New Generation Computing
https://doi.org/10.1007/s00354-023-00224-3

A Novel Model to Predict the Effects of Enhanced Students' Computer Interaction on Their Health in COVID-19 Pandemics

Nidhi Agarwal[1,2] · **Sachi Nandan Mohanty**[3] · **Shweta Sankhwar**[4] · **Jatindra Kumar Dash**[5]

Received: 20 November 2022 / Accepted: 4 May 2023

Abstract

During the COVID-19 pandemic time, educational institutions have really played a good role in imparting online education to students. Their career and academic tenure were not affected as contrary to the past pandemics throughout world history. All this has been possible through long sessions of classes, quizzes, assignments, discussions, chat interactions, and examinations through online video-based learning using computer interactive measures. The students were privileged to utilize digital technologies for longer durations for learning purposes. However, these long stretches have adversely affected their body postures, and physical and mental health as they majorly remain confined to chairs with restricted levels of physical activities. Thus, there is a need to have a model which can act as an insight for parents, doctors (pediatricians), and academic policymakers to decide on maximum hours for online teaching and related activities during future pandemics. The novel model proposed in this work helps to predict the impact of enhanced students' computer interactions on their physical and mental health. The method proposed uses a novel model which is advanced and computationally strong. The model follows a two-step methodology, where at the first level, a variant of already existing machine learning algorithm is proposed and at the next level, it is optimized further using a hybrid bio-inspired optimization algorithm. The model consists of proposing a variant of XGBoost model (step1 optimization) followed by a hybrid bio-inspired algorithm (step2 optimization). The work considers a humongous dataset with varied age groups of students with more than 10 attributes. The proposed model is highly efficient in making predictions with 98.07% accuracy level and 98.43% F1-score. The time complexity of the model obtained is also of order of "n" where "n" depicts the number of input variables. Strong empirical results for other parameters also like specificity (95.63%) and sensitivity (96.74%) ascertain the enhanced predictive power generated using the proposed model. An extensive comparative study with other machine learning models ascertains the elevated accuracy and predictive power using the proposed

Extended author information available on the last page of the article

Published online: 15 June 2023

model. Till now none of the researchers have proposed any such pioneering tool for parents, doctors, and academicians using advanced machine learning algorithms.

Keywords COVID-19 · Effect on health · Machine learning · Academia · Students · Bio-inspired algorithms · XGBoost model · Heuristic approach

1 Introduction

Past pandemics in the world history brought academia to a complete standstill. There were huge academic losses to students in the form of year loss, zero-education, delayed admissions, and even no admissions. But the academia industry played completely different role during the current pandemics.

"COVID-19". The sole credit goes to advancement in technologies related to human–computer interaction. Students were connected virtually to their schools, classrooms, teachers, and most importantly with their peers [30, 36]. The technological advancements even provided them ways to attempt quizzes, assignments, tutorials, etc., both graphically and textually. Even the video-call-based applications provide students with the facility to keep their classroom-customized background where they can have a complete classroom feel (Shao et al. 2019); [15]. All this lead "COVID-19" not hamper the students' academia irrespective of their age groups. The students from class Nursery to class 12 can continue their education. Even the higher education aspiring students can take admissions in various institutes through online mode [2–4, 7]. The entrances of various national and international level institutions are held through online mode [33, 40]. In a nutshell, we can say that unlike previous pandemics, COVID-19 had not disrupted student education. Though there are few exceptions also, as according to UNESCO Institute for Statistics, few countries worldwide could not cope up with online educational system and thus closed the academic setting [35, 49]. But majority of the educational institutions are able to continue with the academics but with enhanced computer and other ICT interaction. Thus, it can be concluded that if students, of any age group, want to continue their academia, then they are compelled to give longer and longer stretches online. Their e-learning or e-education can only be possible if they remain confined to their small sized screens of mobiles, laptops, or tabs [28, 39] for long durations. This enhanced spending of time online can pose adversely on their physical and mental health [20, 23]. The students' physical activities and all sort of movements get restricted [21, 46]. Because of the pandemic spread fear, they even canmot play outside with their peers. Thus, physical movement gets reduced and remaining confined to the couches enhances. In such a scenario, it is the moral duty of the nation as a whole and parents, teachers in general to balance the e-educational benefits with students' health (Pal &Vanijja, 2017a, 2017b). It solely lies upon the nation builders to decide whether this type of e-learning should be made compulsory in case of such serious pandemics or not [31, 53]. In many instances, even the academia could not get ample time to arrange e-classroom mode effectively and efficiently so that students could be benefited to the maximum extend (Dang et al. 2016, Demaidi et al.

2019). Adoption of e-classroom academia setting has resulted in improved outcomes in terms of improving students' knowledge, perception, [8, 27], students' grasping power [11, 51], and relieving academia stress (Hoic-Bozic et al. 2009, Hargittai, 2003. But as a responsible nation builder, teacher, principal, parent, we need to analyze that remaining couch confined with restricted physical movement for kids (who are the future for every nation is also affecting their physical and mental health development [9, 44]. This paper is an effort to make nation builders, academia-related authorities, parents, and doctors aware with the negative effects of e-classroom education to the students.

Though, a handful of work is available which highlights the shortcomings of the e-classroom system where the adverse effect for the students is worked upon. In this work, it is shown that continuous stretches of sitting may retard the response time of the students, which may also discourage them from participating in class and the feeling of loneliness and isolation can also arise. Students may reduce their participation in various activities being done in the class [20, 22]. Another study highlights the requirement of a personalized student feedback system. As we all know that, in offline mode of education, the teacher can instantly understand the various feedbacks related to students understanding but it is not possible in e-classroom education for students [32, 52]. It may also rise the feeling of dejection among the students as the field is isolated and lonely [20]. As we know that some of the students are so much habitual of interacting among each other in offline mode of education and they always try to build their own space among their peers that they feel quite isolated and rejected in online mode of education. This may lead to their social separation from the peers which may affect their learning and educational growth. They may also get physically and mentally disturbed by the isolation faced during online classes [46]. Many a times, even the students are not self-motivated to turn up themselves happily for the e-classes, they are forced to take the online classes because of which many a times they either turn off their cameras or microphones or both. This lack of self-motivation is considered as one of the major reasons for the diminishing importance of e-classroom education during COVID-19 [24, 29]. Thus, if the policy makers make e-classroom concept compulsory, then it may also bring some amount of discontent among the illiterate cult of the country. There is no doubt that the whole world's technology has advanced a lot now a days. The major credit goes to advancements in the domain of human–computer interaction, but still there are a lot many people who are directly or indirectly related to the students like their parents, grandparents but are completely unaware of these technologies. In such a scenario, students cannot get any techno-academic support from them, and this would also raise a feeling of discontent among them. So, there is a dire need of making everybody technology comfortable in addition to the advancement in technology [46]. Some of the studies also reveal that many students are not comfortable in handling the technological aspects of computer-based instruments and thus their learning is hampered during the e-classroom sessions [1, 41].

