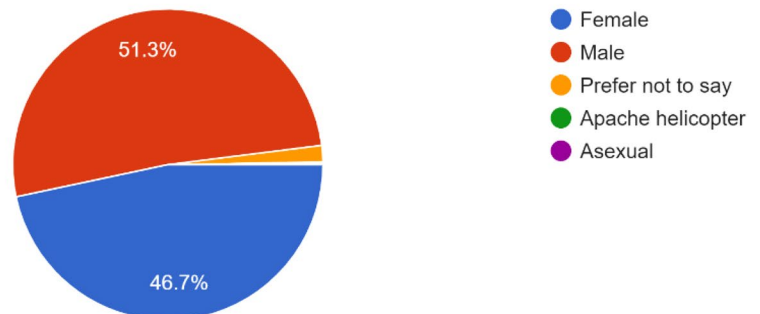
those hours into the question. With just 2.5 h of sleep every day, a PSQI of 10, GAD of 21, this is an example of someone with dangerously high levels of anxiety and poor sleep quality, and who needs immediate attention. There were people in the final year who reported > 90 h of online classes a week, which is simply not possible. It is possible that they misunderstood the question, and hence their hours were changed to an appropriate value.

Since the number of responses in the college and final year student categories are large in number, the modifications to the outliers will not affect conclusions as much as they would if left as it is. Exactly one school student reported 0 h/ week and that school was not functioning, and no online classes were taking place. The Gender column was examined that dubious sounding responses were removed. The final categories after removal were "Male", "Female", and "Prefer to not say". Note that the option of "Other" was given in this question as well. It was observed that the 18 healthcare workers filled up the checkbox question (is your career at risk?) with only one option, as opposed to multiple. So it is noteworthy to observe that the 6 of them, who planned to resign from their job, do so with a clear sense of conviction. The healthcare type column (govt/private service) was removed because of insufficient data (this question was added after the form began circulation). After these steps, the status of data included 704 Original Data from which 5 were dropped leading to current actual data 699.

4 Simulation models

Generalized Anxiety Disorder Scale (GAD-7) [14] was used to measure the quality of sleep of people and classify if they have a sleep disorder. Pittsburgh Sleep Quality Index (PSQI) [10] was used to measure anxiety levels in individuals. Statistical and demographic wise inferences based on the GAD-7 and PSQI scores were investigated to understand the significant factors that affected the most set of people.

4.1 Random forest

Random forests an ensemble method was used for both classification and regression. In Random Forest, multiple trees can be grown as opposed to a single tree in the Decision Tree model. This helps Random Forest overcome the error due to bias which is a major drawback of Decision Tree. The forest takes the average of outputs by different trees as the final output in the case of regression and majority voting is followed in the case of classification.

Let $A = A_1 \dots A_n$ refers the training data and $B = B_1 \dots B_n$, refers the response where random samples of X times are selected from the training set to fit the tree from the

samples. For all the values of m from 1 to X train the tree T_x on the training and response data A_b, B_b the predictions are determined on the regression trees Y':

$$\hat{f} = \frac{1}{X} \sum_{x=1}^{X} f_x(Y')$$

Because it reduces the variance of the model without increasing the bias, this bootstrapping approach leads to greater model performance. This means that while a single tree's predictions are sensitive to noise in their training set, the average of several trees' forecasts are not, as long as the trees are not correlated. Additionally, the standard deviation of the predictions from all of the separate regression trees on x' can be used to evaluate the prediction's uncertainty.

$$\sigma = \sqrt{\frac{\sum_{x-1}^{X} (fx(A^1) - \dot{f})^2}{X - 1}}$$

Random forest was built to study the role the different attributes in causing sleep disorders in college students. Random forest is based on decision tree, decision use reduction in entropy, gini index etc. to form the decision surface. Hence random forest has a tendency to choose high-cardinality categorical variable or numerical variable as important. To overcome this shortcoming permutation feature importance was used, this is computationally expensive when compared to Random forest. The model was built based on the assumption that PSQI score > 7 signified the presence of sleep disorder.

4.2 K-means clustering

K-means clustering is the most widely used algorithm for clustering tasks. The algorithm aims to minimize the variance within a cluster while trying to maintain high variance among data points of different clusters. K-means clustering was used to cluster the participants based on their gad7 (anxiety) score and PSQI (sleep quality) score.

