



Impact of SARS-CoV-2 in Online Education, Predicting and Contrasting Mental Stress of Young Students: A Machine Learning Approach

K. Parthiban¹ · Digvijay Pandey² · Binay Kumar Pandey³

Received: 5 April 2021 / Revised: 22 May 2021 / Accepted: 28 May 2021 / Published online: 8 June 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd. 2021

Abstract

The nation has already had to contend with such a major public health issue when this COVID-19 disease was propagated in March 2020. The corona's propagation does have an effect on all areas of people's lives. The hostile effect on education systems throughout all locations is indeed a diverse expansion impact. A complete shutdown keeps adding fresh problems for students to learn and also for educators to effectively manage the class, likely to result in the conversion of such an offline education system into a virtual class. This study investigated and described various online learning platforms, as well as teaching resource distribution technology and modern technologies used to ensure that students can study. Overall, this relates to channels that can be used to create a confidential environment in order to prevent copying in online tests. This also outlines so many difficulties in managing online teaching methods, like learners' perceptions of e-learning being more traumatic and having a negative impact on general wellbeing and social relationships. So, a technique has been proposed to provide learners with a superior online classroom teaching experience, allowing the online classroom to be as good as, if not better than, a single online classroom. This study focused on daily teaching methods that employ online learning supported by a machine teaching approach to provide an individual with a relevant stress-free solution.

Keywords COVID-19 · Stress · Online class · Respondent · Clustering and classification

Introduction

On Jan 30, 2020, the World Health Organization (WHO) announced coronavirus disease (COVID-19) as just a wide-spread infection. This was first published in China in

December 2019, and since then has rapidly spreading across the world, negatively impacting a number of nations in Europe, America, and Asia [1, 2]. As from the third week of March, UNESCO suggests that almost all colleges and universities stay closed. Bhutan is the first Asian country to the shutdown of educational institutions in the first week of March 2020 as well as a decrease in operating hours [3].

Starting 1 August 2020, the country will be fully shut down [4]. In the meantime, activities are permitted, workplaces resumed, even educational institutions resumed to selected levels while continuing to provide online classes for everyone else. Bhutan's educational institution closure affects over 170,000 students in grades PP–XII nowadays.

COVID-19 appears to have had a much greater impact on education throughout the academic year and will continue to do so in the near future. School systems, institutions, even universities have ceased offering face-to-face instruction. Furthermore, the Ministry of Human Resource Development in India has issued several circulars stating

✉ Digvijay Pandey
digit11011989@gmail.com

K. Parthiban
parthiban14it@gmail.com

Binay Kumar Pandey
binaydece@gmail.com

¹ Department of Computer Science and Systems Engineering, SreeVidyanikethan Engineering College (Autonomous), Tirupati, Andhra Pradesh, India

² Department of Technical Education, IET, Dr. A.P.J. Abdul Kalam Technical University Uttar Pradesh, Lucknow 226021, India

³ Department of Information Technology, College of Technology, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, US Nagar, Uttarakhand, India

that the educational system must not resume routine schoolwork, including examinations, during such a COVID-19 and provided a very good suggestion to implement an online class across all students as in courses available in their institution [5]. There seems to be an urgent need for new instructional and evaluation methods. COVID-19 [6] pandemic now provides a chance to prepare a way for the formation of online learning [7].

This work's primary aim is to evaluate the effects of full shutdown throughout COVID-19 by a new coronavirus on learners' education and general stress [8] and to achieve a high standard of education and student achievement. The number of online courses attended focused on contextual information should be examined using a machine learning technique [9].

Learners' stress can be caused by their electronic communication, their interpersonal lives, their perceptions, their teaching methods, their general stability, and their mood swings. This paper is organised as follows: Related works, proposed methodology, experimental design, result and discussion, conclusion and possible scope, and bibliography.

Related Works

As [10] provide descriptions here on the assessment of stress and anxiety for university students, as well as the investigators' analysis on datasets that included the execution of two tests, and then, based on the results, the researchers measured the adequate versus stress now at the beginning and end of the final examination using a decision tree-based algorithm [11] that helps to determine stress rate. Anxiety about one's life as a result of events that occurred, as well as a lack of sufficient time to complete tasks [12]. This [13] describes information about EEG data utilizing headsets including defined amounts of anxiety and also classification accuracy could be strengthened more.