Brooks et al. [14] have highlighted the psychological, physiological mental, and physical anxiety among the young learners based on e-classroom compulsory education as they are not comfortable in online mode of education and feel that offline mode is the only way for them to learn, understand, and grow. In many cases, the overloading of technological advancements has completely shattered their overall health [16, 19]. There are a lot many new ICT-based technological advances which just keep the students busy, coagulated. Many a times, they go in the anxiety or even depression mode. The examples of this could be the audio support, the video support, and various other attractive hardware instruments related to the computing technology. In addition to this, various applications for assignments tutorials and quizzes also puzzle them as all these things are novice for them. They just spend more time in learning, using or even misusing these ICT-based supporting systems [34]. But parents are bound to arrange these ICT-based hardware for them as it is the need of the hour during pandemics for e-classroom to continue smoothly. Many a times because of the over usage of these ICT-based technology for e-classroom as well as for other purposes, the students may even get bored with them. All this has led to the technological stress among our young learners who are the future of any country because the e-classroom is made compulsory for the students by the national policy makers, parents, and teachers. This is affecting their physical as well as mental health. Long stretches of e-classrooms is also affecting their body postures, body weight is getting increased, and many other ailments are also mushrooming up because of the restricted mobility [26, 43].

If students are compelled to so much technology-related stress, their outcome, performance, analyzing power, academics reduce significantly and the very purpose of continuation of education during e-classroom is diminished [12], (Dhir et al. 2019). During the pandemics, the compulsory continuation of education would be fruitful through e-classroom mode if the subjects are benefited to the utmost level. Parents are doing their best by arranging all the ICT-based items required for the ward to study without any interruption. [50]. But the learning satisfaction along with proper physical and mental development would come only if the subjects were grasping education not at the cost of their health [54]. The subjects also need to be satisfied with the compulsory e-classroom mode and make the most use out of it. They should be matured enough not to misuse the ICTs which their parents have arranged for education continuation. The compulsory e-classroom mode should improve their learning status and at the same time make them more confident and contented [13, 50]. But all this should not affect their physical and mental health.

A novel prediction model is proposed using enhanced model of machine learning, which is optimized at 2 levels, to predict how the long couch stretches during e-classroom is affecting their physical and mental health. This research work would act as an eye opener for all concerned so that necessary precautions could be taken during future pandemics (By the mercy of God, future pandemics should not occur). The model is novel in the sense that none of the researchers till now have proposed such an accurate and highly efficient prediction model using machine learning with two-step optimization.

2 Related work

Though a lot of contribution is made by the researchers to highlight the accomplishments of the e-classroom learning for the benefits of students. A handful of work is available which highlights the shortcomings of the e-classroom system where the adverse effects of this mode of education for the students are worked upon. In this work, it is shown that continuous stretches of sitting may retard the response time of the students, and this may also discourage them from participating in class and the feeling of loneliness and isolation can also arise among students and it may reduce their participation in various activities being done in the class [20, 22]. Another study highlights the requirement of a personalized student feedback system. As we all know that in offline mode of education, the teacher can instantly understand the various feedbacks related to students understanding but it is not possible in e-classroom education for students [32, 52]. It may also rise the feeling of dejection among the students as the field is isolated and lonely [20]. As we know that some of the students are so much habitual of interacting among each other in offline mode of education and they always try to build their own space among their peers that they feel quite isolated and rejected in online mode of education. This may lead to their social separation from the peers which may affect their learning and educational growth. They may also get physically and mentally disturbed by the isolation faced during online classes [46]. Many a times even the students are not self-motivated to turn up themselves happily for the e-classes, they are forced to take the online classes because of which many a times they either turn off their cameras or microphones or both. This lack of self-motivation is considered as one of the major reasons for the diminishing importance of e-classroom education during COVID-19. [24, 29]. Thus, if the policy makers make e-classroom concept compulsory, then it may also bring some amount of discontent among the illiterate cult of the country. There is no doubt that the whole world's technology has advanced a lot now a days. The major credit goes to advancements in the domain of human–computer interaction, but still there are a lot many people who are directly or indirectly related to the students like their parents, grandparents but are completely unaware of these technologies. In such a scenario, students cannot get any techno-academic support from them, and this would also raise a feeling of discontent among them. So, there is a dire need of making everybody technology comfortable in addition to the advancement in technology [46]. Some of the studies also reveal that many students are not comfortable in handling the technological aspects of computer-based instruments and thus their learning is hampered during the e-classroom sessions [1, 41].

Brooks et al. [14] have highlighted the psychological, physiological mental, and physical anxiety among the young learners based on e-classroom compulsory education as they are not comfortable in online mode of education and feel that offline mode is the only way for them to learn, understand, and grow. In many cases, the overloading of technological advancements has completely shattered their overall health [16, 19]. There are a lot many new ICT-based technological advances which just keep the students busy, coagulated. Many a times they go in the anxiety or even depression mode. The examples of this could be the audio support, the video support, and various other attractive hardware instruments related to the computing technology. In addition to this, various applications for assignments tutorials and quizzes also puzzle them as all these things are novice for them. They just spend more time in learning, using or even misusing these ICT-based supporting systems [34]. But parents are bound to arrange these ICT-based hardware for them as it is the need of the hour during pandemics for e-classroom to continue smoothly. Many a times because of the over usage of these ICT-based technology for e-classroom as well as for other purposes, the students may even get bored with them. All this has led to the technological stress among our young learners who are the future of any country because the e-classroom is made compulsory for the students by the national policy makers, parents, and teachers. This is affecting their physical as well as mental health. Long stretches of e-classrooms are also affecting their body postures, body weight is getting increased, and many other ailments are also mushrooming up because of the restricted mobility [26, 43].

All the work which has been done till now is summarized as below in Table 1.

If students are compelled to so much technology-related stress, their outcome, performance, analyzing power, and academics reduce significantly and the very purpose of continuation of education during e-classroom is diminished [12], (Dhir et al. 2019). During the pandemics, the compulsory continuation of education would be fruitful through e-classroom mode if the subjects are benefited to the utmost level. Parents are doing their best by arranging all the ICT-based items required for the ward to study without any interruption. [50]. But the learning satisfaction along with proper physical and mental development would come only if the subjects were grasping education not at the cost of their health [54]. The subjects also need to be satisfied with the compulsory e-classroom mode and make the most use out of it. They should be matured enough not to misuse the ICTs which their parents have arranged for education continuation. The compulsory e-classroom mode should improve their learning status and at the same time make them more confident and contented [13, 50]. But all this should not affect their physical and mental health.