Centroids were initialized by shuffling the dataset and number of clusters (K) was picked at random for the centroids without replacing them. The process was iterated until the centroids do not change. i.e. the clustering of data points does not change. Calculate the total of all data points' squared distances from all centroids. The objective function used to perform as given in Eq. (1) where W_{ij} belongs to the cluster k and C_k refers to the centroid of the cluster.

$$E = \sum_{i=1}^{m} \sum_{j=1}^{Z} W_{ij} \left| \left| X^i - C_k \right| \right|^2 \tag{1}$$

The value of k needs to be specified. Elbow method was used to decide on the optimal value of k. Further hierarchical and affinity propagation clustering were used to see whether the results of K-means was consistent.
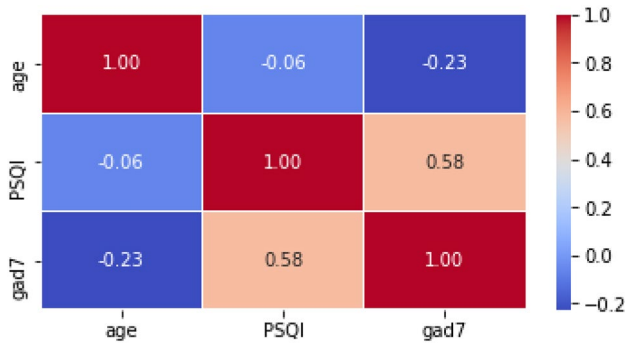
Fig. 5 Correlation between GAD-7, PSQI

Table 1 GAD-7 and PSQ1 Scores

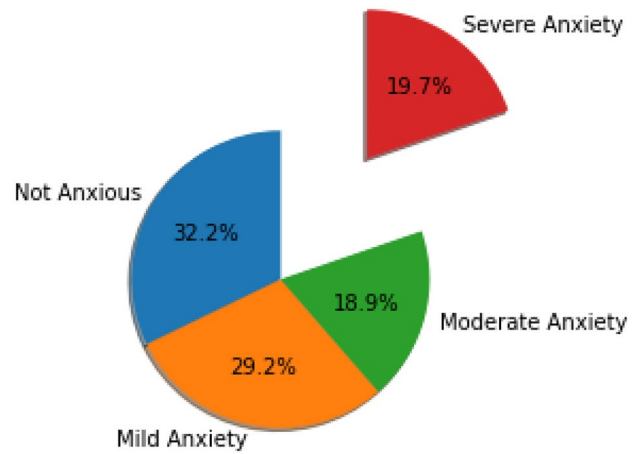| Metrics	GAD-7	PSQI
count	699.000000	699.000000
mean	8.494993	7.402003
std	6.326903	1.957294
min	0.000000	3.000000
25%	3.000000	6.000000
50%	7.000000	7.000000
75%	13.000000	9.000000
max	21.000000	13.000000

Fig. 7 Distribution of GAD-7 scores for the entire population

Table 2 Confusion matrix

	0 (absence of sleep disorder)	1 (presence of sleep disorder)
0 (absence of sleep disorder)	21	7
1(presence of sleep disorder)	2	29

4.3 ANOVA/Kruskal–wallis test

Both Anova and Kruskal–Wallis tests are used to study whether there is a statistically significant difference in the means of a dependent variable of different independent groups. Every data is ranked from 1 to x and assign the values that have not been tied. The test statistics is computed based on the number of observations, number of groups along with number of observations in each group. The average rank of each observation in group are computed using (2) and average of all ranks is calculated using Eq. 3.

$$.\bar{r}_i = \frac{\sum_{j=1}^{n_i} r_{ij}}{n_i} \tag{2}$$

$$.\bar{r} = \frac{1}{2}(N + 1) \tag{3}$$

Correlation among the values are identified where this value usually creates a little difference. Decision is based on the hypothesis values where the hypothesis is greater than null hypothesis then the values are rejected. There is no evidence of stochastic dominance between the samples if the statistic is not significant. If the test is significant, however, at least one sample stochastically dominates the other. Sample pairs are identified based on Kruskal –Wallis test. The Type 1 error rate tends to inflate when performing multiple sample contrasts or tests, raising concerns about multiple comparisons.

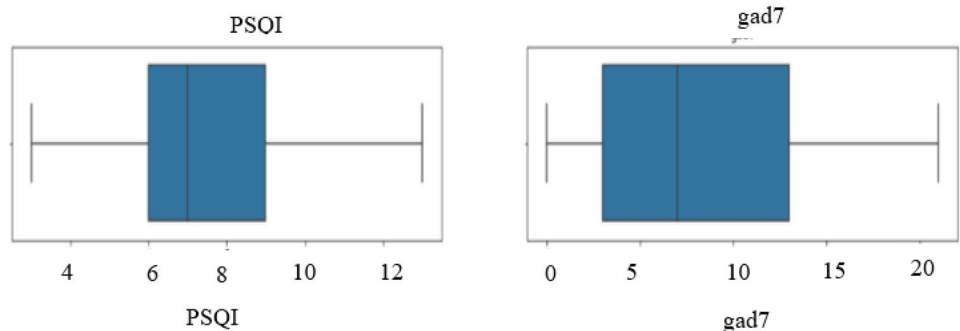**Fig. 6** Distribution of GAD-7 and PSQI scores visualized with box plots