The study relies on Tunisian Schooling information to forecast output utilizing regression tree with Random Forest [14] Moreover, according to [15], a study of new research on machine learning used in relation to COVID-19 is performed, with a focus on the usefulness of machine learning towards two important components: COVID-19 detection, including estimation of overall mortality as well as seriousness using widely available clinical and biochemical details [16].

There is discussion of methodology varieties, training large datasets, also feature extraction. The vast majority of machine learning techniques employed in these two processes were supervised learning algorithms. The developed models are still for use in real-world implementations, and most of the related studies are now in the early stages of

development. Predictive and therapeutic characteristics identified through ML algorithms become compatible with published research findings.

Any use of imbalanced data, that is vulnerable to sampling error, is indeed a disadvantage to existing applications. [17] incorporates predictive methods, and also machine learning-based approaches, which can be specifically applicable to new COVID-19 patterns to estimate patients' discharge capacity and determine whether patient clinical information might affect the duration of hospitalisation. Even though this report provided useful insights into patient healing time. Implementation of such data-driven methods is expected to produce considerable attention in the near future as more accurate clinical evidence becomes accessible. COVID-19, commonly known as coronavirus disease, is indeed a pandemic induced by a coronavirus.

The large number of people infected with this virus will have to face mild to serious viral infection. If it is not properly cared for the elderly, disabled, and convalescing individuals, and those with chronic health issues, such as diabetes, chronic respiratory problems, including cardiovascular problems, are always at risk of catching such an illness. There are currently no specific medications or immunizations against COVID-19. Nonetheless, there are numerous ongoing clinical studies evaluating potential therapies, including vaccines. Identifying the hazards from COVID-19, computational scientists have begun utilizing different types of methods from computational intelligence to determine the presence of the infection leveraging X-rays and CT pictures.

The most significant major obstacle is the scarcity of original dataset and so there are fewer specialists that classify the knowledge specific to such a resistant strain of human infection. Artificially intelligent technologies could be easily built then optimised for incorporating new AI models and using the potential to adjust and connect them with clinical trials knowledge to resolve the emerging COVID-19 community and also the innovative difficulties associated with that as well.

In this study, researchers aimed at a few machine learning and deep learning methods that were used to examine coronavirus information. The aim of this [18] study would be to prepare a multifunctional system to determine the danger of experiencing dangerous situations in COVID-19 patient populations.

Researchers observed a maximum of 1040 patients with such a positive RT-PCR assessment of COVID-19 at a major hospital in So Paulo, Brazil, from March to June 2020, of which 288 (28%) had a serious prognostic value, i.e., intensive care unit admission, the use of respiratory support, and mortality. In this work, five machine learning

techniques are trained using regularly gathered scientific, medical, and population information.

A representative sample of % of patient populations was used to train algorithms, with the remaining % used for work performance, simulating previously unknown information. Each model has been trained through integrating two of the three consequences to forecast another in attempts to see whether the algorithm might achieve common extreme prognostic trends.

All algorithms demonstrated extremely higher prediction accuracy (average AUROC of 0.92, sensitivity of 0.92, and specificity of 0.82). For both the multi-functional algorithms, the three most important parameters are lymphocyte per C-reactive protein ratio, C-reactive protein, and Braden Scale. The results suggest that machine learning models might be possible to forecast non-specific negative COVID-19 consequences from regularly gathered data.

Proposed Methodology

As a result of something like the approaching shutdown of schools and universities due to disease outbreak, a schooling institution has already been in turmoil as of March. There is indeed a concern about completing the curriculum now as the academic year is coming to an end. As just a result, there are a variety of online courses, like Zoom, Google Class, Webex Meet, Microsoft Meet, Teams, Meeting, and so forth. Every coin does have 2 aspects: one part is a good opportunity to implement classes online, whereas the other part is a variety of difficulties with students attending courses. That also causes an amount of anxiety.

In this study, a K-means clustering technique is used to divide student groups into an adequate number of clusters, and the harvested cluster would provide information about students with special needs by displaying the highest and lowest ratios of online courses throughout COVID-19.

As given in Fig. 1, few of the methods throughout machine learning methods are indeed the classification technique. k-NN, logistic regression, naive Bayes, support vector machine, decision trees, and Random Forest do seem to be examples of different classifiers. It correctly classifies classes with such a low degree of error to see if there is any link here between various personality characteristics but also considerations linked with entrepreneurial activities. A confusion matrix correctness has been used as a metric for analysing the different classification techniques.