2.1 Gaps in the Existing Work

As is evident quite clearly from the work which has been done till now on the adverse effects of COVID-19 pandemics on the health of students of all groups, almost all the studies revolved around theoretical concepts. The studies done in this domain are majorly revolving around proposing a set of strategies that could be implemented. The actual implementation in the form of tool, technique is missing. There is no empirical study to mitigate the adverse effects of enhanced ICT tools usage for budding students who are the future of any country. The progress of any country and its international place are decided by how well the young blooming generation are raised up in which their health plays a crucial role. All the work is based on researchers' illustrations based on adverse online learning effects on both physical and mental health of the students. There is no tool available based on empirical studies which can support the claims and/or suggest ways to deal with such type of situations in the near future. It is imperative for the countries to have such a tool which can deal with the future pandemics. This research work is a solution to the problems in dealing with the future pandemics. Forcing students and making e-education compulsory with stretched ICT usage has a lot many implications. The main gaps which are existing in the studies done by researchers till now are summarized as below. In the next section, we will discuss how this research work is able to fill the gaps which are existing in current studies.

Table 1 Work done till date

Work	Central idea of the work
[20]	Discussed regarding continuous stretches of sitting may retard the response time of the students and this may also discourage students from participating in class and the feeling of loneliness and isolation can also arise
[22]	Highlighted the side effects of continuous stretches of sitting which may reduce the academic response time of the students as if they are not physically fit, they cannot remain mentally fit also
[32, 52]	Highlighted the requirement of a personalized student feedback system as in offline mode of education, the teacher can instantly understand the various feedbacks related to students understanding but it is not possible in e-classroom education for students
[20]	Emphasized on the feeling of loneliness and isolation can also arise among young learners during online classes
[46]	Concluded that the social separation of students from their peers may affect their learning and educational growth. They may also get physically and mentally disturbed by the isolation faced during online classes
[1, 41]	Studies reveal that many students are not comfortable in handling the technological aspects of computer-based instruments, and thus their learning is hampered during the e-classroom sessions
[14]	Highlighted the psychological, physiological mental, and physical anxiety among the young learners based on e-classroom compulsory education as they are not comfortable in online mode of education and feel that offline mode is the only way for them to learn, understand, and grow
[16, 19]	Revealed that in many cases, the overloading of technological advancements has completely shattered students' overall health as there are a lot many new ICT-based technological advances which just keep the students busy and coagulated. Many a times they go in the anxiety or even depression mode
[26, 43]	Claimed that long stretches of e-classrooms are also affecting their body postures, body weight is getting increased, and many other ailments are also mushrooming up because of the restricted mobility
[12], Dhir et al. (2019)	Concluded that if students are compelled to use extensive technology then their stress outcome, performance, analyzing power, and academics reduce significantly and the very purpose of continuation of education during e-classroom is diminished
[54]	Claimed that the learning satisfaction along with proper physical and mental development would come only if the subjects were grasping education not at the cost of their health
[13, 50]	Highlighted that the subjects need to be satisfied with the compulsory e-classroom mode and make the most use out of it. They should be matured enough not to misuse the ICTs which their parents have arranged for education continuation. The compulsory e-classroom mode should improve their learning status and at the same time make them more confident and contented. But all this should not affect their physical and mental health

1. All studies are based on suggestions and findings.
2. Unavailability of empirical evidence available behind previous findings.
3. Unavailability of dataset being used in previous studies.
4. Unavailability of a precise and accurate tool or model which can predict the side effects of excessive usage of ICTs on students.
5. Unavailability of an automated feedback-based tool which can help combat the situation during future pandemics.

6. Unavailability of an insight for policy makers which can fix the maximum timings for e-education so that students' health is also not affected.

2.2 Main Findings, Implications, and Contributions of the Present Study

In this research work, a novel model is proposed to predict the aftereffects of long couch stretches during e-classrooms on the physical and mental health of the students. A variant which is highly accurate and computationally stronger than the already existing machine learning algorithm is used for this work in the first step of optimization. To improve the predictive efficiencies further, the model is optimized at the second step also using a hybridized bio-inspired algorithm. This research work would act as an eye opener for all concerned so that necessary precautions could be taken during future pandemics. The model is novel in the sense that none of the researchers till now have proposed such an accurate and highly efficient predictive model using two-step optimization techniques. The main findings and contributions of this work are summarized as below:

1. A novel two-step optimization-based tool to predict the side effects of excessive usage of ICTs on students.
2. Highly accurate and efficient two-step-based model optimized model with strong empirical support.
3. Usage of a humongous data set to obtain unbiased results.
4. Proposal and usage of an enhanced variant of a highly computationally strong machine learning model to obtain highly accurate results by the end of step 1.
5. Proposal and usage of a hybridized bio-inspired optimization algorithm to improve the results further by the end of step 2.
6. First work of its type which can act as an insight to deal with future pandemics.
7. Would help countries worldwide to take preventive measures to help us combat with the future pandemics.
8. Would act as an insight for the policy makers to define and limit e-education duration for young blooming kids to maintain a perfect balance between their health and education.

3 Statutory E-classroom During Pandemic

If we analyze the situation of current pandemic COVID-19 with the other pandemics in the world history ages back, then we see that the previous ones have led to zero academic years. The students were unable to continue their academia, their learning outcomes and academic excellence learning growth were all diminished. The major credit goes to the technological advancements that the parents of current pandemics can continue with the education of their wards with full support. But as see that the enhanced inclination and usage toward the e-classroom-related ICTs have led to elevated physical and mental stress among the subjects. So, parents are also bound to keep e-classroom continue with full support for their children, otherwise they will not be able to cope up with the academia with their peers.

3.1 Helpless Parents: Fight for Academia Existence and Success in Education

For the pandemic parents, the term "academic excellence" is not subtle. All the parents want their kids to excel in academia, be it at the cost of their own comfort. The parents prefer to arrange all the ICT tools and technology for their wards so that their education is not hindered and hampered during the pandemics time like COVID-19. None of the parents want their kids to not perform good in education during the pandemics. Though it is a big challenge for the home-restricted parents as well as their wards when the salaries of parents also get reduced or even zero in many cases. But it is the demand of the schools, colleges, and various higher educational institutions that academia should continue. Many times, the students are also overburdened due to the home isolation and confiscation restrictions, that the pressure among students during online classes is added on as compared to the offline one. One major reason for this is the "documentation" to ascertain various proofs which every school, college, and education imparting institution wants to keep as perfect during e-classroom education impartment mode. The pressure is more on the students to document everything regarding assignments, tutorials, deadline-based CBA sheets that it elevates the stress among the young learners. They start feeling that online classes' mode of education is enhancing their stress levels.