Fig. 8 **a** Elbow method using distortion. **b** Elbow method using Intertia

(a) Elbow method using distortion

(b) Elbow method using Intertia

Fig. 9 **a** K-means clustering. **b** Agglomerative clustering
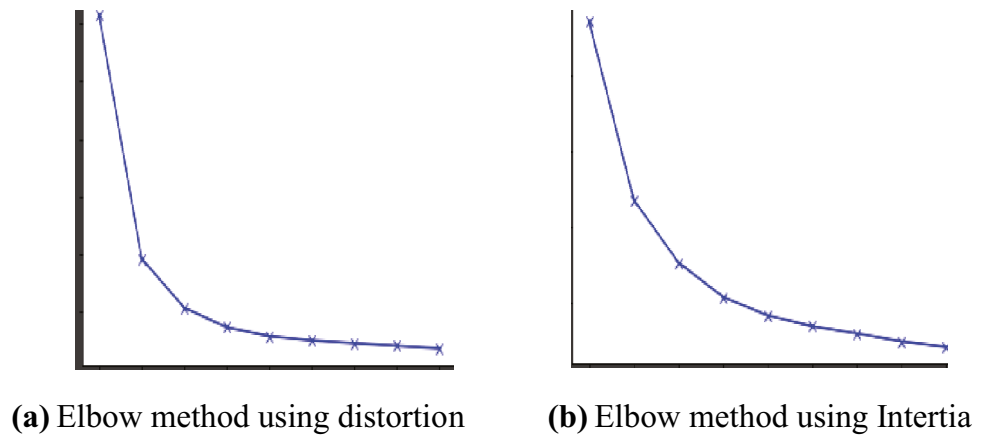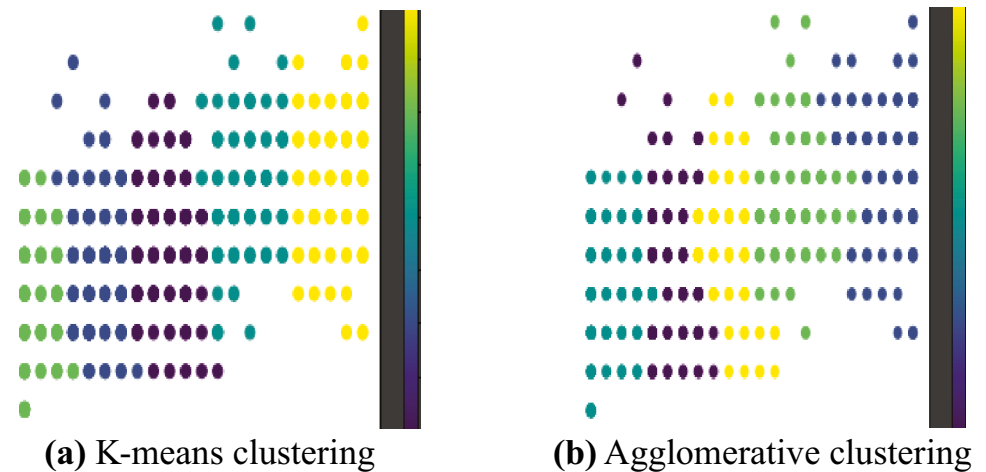

(a) K-means clustering

(b) Agglomerative clustering

For ANOVA to be applicable certain conditions have to be met. These conditions are data needs to be drawn independently and randomly, homogeneity of variance and normal distribution of data. ANOVA is robust to the assumption of normality, hence even if that condition isn't met, ANOVA can give meaningful results. Kruskal–Wallis test is a non-parametric alternative to ANOVA that can be used when data are not drawn from normal distribution. Levene's test was used to test homogeneity of variance. Shapiro–Wilk test was used to test whether data points belonged to normal distribution.

5 Results and analysis

The General Anxiety Disorder questionnaire (GAD-7) consists of a 4-point scale for a total of 7 questions. The scores hence range from 0–21 with the following ranges:

- $5 <= x < 10$—Mild Anxiety
- $10 <= x < 15$—Moderate Anxiety
- $x >= 15$—Severe Anxiety

The PSQI (Pittsburgh Sleep Quality Index) is also a self-reporting questionnaire that assesses sleep quality over a month, containing 19 items, creating 7 components. Each item is weighed on a 0–3 interval scale. The overall score ranges from 0–21, where lower scores denote better sleep quality. Specifically:

- $x > 6$ – Indication of sleep disorder, worsening with higher scores.