The classification model is a special data mining method that disintegrates given information in individual cases. This assigns a particular piece to an instance with a really

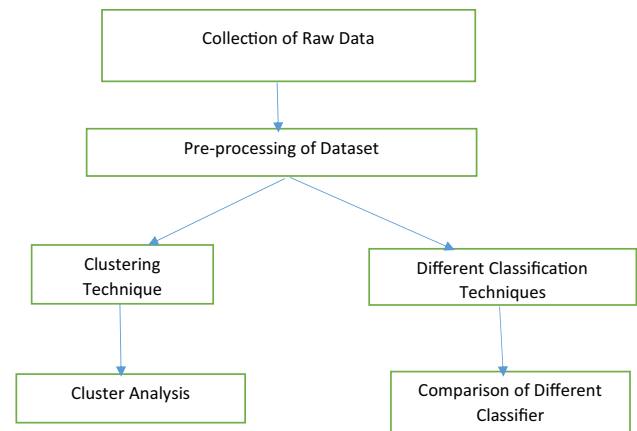


Fig. 1 Machine learning general methodology

low risk of failure. It indeed would get rid of modelling crucial information classes within such a provided information set. In this work, a classification algorithm is used to predict people's anxiety levels. A data ratio of 1:3 is used to perform training and testing.

Experimental Setup and Results

Data Collection and Pre-process

The dataset has been compiled from 647 understudies' reactions to online course suggestions. A survey was conducted via a list of questions. Participants, or learners, are requested to fill in a form based according to their own experiences. There has been a maximum of 45 questions framed and split the responses of participants into two categories: ones collected prior to online courses and those collected during online courses.

The very first category includes collected data prior to online courses, whereas the second type includes collected data during online activity, that also causes students to explore pressure. Each question would have five possible responses: outstanding, very good, good, satisfactory, and negative. On the basis of responses, participants were grouped as extremely distressed, frustrated, or ordinary. As from the Adolescent Learners' Survey, that also aids in the analysis of levels of stress.

Data Preparation and Data Visualization

Further to the collection of information, a procedure with cleanup will start to erase mistakes, multiple entries, and inconsistent data. A primary objective thereafter is visualisation. This must be done in order to identify connections among attributes and to carry out data exploration

assessment. Furthermore, data are divided into two parts: data for training and data for testing.

Tools Used

R and Python have been preferred owing to the accessibility of popular libraries. Pandas, Scikit, and matplotlib are Python libraries for statistical analyses, visual analytics, clustering, and classification. Through R, dplyr, caTools, e1071, factoextra, and ggplot2 libraries are used.

Clustering

The K-means classifier has been the most effective way for cluster analysis [19]. Based on the feedback from students in online classrooms throughout lockdown, set clusters of connected learners have been created. This assists in determining students who may have similar interest as well as anxiety. Typically, to instantiate a method, a fixed number of clusters, such as 'k,' must be stipulated. A famous Elbow method is used for appropriate clustering experimentation.

The optimum level of 'k' could be ascertained here. In this scenario, this entails the processing of complex intervals where the clustering algorithm begins to grow. Furthermore, a clustering scoring system is recorded as a function of number of clusters. This Elbow process achieves exactly 3 or 4 clusters (see Fig. 2).

Only after clusters were evaluated, as shown in Fig. 3, was it discovered that the majority of respondents who attended online courses throughout COVID-19 now fall into Cluster 2. Outcome has been described in terms of the number and percentages of students who have experienced anxiety in such clusters. A proportion visualisation is by far the most useful technique for accurately assessing the total number of students under mental anguish.

According to Table 1, the greatest number of students has been dissatisfied, primarily during online sessions of courses during the COVID-19 outbreak lockdown. Second, a cluster of the distressed was constructed. Furthermore, the reduced number of participants out of three with cluster is students with a typical mind set.