3.2 Elevate Learning Satisfaction and Learning Performance

All the parents, teachers, and the students themselves want to elevate their learning experience and performances during the e-classroom scenario. Because there is a lot of competition among the students now a days, everybody wants to excel in academics. Also, the parents and teachers want every student to be equivalent in academics and to excel in all the activities. The whole pressure comes to the physical and mental well-being of the students where a lot of expectations are on their way and it depends upon them whether they are able to fulfill all the types of requirements or not. Many cases have been seen during the pandemics where students were highly stressed out and they lost their efficiency of working as compared to offline classes' mode just because of the fear of gaining excellence in academic sphere.

3.3 Digital Inequalities

It is a bitter fact that money brings us closer to the digital world as the more money we have, the better is our affordability to explore, purchase, and use various digital technologies. A very dark side of e-classroom education is the availability of digital technology-based ICT tools to only few students who can afford it. There are many students who cannot afford each type of e-classroom-based ICT technology, be it a hardware or a software. This all brings a level of dissatisfaction among the non-afforders and it leads to a competitive sort of feeling among the students who were treated equally during the offline classes. There is no such disparity in the offline classes based on the various modes of digital technologies. All this adds on to stress regarding the physical and mental health of the students and, of course, it hampers their learning chances also. The better the internet connection, the better various other modes of tools attached with the laptops or the mobiles, the better would be the grasping power of the students. There are more chances of enhancement in learning and understanding if we have better digital technology supporting tools which are no-doubted expensive. This e-classroom scenario has led us to think about the side effects of economic disparity among the students' parents.

3.4 Technology Savvy and Non-savvy Parents

Although parents try to do their best to provide uninterrupted e-education-based ICT tools to their kids, many a times the technology savviness of the parents also hampers the uninterrupted e-classroom continuation of the students. All the parents are not very much comfortable with the new tools and technologies related to both hardware and software and it impacts the e-classroom continuation of their wards with respect to efficiency. The parents either try to gain techno knowledge from their colleagues or the students are completely deprived of the usage of new tools and technologies. But there is no such hindrance in the way of students learning if we talk about the offline mode of education. This also brings a sort of dissatisfaction among the students.

3.5 Bad Sitting Postures

The need of education through e-classroom demands the prolonged sitting of students at one position because that is the only way through which they can grasp the learning concepts, attempt assignments, quizzes, and tutorials. All this is affecting the body postures of the young budding students who are the future of any country to a great extent. There spinal cords are also getting aligned to the same position which is affecting their physical health and once the physical health is affected it also affects the mental health. Students are developing lot many ailments, injuries related to their backbone muscles and tissues. This is their age when they must be physically fit and fully energetic health wise to do their studies in an interrupted way but because of the bad postures and prolonged sittings during e-classrooms, they are unable to cope up with this in many situations. The nation builders need to device the methodologies through which the education during the future pandemics is not hampered as well as their sitting positions are also proper.

3.6 Semi-laid and Fully Laid Postures Which Highly Stresses Eyes and Headaches

The continuation of education throughout the pandemics is also affecting the backbone and the whole-body posture of the students. Students, getting in the comfort zone, many times do not sit on the table and the chair mode rather they prefer to sit on couches, or beds in a semi-laid or fully laid positions. All this effects their eyes as well as their whole-body postures. There are even seen fewer symptoms of cervical spondylitis and other spine related diseases. Even their eye sights are also getting affected because of all this. But they are bound to continue their education through e-classroom mode during the pandemics. The solution to this can be the strict adherence to the table and chair mode by the parents so that the students' postures are not getting affected. They also develop severe headaches because of wrong postures and eyesight-related issues.

3.7 Low Mobility Signs: Indigestion Issues

As the mobility of the students is getting restricted, they cannot go out and roam much frequently during the pandemics. They are restricted completely to their homes, so their body movements become reduced. In such a scenario, digestion of food is a big problem among the young as well as the little students. Whatever they eat is not digested properly as kids are also habitual of playing a lot but during pandemics, their all-physical movements are restricted. In such a scenario, their physical health is affected a lot and they can develop various major elements also if their food is not digested properly if it is for long durations and for recurrent pandemics. But because of the compulsory e-education scenario during pandemics, the students are bound to go for such a mode of education. This scenario was not there during the offline classes as when a student goes to school in morning, he gets up early, dresses himself, moves to his stop to board the bus or goes on foot to the school. So, a lot of physical activity is involved which is completely restricted during the pandemic time. So, their body which is tuned to digestion only after a heavy physical work is changed completely and indigestion-related issues start mushrooming.

4 Research Methodology

In this section, we discuss the methodology adopted to propose the novel model used in this research work. An enhanced and highly efficient variant of an already existing algorithm of machine learning called XGBoost is proposed and then used for this work at first step of optimization. It is optimized further to improve the predictive efficiencies by proposing and then using a hybridized bio-inspired algorithm. XGBoost being a computationally strong machine learning model requires a very high RAM, CPU, and GPU to run. Though, it takes some time to learn the feature variables through the training data and then make predictions on the test data. Hence, we use it only when all the other lightweight (requiring less RAM, CPU, and GPU) prediction models fail. The methodology used in this research work is shown in Fig. 1, with all the intermediate steps. First, the data of students are collected and arranged in the form of a dataset. Then these data are pre-processed to deal with missing values, outliers, and duplicate records. The data are also checked for imbalance problem and are balanced if they have vast differences in the number of values in the two classes. In such a scenario, the class with greater number of data values is called the majority class and the one having lesser number of attributes is called the minority class. It is followed by application of the proposed model which involves optimization at 2 levels.

This is then followed by generating the predictive results and then comparing the obtained values with the already existing models. Here, we tried to apply various other models like Random Forest, etc., but none of them is yielding satisfactory results. So, we proposed and then adopted the enhanced version of XGBoost model in the research work. A comparative analysis is also shown in Sect. 4 to ascertain the effectiveness of this model over other machine learning models. The comparative study also involves results of execution of the proposed model on various other datasets in addition to the COVID-19 student dataset.