Table 3 Shapiro–Wilk test for Occupation-PSQI

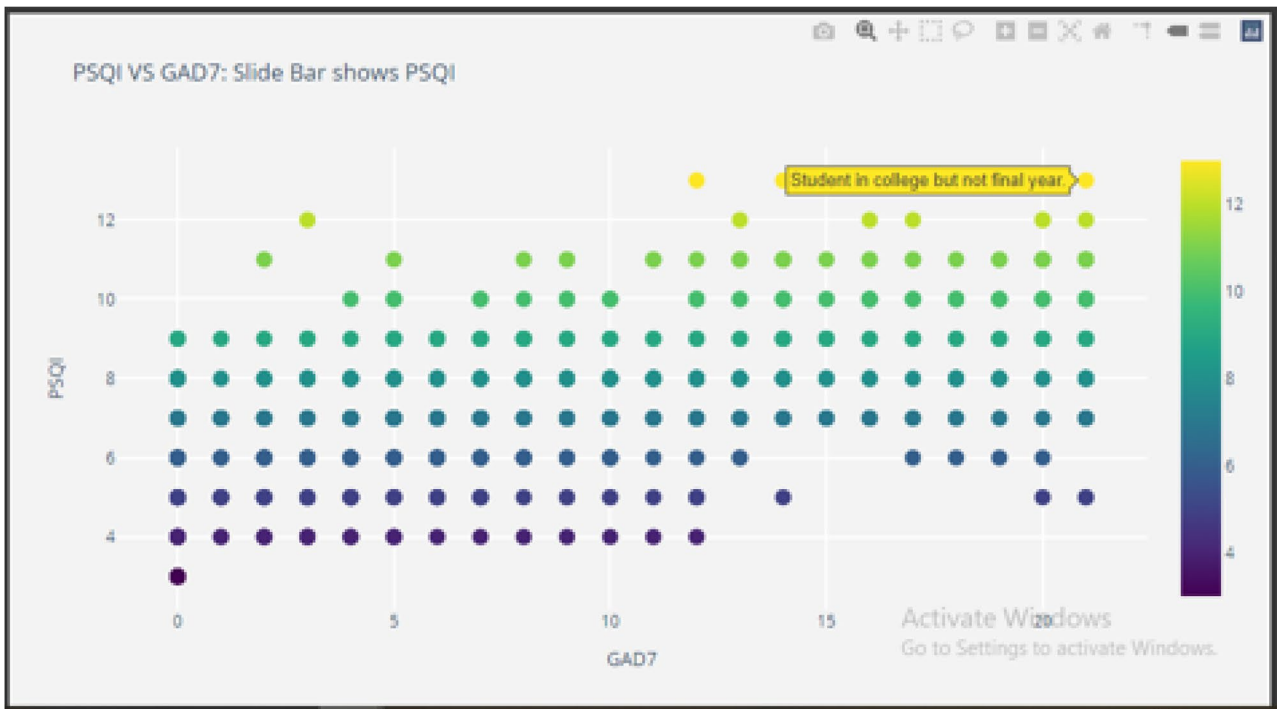| Occupation	p-value	Null hypothesis
Final year of college	0.0007	Reject
College but not final year	0.00001	Reject
Working professionals	0.0002	Reject
Unemployed	0.55	Accept
Just finished 12th grade	0.08	Accept
Home-maker	0.2	Accept
Healthcare professionals	0.43	Accept
School students	0.1	Accept
PhD researchers	0.4	Accept

Fig. 10 GAD7 vs PSQI

The validity of PSQI scores for Indian students (can be extended to population) are explored in this paper [11] and are consistent.

The correlation between GAD-7, PSQI and age are shown in the Fig. 5.

A positive correlation of 0.58 between GAD-7 and PSQI scores indicating a relationship between the two. Higher PSQI scores (worsening sleep quality) correlate to high GAD scores (worsening anxiety) are shown in the Table 1. Figure 6 shows the Distribution of GAD-7 and PSQI scores visualized with box plots.

19.7% of the participants show highly anxious GAD-7 scores, and 67% of the population have poor sleep quality (PSQI > 6). The distribution of GAD-7 scores for the entire population is examined and is shown in Fig. 7.

Random forest was used to build a classification model. The classes were presence of sleep disorder (PSQI > 7) and absence of sleep disorder (PSQI < 7). The objective of this is to study the factors affecting sleep. The model was built only for college students. Similarly, models can be built for other occupations as well. Table 2 shows the confusion matrix that is outlined post-training on the respective validation data.