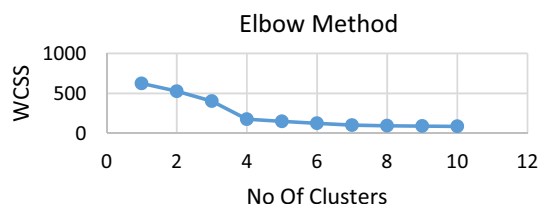


Fig. 2 Elbow method used to initialize the number of centroids

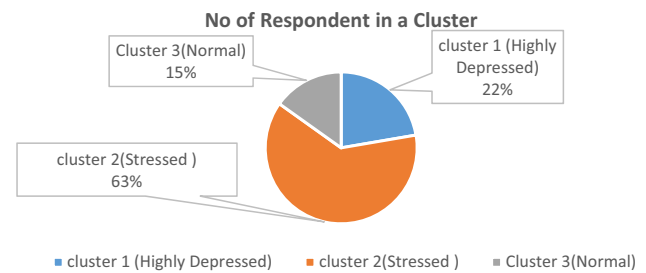


Fig. 3 Number of respondent in a cluster

Classification

As shown in Table 2, Classification is indeed a distinguishable methodology throughout machine learning approaches. In this, classification methods like Random Forests, k-nearest neighbors ($k = 5$), support vector machine, logistic regression, naive Bayes, and decision tree have been used as the Students' responses depending on the responses provided mostly during web-based discussion handled throughout COVID-19 shutdown.

A training-test dataset would be used to feed a classification method. The training-test set of data separation with 84:33 is often used. A study provides a much more qualitative method here. The variables then are considered for classification methods.

Random Forest

The Random Forest algorithm is indeed a tree-based ensemble. These are composed of multiple decisions that also combine to create a forest. This Random Forest model can be used for classification as well as regression [20]. The whole algorithm processes a large amount of data. In our study, 600 decision trees and Ginni for the imperfection indices are used to improve classification efficiency.

K-Nearest Neighbour

The KNN method is utilized for sharing relevant information. Student groups are classified here depending on the responses provided. If there are three different classification groups, the data would be segmented into the respective components [21]. k's valuation is always positive. The number of nearby neighbourhoods would be investigated.

Support Vector Machine

SVM has indeed been efficiently and frequently adopted for experimentally validated problems in various areas, including image classification, text categorization, medical

Table 1 Competency of students in a cluster

Cluster	Percentage of Students	No of Students
Cluster 1 (highly depressed)	22.30%	144
Cluster 2(stressed)	62.57%	405
Cluster 3(normal)	15.13%	99

Table 2 Comparison of different classification algorithms

SL. No	Algorithm	Sensitivity (%)	Specificity (%)	Accuracy (%)
1	Random Forests(500 trees)	100	66.6	83.3
2	k-nearest neighbors ($k = 5$)	70	44	55.18
3	Support vector machine	75	98	87.75
4	Logistic regression	71	91.27	79.13
5	Naïve Bayes	66	75	71
6	Decision tree	61	65	57.14

science, and mechanical machine diagnostic testing. This classifier's normal response is mostly on hyperplane [22].

Logistic Regression

Logistic regression was chosen since it performs well enough in binary classifiers. It isolates actions by establishing a threshold between positive and negative responses. Logistic functions have been used to estimate the probability distribution of a positive or negative tag [23].

Naïve Bayes

Naive Bayes classification algorithms have been referred to as “classification algorithms” collaboratively. They operate with probabilities. Naive Bayes classifiers are incredibly versatile. For just a specified target value, they are often predicted to have been provisionally independent [24].

Decision Tree

The ID3 algorithm is being used to build a decision tree by conducting a top-down, greedy search via the given sets to assess every trait at every tree node. The information gain model is normally used during decision tree measuring to assess good characteristics for every node of the created decision tree [25].

Performance Parameters

Sensitivity

The true positive rate is used to determine sensitivity. It's the proportion of true positives (TP) to true positives and false negatives (TPFN) (FN). As shown in Eq. 1, this

determines a model's ability to accurately classify stressed students due to online courses throughout COVID-19.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (1)$$

Specificity

This is referred to as true negative rate. As seen in Eq. 2, it should be the proportion of true negative (TN) to true negative and false positive (FP). It requires the ability to recognize with accuracy as part of digital classrooms in COVID-19.