4.1 Data Collection and Procedure

The data are taken from a vast dataset comprising 1182 students, all belonging to the academia in and around Delhi-NCR and from varied age groups. The vast set of data are analyzed to remove the impurities in the form of redundancies, missing values, dealing with imbalanced data values. A lot of machine learning researchers are working on the ways to find solutions for missing values. The best way found till now which does not affect the data much is to replace the values with their medians. This strategy is adopted by us also in this research work. If we see the descriptive statistics from Fig. 2 and Table 2 for the dataset, we realize that it is highly imbalanced. Another major problem is of data imbalance. One class contains proportionally higher values as compared to other class or classes. The artificial data points are synthesized or manufactured in the dataset by applying an oversampling technique. Generally, in classification problems such as this one, the data are highly imbalanced. In such a scenario of highly imbalanced data, the number of data points for one class is very high compared to another class. The class having the greatest number of data points is called the majority class whereas the class having the least number of data points is called the minority class. Al-Asadi and Tasdemír [5, 6]

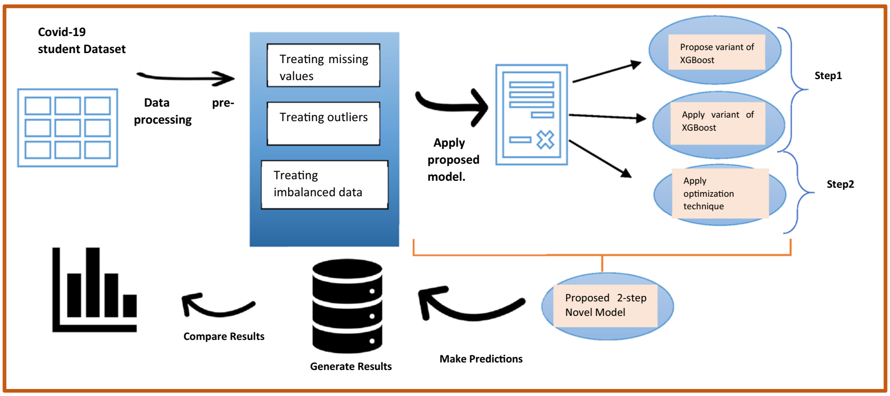


Fig. 1 Proposed methodology

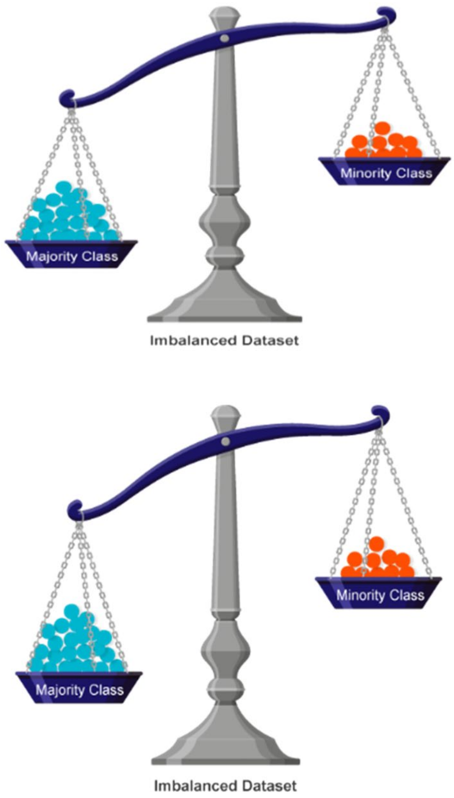


Fig. 2 Majority and minority lasses (imbalanced dataset)

Table 2 Descriptive statistics (COVID-19 student dataset)

	Age of subject	Time spent on online class	Time spent on self-study	Time spent on fitness	Time spent on online sleep	Time spent on social media	Number of meals per day
Count	1182	1182	1182	1182	1182	1182	1182
Mean	20.17	3.21	2.91	0.77	7.87	2.37	2.92
Std	5.52	2.10	2.14	0.72	1.62	1.77	0.83
Min	7.00	0.00	0.00	0.00	4.00	0.00	1.00
Max	59.00	10.00	18.00	5.00	15.00	10.00	8.00
25%	17	2	2	0	7	1	2
50%	20	3	2	1	8	2	3
75%	21	5	4	1	9	3	3

The major problem with imbalanced data is that a prediction model will always be biased in favor of the majority class in making predictions.

An oversampling technique synthesizes the artificial data points for the minority class data to balance a highly imbalanced dataset. An oversampling technique is required to remove the bias in favor of the majority class in a dataset. Hence, using an oversampling technique, we can artificially synthesize the minority class data in a training dataset so that both the classes have equal representation in the dataset. The oversampling technique is applied only to the training dataset. It is never applied to the test dataset.

Table 2 illustrates the descriptive statistics of the data, oversampling technique is applied only to the training COVID-19 student dataset. It is never applied to the test COVID-19 student dataset.

Here the 25%, 50%, and 75% are the quartile values which are called first, second, and third quartiles, respectively. The second quartile, i.e., 50% is the median value which divides the data series into 2 halves. Now, 25% quartile depicts the median value of the first half of the series and the 75% depicts the median value for the second half of the series. As one can infer from Table 2 that the data are imbalanced, we need to balance it using suitable techniques. If we see the outliers in the age of students, it is also remarkable. There is also a lot of variation in the time spent by students in online classes as well as for self-study mode. Table 2 also reveals that some of the students are very keen for physical activities during lockdown and have changed their habits, while some are still sluggish and prefer to remain confined to their "couches". Some of the data such as the subjects' age is distributed uniformly among all age groups (Fig. 3). The time which students have spent on online classes is of varied range (Fig. 4).

As we observe from Fig. 4 above, there are lot many outliers available in the COVID-19 student dataset. Also, the modes for taking online classes are also different—around 40%–40% opted for laptop/desktop and smartphones, respectively. Overall 5% opted for only tablet mode to attend online classes.

4.2 Treating the Missing Values

The data first need to be pre-processed to remove the impurities in the form of redundancies, missing values, dealing with imbalanced data values. A lot of machine learning researchers are working on the ways to find solutions for missing values. The best way found till now which does not affect the data much is to replace the values with their medians. This strategy is adopted by us also in this research work.