The important features are anxiety score, worry about inability to understand concepts taught online, involvement of parents, college hours, worrying about other work load and deadlines. Clustering was used to get insight into the

Table 4 Mean PSQI for various occupation

	Occupation	Mean	Count
0	12 grade	6.888889	18
1	PhD Student	7.571429	7
2	College	7.624573	293
3	Emp	7.147727	88
4	Final Year	7.410405	173
5	Health Care	6.777778	18
6	Home	7.055556	18
7	School	7.270833	48
8	Unemp	6.809524	21

Table 5 Shapiro–Wilk test for Occupation-GAD-7

Occupation	p-value	Null hypothesis
Final year of college	0.00002	Reject
College but not final year	0.0000000008	Reject
Working professionals	0.0000003	Reject
Unemployed	0.02	Reject
Just finished 12th grade	0.02	Reject
Home-maker	0.002	Reject
Healthcare professionals	0.055	Accept
School students	0.004	Reject
PhD researchers	0.2	Accept

Table 6 Mean GAD-7 for various occupation

	Occupation	Mean
0	12 grade	8.166667
1	PhD Student	7.285714
2	College	9.300341
3	Emp	6.772727
4	Final Year	9.271676
5	Health Care	5.333333
6	Home	4.500000
7	School	8.791667
8	Unemp	7.095238

structure of dataset. Elbow method using distortion and Inertia was performed to determine the number of clusters to be initialised and is shown in Fig. 8(a) and (b). From the plots the optimal value of k is concluded as 5. The K-means clustering and the Agglomerative clustering are presented in the Fig. 9(a) and (b). From the visualization seen in Fig. 9, we can see that the decision surface is vertical, therefore clustering is done based on GAD-7 score, and we can conclude that PSQI score does not have significant variance.

The results of ANOVA/Kruskal–Wallis Test are presented to show the effect of occupation on PSQI score: the

Table 7 Paired t-tests for Occupation-GAD-7

Multiple Comparison of Means – Tukey HSD, FWER = 0.05

Group1	Group1	Mean diff	p-adj	Lower	Upper	Reject
12 grade	PhD Student	-0.881	0.9	-9.5141	7.7521	False
12 grade	College	1.1337	0.9	-3.5728	5.8401	False
12 grade	Emp	-1.3939	0.9	-64,076	3.6198	False
12 grade	Final Year	1.105	0.9	-3.695	5.905	False
12 grade	Health care	-2.83333	0.9	-9.2938	3.6271	False
12 grade	Home	-3.6667	0.6789	-10.1271	2.7938	False
12 grade	School	0.625	0.9	-4.7317	5.9817	False
12 grade	Unemp	-1.0714	0.9	-7.2968	5.154	False
PhD student	College	2.0146	0.9	-5.3978	9.427	False
PhD student	Emp	-0.513	0.9	-8.1242	7.0982	False
PhD student	Final Year	1.986	0.9	-5.4862	9.4581	False
PhD student	Health care	-1.9524	0.9	-10.5855	6.6807	False
PhD student	Home	-2.7857	0.9	-11.4188	5.8474	False
PhD student	School	1.506	0.9	-6.3355	9.3474	False
PhD student	Unemp	0.1905	0.9	-8.6492	8.2682	False
College	Emp	-2.5276	0.0248	-4.8836	-0.1716	True
College	Final Year	-0.0287	0.9	-1.887	1.8296	False
College	Health care	-3.967	0.1787	-8.6734	0.7394	False
College	Home	-4.8003	0.0416	-9.5068	-0.0939	True
College	School	-0.5087	0.9	-3.5266	2.5092	False
College	Unemp	-2.2051	0.7991	-6.5834	2.1732	False
Emp	Final Year	2.4989	0.0575	-0.0387	5.0366	False
Emp	Health care	-1.4394	0.9	- 6.4531	3.5743	False
Emp	Home	-2.2727	0.8938	-7.2864	2.741	False
Emp	School	2.0189	0.6545	-1.4587	5.4966	False
Emp	Unemp	0.3225	0.9	-4.3845	5.0295	False
Finalyear	healthcare	-3.9383	0.2088	-8.7383	0.8616	False
Finalyear	home	-4.7717	0.0528	-9.5717	0.0283	False
Finalyear	school	-0.48	0.9	-3.6418	2.6818	False
Finalyear	unemp	-2.1764	0.8324	-6.6551	2.3022	False
Healthcare	home	-0.8333	0.9	-7.2938	5.6271	False
Healthcare	school	3.4583	0.5321	-1.8984	8.815	False
Healthcare	unemp	1.7619	0.9	-4.4635	7.9873	False
Home	school	4.2917	0.2371	-1.065	9.6484	False
Home	unemp	2.5952	0.9	-3.6302	8.8207	False
School	unemp	-1.6964	0.9	-6.7672	3.3744	False

Table 8 Shapiro–Wilk test for Occupation-GAD-7

Contacts (Whether anyone close to participant has been tested positive)	p-value	Null hypothesis
Yes. But I am not fazed by it	0.000000001	Reject
Yes. And it is worrying	0.00000006	Reject
No	0.000000002	Reject

Shapiro–Wilk test are shown in Table 3, Levene's test achieved a p-value = 0.97 showing Null hypothesis is accepted. Since ANOVA is robust to assumption of normality, we can proceed with ANOVA p-value = 0.195 (Null hypothesis accepted). Kruskal–Wallis Test attained p-value = 0.192 (Null hypothesis accepted). p-value of both ANOVA and Kruskal–Wallis Test are greater than 0.05, therefore we can conclude that occupation has no effect on PSQI score.