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2)$$

Accuracy

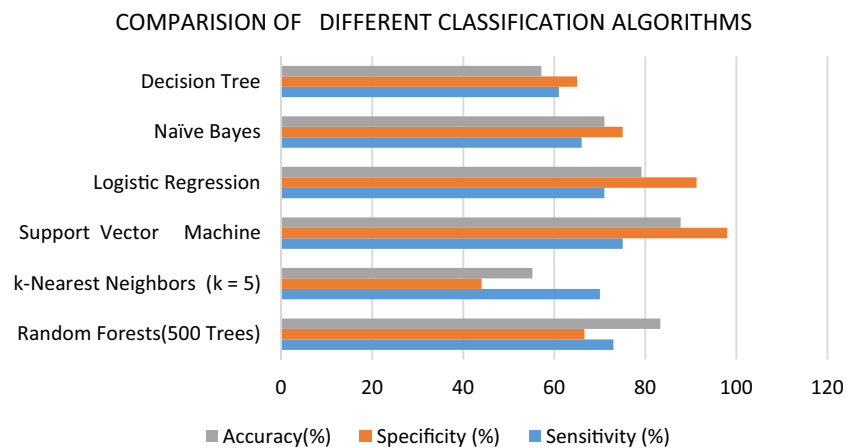
As shown by Eq. 3, this is now the proportion of true positives plus true negatives to true positives plus true negatives plus false positives plus false negatives. It computes the number of cases that are classified correctly.

$$\text{Accuracy} = \frac{TP + TN}{TN + FP + TP + FN} \quad (3)$$

SVM [26] is the best performing algorithm among the six machine learning algorithms that are based on comparative calculations, but the accuracy noted by respondent for having attended online courses is significant.

Conclusion and Future Work

In today's world, students are overly supported for their education by mentors and lecturers. Not only the lessons, but also awareness of COVID-19 and lockdown measures.

Fig. 4 Comparison of different classification algorithms

The primary responsibilities of a mentor are to assist students in relieving stress and providing double friendship for their emotional feelings. Due to COVID-19, the uncertainties related to their examinations and growth path by way of internships, jobs, etc., are a major cause of mental stress for the students. The results show that the student believes that an online class can be used to supplement knowledge but cannot replace classroom learning due to one-on-one interaction.

Based on 647 students' feedback for online classes during COVID-19, a dataset was created. The parameters such as sensitivity, specificity, and accuracy are computed, and both clustering and classification methods are discovered by using machine learning techniques, which help to improve the mental health of future pillars of nations and make students mentally sound (see Fig. 4). The inherent strength, positivity, and maturity of the students are combined in their one-line message to the world whose spirit is Stay Home, Stay Safe, and Stay Together as depicted.

Acknowledgements Many thanks Department of Technical Education for continuous support.

Funding No funding is provided for the preparation of manuscript.

Declarations

Conflict of interest All authors declare that they have no conflict of interest.

Availability of Data and Material The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Ethical Approval This article does not contain any studies with human. Participants or animals performed by any of the authors.

References

- Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR (2020) Severe acute respiratory syndrome coronavirus (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int J Antimicrob Agents* 55(3):105924
- Khongsai L, Anal TSSC, Rapheileng AS et al (2020) Combating the spread of COVID-19 Through community participation. *Glob Soc Welf*. <https://doi.org/10.1007/s40609-020-00174-4>
- Kuensel (2020) First confirmed coronavirus case in Bhutan. Kuensel. <https://kuenselonline.com/first-confirmed-coronavirus-case-in-bhutan/>
- Palden T (2020) Women test COVID-19 positive after five tests locking down entire country. Kuensel 1–2
- <https://mhrd.gov.in/circulars-orders-notification>
- Manne R, Kantheti S (2020) COVID-19 and its impact on air pollution. *Int J Res Appl Sci Eng Technol (IJRASET)*. <https://doi.org/10.22214/ijraset.2020.32139>
- Dhawan S (2020) Online learning: a panacea in the time of COVID-19 crises. *J Educ Technol* 49(1):5–22. <https://doi.org/10.1177/0047239520934018>
- Kassaw C, Pandey D (2020) The current mental health crisis of COVID-19 pandemic among communities living in Gedeo Zone Dilla, SNNP, Ethiopia, April 2020. *J Psychosoc Rehabil Ment Health*. <https://doi.org/10.1007/s40737-020-00192-7>
- Khattar A, Jain PP, Quadri SMK (2020) Effects of the disastrous pandemic COVID 19 on learning styles, activities and mental health of young Indian students-a machine learning approach. In: *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020)*, IEEE Xplore Part Number: CFP20K74-ART; ISBN: 978-1-7281-4876-2
- Ahuja R, Bangab AA (2019) Mental stress detection in university students using machine learning algorithms. *Procedia Comput Sci*. <https://doi.org/10.1016/j.seps.2019.06.009>
- Adnan, Nadia, et al (2012) University students stress level and brainwave balancing index: comparison between early and end of study semester. *Research and Development (SCoReD)*, 2012. In: *IEEE Student Conference on*. IEEE 2012
- Slavich GM (2019) Stressnology: The primitive (and problematic) study of life stress exposure and pressing need for better measurement. *Brain Behav Immun* 75:3–5. <https://doi.org/10.1016/j.bbi.2018.08.011>
- Gaikwad, Paithane (2017) Novel approach for stress recognition using EEG signal by SVM classifier. In: *Proceedings of the IEEE 2017 international conference on computing methodologies and communication IEEE 2017*