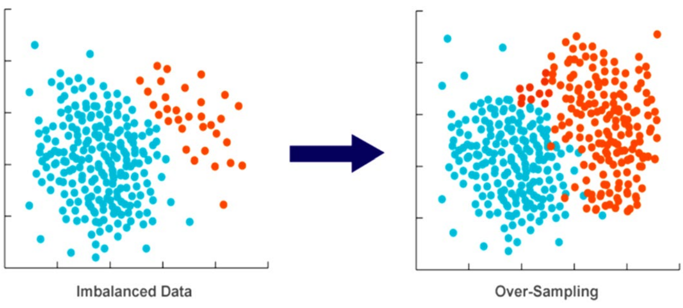


Fig. 3 Data balancing after over-sampling

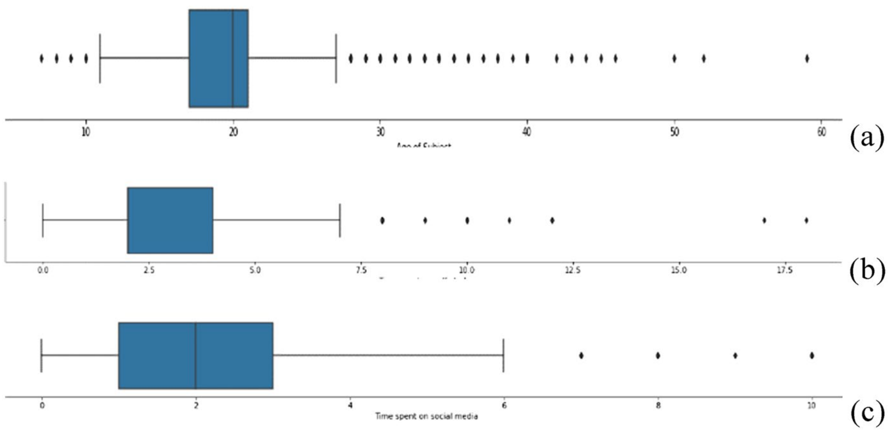


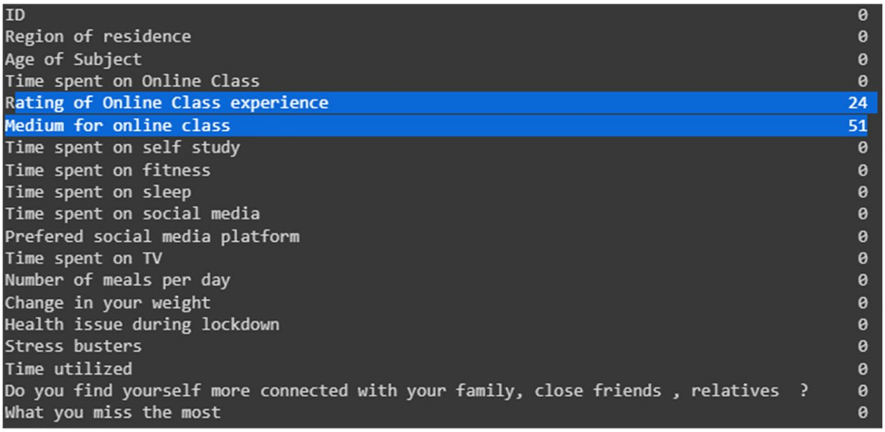
Fig. 4 Outliers in COVID-19 student dataset a: age of subject; b: time (self-study) C: time (social media)

First, we need to ascertain that the data contain missing values. The missing values are generally reported as NaN values. We can find the rows or columns containing the NaN values using either the isnull() or the isna() function. They both return True for the NaN (or null or missing) values. If we check the missing values using Fig, we observe that only 2 columns named "Rating of Online Class experience" and "Medium for online class" contain 24 and 51 missing values, respectively.

Since these two columns are non-numeric, it is not possible to convert them to median values. So, we replaced them with the majority values at respective columns (Fig. 5).

4.3 Recursive Feature Elimination (RFE)

Recursive feature elimination (RFE) is a feature selection (or elimination) method that fits a model and removes the weakest feature (or features) which are not of more importance for us. Here, one first needs to decide the number of features he wants to select to build the model. Then one can validate his choice of number of features and increase or decrease them (if required). Features are ranked by the model's significantly important attributes, and by recursively eliminating a small number of features per loop, RFE attempts to eliminate dependencies and collinearity that may exist in a machine learning model. RFE requires a specified number of features to keep; however, it is often not known in advance how many features are valid. In this work, we have selected the best 10 features out of the highly correlated features that are stored in the dictionary. After applying RFE to our model, we are able to eliminate the least significant attributes "Number of meals per day" and "Time spent on social media" for the COVID-19 student dataset by following repetitive iterative cycles. Thus, we are subsequently selecting the important attributes "Age of Subject", "Time spent on Online Class", "Time spent on self-study', "Time spent on fitness", "Time spent on Online sleep" and "Number of meals per day".



ID	0
Region of residence	0
Age of Subject	0
Time spent on Online Class	0
Rating of Online Class experience	24
Medium for online class	51
Time spent on self study	0
Time spent on fitness	0
Time spent on sleep	0
Time spent on social media	0
Prefered social media platform	0
Time spent on TV	0
Number of meals per day	0
Change in your weight	0
Health issue during lockdown	0
Stress busters	0
Time utilized	0
Do you find yourself more connected with your family, close friends , relatives ?	0
What you miss the most	0

Fig. 5 Missing data values

4.4 Oversampling for Classification Problems—SMOTE

There are 3 different methods to synthesize the artificial data points for a classification problem. They are random oversampling, SMOTE, ADASYN. We have applied the SMOTE method to synthesize the artificial data points in the training dataset. The term SMOTE stands for Synthetic Minority Over-Sampling Technique. The SMOTE method is the easiest one to understand and here in this work, it proves to be better than other methods [10, 17, 18]. As our data are imbalanced, we have used SMOTE for balancing our data across all the classes. Those data values are used to train different models for prediction. As is evident from Fig. 6, SMOTE works as follows:

1. A random data point or an instance from the minority class (a green circle) is first chosen.
2. Then the nearest neighbors for that instance are found (usually 5 neighbors)
3. One of the neighbors of that instance is randomly chosen and an artificial data point (red circle) is generated between that neighbor and the instance.
4. This process keeps on repeating to create the required number of synthetic data points for the minority class.