From the Fig. 10, it can be observed that Student in college have high PSQI score but the difference in means of different occupation is not statistically significant. The mean PSQI for various occupation is shown in Table 4.

Similarly, the effect of occupation on GAD-7 score is detailed in Table 5 and 6. Levene's test attained p-value = 0.97 Null hypothesis is accepted, One-way ANOVA: Since ANOVA is robust to assumption of normality, we can proceed with ANOVA p-value = 0.0006 (Null hypothesis rejected). Kruskal–Wallis Test: p-value = 0.0002 (Null hypothesis rejected). p-value of both ANOVA and Kruskal–Wallis Test are less than 0.05, therefore we can conclude that occupation has effect on gad7 score.

To study which groups are different, paired t-tests were used and the values are shown below Table 7.

Effect of contacts being tested positive for COVID-19 on PSQI score are analysed and the Shapiro–Wilk test for Occupation-GAD-7 is shown in Table 8 and the mean PSQI for various contacts is listed in Table 9. Levene's test: p-value = 0.44 Null hypothesis is accepted. One-way ANOVA: Since ANOVA is robust to assumption of normality, we can proceed with ANOVA p-value = 0.0006 (Null hypothesis rejected). Kruskal–Wallis Test: p-value = 0.01 (Null hypothesis rejected). p-value of both ANOVA and Kruskal–Wallis Test are less than 0.05, therefore we can conclude that contacts has effect on GAD-7 score. Table 10 also specifies the recent work compared with the proposed work.

Table 9 Mean PSQI for various contacts

	Contacts	Mean
0	No	7.236052
1	Yes. And it is worrying	7.718876
2	Yes. But I am not fazed by it	7.216590

Table 10 Summarisation of the recent works along with the proposed work

Ref No.	Participants	Issues	Methodology	Assessment
44	Normal & COVID-19	Sleep problems with psychological distress	Observational studies (*Scopus, PubMed Central, ProQuest, ISI Web of Knowledge,* and *Embase*)	Pittsburgh Sleep Quality Index or Insomnia Severity Index
45	Chronic Rheumatic Diseases of Italian citizen	Poor mental health, sleep disorders among RDs during the pandemic	Online survey including the Perceived Stress Scale (PSS) and Impact Event Scale-Revised (IES-R)	Goodness-of-fit test, inal multivariate regression model
46	Italian population during lockdown	Prevalence of depressive symptoms, anxiety symptoms and sleeping disturbances	Questionnaires assessed socio-demographics characteristic, behaviors and healthcare access	Insomnia Severity Index
47	Washington State Twin Registry (WSTR)	Increased stress, anxiety, and sleep disruptions	Survey	Statistical Analysis
48	Bangladesh	Sleep disturbance proven risk factors for suicide	Web-based survey	multivariable logistic regression model
49	UK	Impact on mental health, self-isolation, keyworker status, suspected COVID-19 or ongoing COVID-19	Online survey-based study	Statistical analysis
Proposed	Indian	Anxiety and Sleep Disorders during COVID-19 lockdown	Online survey-based study	Machine Learning Model

6 Conclusion

This study investigated a systematic review and meta-analysis to discover the critical factors causing sleep and anxiety disorders during COVID-19 pandemic lockdown and the results were evaluated in terms of in GAD, PSQI, and machine learning models. The study examined a significant association between the symptoms such as sleep quality, anxiety, depression, and insomnia with the COVID-19 pandemic lockdown. Experimental results conclude that the participants showed prevalence of anxiety and sleep disorder to be high during the COVID-19 pandemic. 19.7% of the participants found to show high anxious GAD-7 scores and 67% of the population have poor sleep quality (PSQI > 6). There is a positive correlation of 0.58 between PSQI and GAD-7 scores of the participants. ANOVA results highlights the anxiety (GAD-7) to have effect on sleep quality (PSQI) whereas occupation is not a contributing factor to sleep quality. The higher score of PSQI findings summarized that people who have contacts with COVID-19 patients were found to be more anxious that contributed to worser sleep quality. The repercussion of strict lockdown, isolation and lifestyle changes affected young students with anxiety and sleep disorders. K-means clustering and the Agglomerative clustering confirmed that the anxiety score, worry about inability to understand concepts taught online, involvement of parents, college hours, worrying about other work load and deadlines as important features for college student dataset. This study is conducted in India with limited sample size covering different geographical conditions. In future, a larger cohort study with participants from different parts of the world will help to find extensive insights of different characteristics, behaviour and factors associated with sleep anxiety, depression during the global pandemic lockdown.