14. Rebaia S, Yahia FB, Essidb H (2020) A graphically based machine learning approach to predict secondary schools performance in Tunisia. *Socio Econ Plan Sci* 70:100724
15. Alballa N, Al-Turaiki I (2021) Machine learning approaches in COVID-19 diagnosis mortality, and severity risk prediction: a review. *Inform Med Unlocked*. <https://doi.org/10.1016/j.imu.2021.100564>
16. Bokam Y, Guntupalli C, Gudhanti S, Kulandaivelu U, Alavala R, Alla N, Manne R (2021) Importance of pharmacists as a front-line warrior in improving medication compliance in Covid 19 patients. *Indian J Pharm Sci* 83(2):393–396
17. Nemati M, Ansary J, Nemati N (2020) Machine-learning approaches in COVID-19 survival analysis and discharge-time likelihood prediction using clinical data. *Patterns*. <https://doi.org/10.1016/j.patter.2020.100074>
18. Fernandes FT, de Oliveira TA, Teixeira CE et al (2021) A multipurpose machine learning approach to predict COVID-19 negative prognosis in São Paulo. *Brazil Sci Rep* 11:3343. <https://doi.org/10.1038/s41598-021-82885-y>
19. Singh S, Gill NS (2013) Analysis and study of K-Means Clustering Algorithm. *Int J Eng Res Technol (IJERT)* 2:2546–2551
20. Khosrowabadi R, Quek C, Ang KK, Tung SW, Heijnen M (2011) A brain-computer interface for classifying EEG correlates of chronic mental stress. In: *IJCNN*. p. 757–762
21. Zheng S, Ding C (2020) A group lasso based sparse KNN classifier. *Pattern Recognit Lett* 131:227–233
22. Pirra M, Diana M (2019) A study of tour-based mode choice based on a Support Vector Machine classifier. *Transp Plan Technol* 42(1):23–36
23. Omari MA, Al-Hajj M, Hammami N, Sabra A (2019) Sentiment classifier: logistic regression for Arabic services Reviews in Lebanon. In: 2019 international conference on computer and information sciences (iccis)
24. Subhani AR, Mumtaz W, Saad MNBM, Kamel N, Malik AS (2017) Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access* 5:13545–13556
25. Sharma H, Kumar S (2016) A survey on decision tree algorithms of classification in data mining. *Int J Sci Res (IJSR)* 5(4):2094–2097
26. Pandey D, Pandey BK, Wairya S (2020) Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images. *Soft Comput*. <https://doi.org/10.1007/s00500-020-05245-4>
27. UNESCO-COVID-19 Educational disruption and response. [Online]. Available: <https://en.unesco.org/covid19/educationresponse>
28. Aishwarya T, Kumar VR (2021) Machine learning and deep learning approaches to analyze and detect COVID-19 a review. *SN COMPUT SCI*. <https://doi.org/10.1007/s42979-021-00605-9>
29. Istaiteh O, Owais T, Al-Madi N, Abu-Soud S (2020) Machine learning approaches for COVID-19. In: 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA). p. 50–57, Online Presentation
30. Rustam F, Reshi AA, Mehmood A, Ullah S, Wonon B, Aslam W, Choi GS (2020) COVID-19 future forecasting using supervised machine learning models. *IEEE Multidiscipl Open Access J* 8:101489–101499

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.