5 Implementation Details Using the Two-Step Model

The work proposes a novel model which functions at two levels. At the first level, an enhanced variant of XGBoost algorithm is proposed. At the second level, this variant is optimized further using a bio-inspired algorithm. Thus, the efficiency of the proposed model is improved at both the levels. In this way, we are able to obtain a highly efficient and accurate model at the end of both the levels. Both the levels

Fig. 6 Resampled COVID-19 student dataset

enhance the predictive accuracies to a great extent, thus obtaining a highly efficient and accurate model at the end of the two steps of the proposed model. Both the levels with the algorithms being used are explained below:

5.1 Step 1 Optimization: Using Variant of XGBoost Model

The original XGBoost algorithm is quite computationally strong and is expected to yield much better predictive results as compared to its counterparts. But since it deals with a lot many parametric quantities, so adjusting all these parameters together for the model has always been a challenge among the researcher community. The overall time consumed increases; thus, the predictive power is affected by enhanced time duration. The efficiency may also get affected in adjusting a large number of parameters. Thus, there has always been a need to improve the model to provide best results. Many works have been proposed to combat with this situation by the researchers and improve the already improved model to have best combination of results [25]. The improvement methods are being suggested by incorporating grid search, parametric optimizations, hyperparametric optimizations (Bayesian based), and cross-validation-based techniques. But none of them are able to enhance the existing shortcomings significantly. Thus, to improve the XGBoost algorithm further, we propose to hybridize it with “Differential Evolution method”. It is an evolutionary algorithm which uses the probability model to relate the COVID-19 student dataset values directly with its output results. The objectivity function of the algorithm is linked with its labeled attributes by “n” number of folds per process. The parameters are hybridized further into various sub-parameters to obtain the optimal parameters at every step. Probability function is applied at every level to ascertain the validity of various step-wise results. It is a novel concept which is proposed and used by us as the results being generated are enhancing the predictive accuracy and reducing the prediction time of the model. Thus, it is being used in step 1 as the first level of optimization. The objectivity function is applied in such a way that it can quantify the values so as to generalize the model for any ML algorithm and any COVID-19 student dataset with the sole aim of enhancing the predictive accuracies. The Gaussian functionality is adopted to distribute hypothesis related to process which is followed by acquisition process for determining the frontal points which need evaluation as they are encountered iteratively during the whole cycle.

5.1.1 Proposed Variant of XGBoost Algorithm

The Gaussian functionality is adopted to distribute hypothesis related to process which is followed by acquisition process for determining the frontal points which need evaluation as they are encountered iteratively during the whole cycle.

The Gaussian function is depicted and discussed as below:

$$f(z) :\ : FP(me(z), Q(z1, z2)) \tag{1}$$

$$Q(z1, z2) = \exp\left(-\frac{1}{2}(|z1 - z2|pow2)\right) \tag{2}$$

Here $me(z)$ depicts the mean functionality value, $Q(z1, z2)$ depicts the mean squared error functionality. Assuming that we have a dataset having "t" values in train dataset. The Gaussian distribution for this scenario is done assuming that all the samples to be trained belong to the distribution functionality. To obtain the target attributes, Gaussian function builds covariance matrix and obtains a fresh additional dimension distribution iteratively in every level. In the iteration process of XGBoost algorithm, Bayesian optimization uses the acquisition function to obtain the global optimal solution of the algorithm. First, the unidentified points are evaluated to avoid the local optimal solution. Second, the global optimal solution can be obtained using multiple acquisition near the current solution. By maximizing the acquisition function, the next point to be evaluated can be selected, thus reducing the total number of iterations, and improving the efficiency of the algorithm. On the basis of Bayesian optimization method, the proposed algorithm finds out the optimal attributes. In this way, the improved variant of XGBoost model continuously trains itself using the Bayesian-based optimized methodology. Gaussian method distributes iterations over the original XGBoost method, giving birth to its variant, thus enhancing its predictive capabilities. During iterative processing, best attributes based on topmost scores are selected iteratively, thus providing the XGBoost model having maximum score value. Thus, Bayesian optimization enhances predictive efficiency and accuracy of the model. The algorithm (algorithm1) works as follows:

1. Initialize all data points according to the output values and construct the database $Q'\left(x', y'\right)$.
2. Construct the regression model using Gaussian method for the database $Q'\left(x', y'\right)$.
3. Apply accession function for obtaining subsequent accession points as $\left(x^{'} + 1\right)$.
4. Perform evaluation based on $\left(x^{'} + 1\right)$andkeeponaddingpoints$\left\{\left(x^{'} + 1\right), \left(y^{'} + 1\right)\right\}$ to the database $Q'\left(x', y'\right)$ and thus build a fresh database in every iteration.
5. Finally, obtain optimized results.

5.2 Step 2 Optimization: Using Bio-inspired Algorithm

Most of the techniques employed to optimize the weighted measure in the previous research performed by various researchers revolve around improving the XGBoost model by applying various improvement techniques to it. None of the researchers have adopted a highly efficient heuristic-based whale optimization algorithm with elevated optimized values. To enhance the performance of the measure to propose the variant of the specified model, the traditional whale optimization algorithm is implanted with Hill-climbing algorithm. This implantation helps in exploiting the advantageous domains of both the concepts by switching between them as per the requirements. We work on two vector operators X and E (one from each algorithm), to discover new areas near each whale to explore and exploit the best prey catching possibility domains. As the Hill climbing works well with local search, so its vector variable will be used during the process to search for the most optimal local and another variable will continue searching till the end. This also helps in avoiding the early convergence problem, thus allowing the implanted algorithm to continue exploring and exploiting all the possible combinations. Using the Hill-climbing concept, the superior most value for forward and/or backward movement is obtained, thus eliminating the inferior ones. The proposed implantation algorithm is applied on various datasets which confirm its enhanced optimization $\in$.

Optimization Algorithm. The hunting methodology of whales is used for the formulation of whale optimization algorithm. Whales use the bubble-net methodology comprising flinching, neighboring, and updating spiral positions to catch their preys successfully without fail. Whales neighbor their prey by constantly moving in the most optimized direction. Local search parametric value using Hill climbing for the proposed variant of whale algorithm helps them to find the best values to surround the prey. Now the second controlling vector helps in controlling the exact attacking position. During the complete attacking position, a controlling vector attribute ca is constantly decreased, thus helping in catching the prey without miss. This controlling vector lies in the range $[-ca, ca]$. The exploitation level starts with capricious points $p = \left(p_1, p_2, \dots p_n\right)$ where n is the total number of capricious points. Two vectorized parametric values X and E are taken to find the direction to search for fresh prey and guarantee local search optima for current prey. The optimal parameter “q” choosing nearest point “p” is represented as $p_i = p_i + -$ B (0, 1)* bandwidth, i $\in [1, 2, \dots, n]$. The implantation of Hill climbing with whale optimization algorithm enriches the exploration and enhances the shifting between the two as and when required to optimize the two vector variables. These variables apply on every whale to explore new areas in the vicinity of whales for exploitation and exploration purpose. Following this, it helps whales in retaining the superior most downhill values, thus neglecting the inferior ones. In this manner, the two vectorized operators help in renewing the spatial position, thus finally reaching to the optimized solution. The algorithm for the implanted modified whale optimization algorithm (M’WOA) is discussed as below:

Algorithm to compute the modified whale optimization algorithm (M’WOA).