Declarations

References

1. Pappa S, et al. Prevalence of depression, anxiety, and insomnia among healthcare workers during the COVID-19 pandemic: A systematic review and meta-analysis. Brain Behav Immun. 2020;88:901–907.

2. Chen Q, et al. Mental health care for medical staff in China during the COVID-19 outbreak. Lancet Psychiatry. 2020;7(4):e15–e16.

3. Maunder RG, et al. Factors associated with the psychological impact of severe acute respiratory syndrome on nurses and other hospital workers in Toronto. Psychosoma Med. 2004;66 (6):938–942.

4. Bao Y, Sun Y, Meng S, Shi J, Lu L. 2019-nCoV epidemic: address mental health care to empower society. Lancet. 2020. https://doi.org/10.1016/S0140-6736(20)30309-3.

5. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, Rubin GJ. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet. 2020. https://doi.org/10.1016/S0140-6736(20)30460-8.

6. Mak IWC, Chu CM, Pan PC, Yiu MGC, Ho SC, Chan VL. Risk factors for chronic posttraumatic stress disorder (PTSD) in SARS survivors. Gen Hosp Psychiatry. 2010;32(6):590–8. https://doi.org/10.1016/j.genhosppsych.2010.07.007.

7. Wu KK, Chan SK, Ma TM. Posttraumatic stress, anxiety, and depression in survivors of severe acute respiratory syndrome (SARS). J Trauma Stress. 2005. https://doi.org/10.1002/jts.20004.

8. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological efects of quarantine, Toronto, Canada. Emerg Infect Dis. 2004. https://doi.org/10.3201/eid1007.030703.

9. Wu KK, Chan SK, Ma TM. Posttraumatic stress after SARS. Emerg Infect Dis. 2005. https://doi.org/10.3201/eid1108.04108 36.genhosppsych.2010.07.007.

10. Xiao H, Zhang Y, Kong D, Li S, Yang N. The Effects of Social Support on Sleep Quality of Medical Staff Treating Patients with Coronavirus Disease 2019 (COVID-19) in January and February 2020 in China. Med Sci Monit. 2020.

11. Xiao H, Zhang Y, Kong D, Li S, Yang N. Social Capital and Sleep Quality in Individuals Who Self-Isolated for 14 Days During the Coronavirus Disease 2019 (COVID-19) Outbreak in January 2020 in China. Med Sci Monit. 2020;26. https://doi.org/10.12659/MSM.923921.

12. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, Ho RC. Immediate psychological responses and associated factors during the initial stage of the 2019 Coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health. 2020. https://doi.org/10.3390/ijerph17051729.

13. Li Z, Ge J, Yang M, Feng J, Qiao M, Jiang R, Bi J, Zhan G, Xu X, Wang L, Zhou Q, Zhou C, Pan Y, Liu S, Zhang H, Yang J, Zhu B, Hu Y, Hashimoto K, Jia Y, Wang H, Wang R, Liu C, Yang C. Vicarious traumatization in the general public, members, and non-members of medical teams aiding in COVID-19 control. Brain Behav Immun. 2020. https://doi.org/10.1016/j.bbi.2020.03.007.

14. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res. 2020;288:112954.

15. Altena E, Baglioni C, Espie CA, Ellis J, Gavrilof D, Holzinger B, Schlarb A, Frase L, Jernelöv S, Riemann D. Dealing with sleep problems during home confnement due to the COVID19 outbreak: practical recommendations from a task force of the European CBT-I Academy. J Sleep Res. 2020. https://doi.org/10.1111/jsr.13052.

16. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. J Sleep Res. 2020. https://doi.org/10.1111/jsr.13074.

17. Li Y, Qin Q, Sun Q, Sanford LD, Vgontzas AN, Tang X. Insomnia and psychological reactions during the COVID-19 outbreak in China. J Clin Sleep Med. 2020. https://doi.org/10.5664/jcsm.8524.

18. Bekhet S, Alkinani, MH, Tabares-Soto R, Hassaballah M. An efficient method for covid-19 detection using light weight convolutional neural network. Comput Mater Contin. 2021;2475–2491.

19. Roy D, Tripathy S, Kar SK, Sharma N, Verma SK, Kaushal V. Study of knowledge, attitude, anxiety & perceived mental healthcare need in Indian population during COVID-19 pandemic. Asian J Psychiatr. 2020;102083.

20. Zhuo K, Gao C, Wang X, Zhang C, Wang Z. Stress and sleep: a survey based on wearable sleep trackers among medical and nursing staff in Wuhan during the COVID-19 pandemic. General Psychiatry. 2020;33:3.

21. Lin LY, Wang J, Ou-yang XY, Miao Q, Chen R, Liang FX, Zhang YP, Tang Q, Wang T. The immediate impact of the 2019 novel coronavirus (COVID-19) outbreak on subjective sleep status. Sleep Med. 2020.

22. Casagrande M, Favieri F, Tambelli R, Forte G. The enemy who sealed the world: Effects quarantine due to the COVID-19 on sleep quality, anxiety, and psychological distress in the Italian population. Sleep Med. 2020.

23. Mohamed A, Kamal M. Association of Student's Position in a Classroom and Student's Academic Performance Using ANOVA. In: 2015 Fifth International Conference on e-Learning (econf). IEEE; 2015. pp. 392–395.

24. Ibarra FA, Turizo D, Orozco-Henao C, Guerrero J. Generator Controller Tuning Considering Stochastic Load Variation Using Analysis of Variance and Response Surface Method. In: 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe). IEEE; 2019. pp. 1–5.

25. Jaber A, Abou Taam M, Makhoul A, Abou Jaoude C, Zahwe O, Harb H. Reducing the data transmission in sensor networks through Kruskal-Wallis model. In: 2017 IEEE 13th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). IEEE; 2017. pp. 71–78.

26. Al-Ghussain L, El Bouri S, Liu H, Zheng D. Clinical evaluation of stretchable and wearable inkjet-printed strain gauge sensor for respiratory rate monitoring at different measurements locations. J Clin Monit Comput. 2020;1–10.

27. Adhikari BK, Zuo W, Maharjan R, Han X, Amatya PB, Ali W. Statistical Analysis for Detection of Sensitive Data Using Hadoop Clusters. In: 2019 IEEE 21st International Conference on High Performance Computing and Communications; IEEE 17th International Conference on Smart City; IEEE 5th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE; 2019. pp. 2373–2378.

28. Sylvester EV, Bentzen P, Bradbury IR, Clément M, Pearce J, Horne J, Beiko RG. Applications of random forest feature selection for fine-scale genetic population assignment. Evol Appl. 2018;11(2):153–65.

29. Dimitriadis SI, Liparas D, Tsolaki MN. Alzheimer's Disease Neuroimaging Initiative, 2018. Random forest feature selection, fusion and ensemble strategy: Combining multiple morphological MRI measures to discriminate among healhy elderly, MCI, cMCI and alzheimer's disease patients: From the alzheimer's disease neuroimaging initiative (ADNI) database. J Neurosci Methods. 2018;302:14–23.

30. Huang N, Lu G, Xu D. A permutation importance-based feature selection method for short-term electricity load forecasting using random forest. Energies. 2016;9(10):767.

31. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing; 2018. vol. 336(1), p. 012017. https://ui.adsabs.harvard.edu/link_gateway/2018MS&E..336a2017S/doi:10.1088/1757-899X/336/1/012017.

32. Marutho D, Handaka SH, Wijaya E. The determination of cluster number at k-mean using elbow method and purity evaluation on headline news. In: 2018 International Seminar on Application for Technology of Information and Communication. IEEE; 2018. pp. 533–538.

33. Ding X, Yao J. Peer education intervention on adolescents'anxiety, depression, and sleep disorder during the Covid-19 Pandemic. Psychiatr Danub. 2020;32(3–4):527–35.

34. Fu W, et al. Psychological health, sleep quality, and coping styles to stress facing the COVID-19 in Wuhan, China. Transl Psychiatry. 2020;10(1):1–9.

35. Marelli S, Castelnuovo A, Somma A, et al. Impact of COVID-19 lockdown on sleep quality in university students and administration staff. J Neurol. 2021;268:8–15. https://doi.org/10.1007/s00415-020-10056-6.

36. Zhou S-J, et al. Sleep problems among Chinese adolescents and young adults during the coronavirus-2019 pandemic. Sleep Med. 2020;74:39–47.

37. Beck F, et al. Would we recover better sleep at the end of Covid-19? A relative improvement observed at the population level with the end of the lockdown in France. Sleep Med. 2021;78:115–119.

38. Salehinejad MA, et al. Circadian disturbances, sleep difficulties and the COVID-19 pandemic. Sleep Med. 2021.

39. Labarca G, et al. Undiagnosed sleep disorder breathing as a risk factor for critical COVID-19 and pulmonary consequences at the midterm follow-up. Sleep Med. 2021.

40. Zhou J, et al. Mental health response to the COVID-19 outbreak in China. Am J Psychiatry. 2020;177(7):574–575.