Initialize:
Parameters X, E, Bandwidth, max number of iterations denoted as n
Whales' positions p_i
Exploration agents' fitness value
p_I *as the best solution*
Output:
Final Score after Optimization
Begin
While (t<T)
Renew operators
For every whale
If p< 0.5
if | A | > 1
If e<0.5
Search for a leading value $p_{methodical}$
Update the position using $\overrightarrow{P(t+1)} = \overrightarrow{P_{methodical}} - \vec{A} \cdot \vec{D}$
Else If e>0.5
Renew positions using $p_i\,(t+1) = p_i\,(t) + 0.001\, G(p_i\,(t) - lower\ bound,\ upper\ bound - p_i\,(t) + FF_1R_1 + FF_2R_2)$
end if
Else if | A | <1
If r < 0.5
Update the positions using $\overrightarrow{P(t+1)} = \overrightarrow{P(t)} - \vec{A} \cdot \vec{D}$
Else if r > 0.5
Compute the parametric value using $p_i = p_i \pm U(0,1) * bandwidth * (2 * r - 1), i \in [1,2, \ldots N]$
Compute the E-parameter
Compute the X-parameter
End if
If p> 0.5
Renew positions using $\overrightarrow{P(t+1)} = \overrightarrow{D * cos(2\pi l)} + \overrightarrow{p_t(t)}$
End If
End for
Check for spatial constraints
Update $p_{methodical}$
t=t+1
End While
Revisit $p_{methodical}$

6 Data Analysis and Results: Evaluating the Proposed Model

The data from the COVID-19 student dataset is split into training and test dataset. Around 67% values are assigned to training dataset and remaining 33% to test dataset. A highly efficient algorithm of machine learning is proposed which is optimized at 2 steps, In the first step, a variant of already existing XGBoost is proposed and used for this purpose and in step 2, a bio-inspired algorithm which is being hybridized with another heuristic technique is proposed. Thus, the proposed model in this research work handles optimization at two steps, thus enhancing the predictive accuracy and efficiency of the model to a great extent. XGBoost being a computationally strong machine learning model, requires a very high RAM, CPU, and GPU to run. It takes some time to learn the feature variables through the training data and

```
array([2, 2, 1, 2, 0, 1, 1, 0, 0, 2, 0, 1, 0, 1, 1, 2, 1, 0, 0, 2, 2, 2,
       2, 2, 0, 2, 2, 2, 1, 2, 1, 0, 1, 1, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2,
       0, 1, 2, 2, 2, 2, 0, 2, 2, 0, 2, 0, 2, 0, 1, 2, 2, 0, 1, 1, 2, 2,
       1, 2, 2, 1, 0, 2, 1, 2, 2, 2, 0, 0, 2, 1, 2, 1, 1, 1, 1, 1, 2, 0,
       2, 0, 1, 0, 0, 0, 2, 1, 0, 0, 0, 0, 2, 1, 1, 0, 1, 1, 1, 1, 0, 0,
       1, 1, 1, 0, 2, 1, 2, 2, 1, 2, 2, 1, 1, 0, 0, 2, 0, 1, 1, 1, 1, 2,
       1, 2, 2, 0, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 1, 1, 0, 2, 1, 1, 1,
       2, 0, 0, 1, 1, 0, 1, 0, 0, 1, 2, 0, 1, 0, 0, 0, 0, 2, 2, 0, 1, 0,
       2, 2, 2, 0, 0, 2, 2, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 2, 2, 0, 0,
       0, 0, 2, 1, 2, 0, 0, 0, 0, 1, 0, 2, 2, 0, 2, 0, 1, 2, 2, 2, 1, 0,
       0, 0, 1, 2, 0, 1, 2, 1, 1, 1, 0, 2, 2, 0, 1, 0, 1, 1, 2, 0, 1, 0,
       2, 0, 2, 2, 0, 0, 2, 0, 1, 1, 2, 2, 2, 2, 0, 1, 2, 0, 0, 2, 1, 0,
       1, 0, 0, 1, 2, 2, 0, 2, 2, 0, 0, 1, 2, 0, 0, 0, 2, 0, 0, 1, 2, 1,
       1, 0, 1, 2, 1, 0, 2, 2, 2, 2, 1, 2, 2, 0, 1, 1, 2, 2, 0, 1, 1, 0,
       2, 2, 0, 1, 0, 2, 0, 2, 1, 2, 0, 1, 0, 0, 0, 1, 2, 2, 2, 0, 2, 1])
```

Fig. 7 Predicted values

then make predictions on the test data. Hence, its usage is preferred only if all the other lightweight (requiring less RAM, CPU, and GPU) prediction models do not generate satisfactory results. Here, we tried to apply various other models like Random Forest, etc., but none of them is yielding satisfactory results. So, we adopted the XGBoost model in our proposed methodology. A comparative analysis is also shown in Sect. 4 to ascertain the effectiveness of this model over others. From the prediction of the model, we observe that the values fit into the XGBoost model as depicted by Fig. 7.

From Fig. 7 we evaluate the values predicted using our model. Here we have a multi-classification model using values 0,1,2 which signify the health conditions during the pandemics as stressed, relaxed, or neutral. The above Fig. 7 can also be converted into a matrix form. All these data are highly important to evaluate the performance of the model. As is established empirically that the parametric values like accuracy of the model comes out to be 98.07%. Also, we notice that the sensitivity, specificity, and F1-score have attained values 96.74%, 95.63%, 98.43%, respectively, which ascertain the high predictive performance of our model on the basis of these parameters. This higher predictive parameters results achievement using our model, as compared to the other models used in literature till now, is possible only using the novel approach suggested in this research paper. All these parameters are discussed in detail in the following section. The empirical results are also established IN Table 3.

6.1 Statistical-Based Evaluative Measures

The various statistical measures which are used to evaluate the proposed model are based on major 4 values, truly_productive, falsely_productive, truly_non_productive, and falsely_non_productive. Truly_productive refers to the number of circumstances which are present in the test data set and are also being discovered positively. Falsely_productive refers to the number of circumstances which are not present and

Table 3 Empirical results using proposed model (after 2-step optimization)

Base model	Specificity (%)	Sensitivity (%)	PPV (%)	NPV (%)	Accuracy (%)± *SD*	F1-score (%)
Proposed model	95.63	96.74	97.77	96.34	98.07 ± 0.02	98.43

are discovered positively. Similarly, Truly_non_productive implies to those circumstances which are present in test dataset but are being discovered negatively. Falsely_non_productive refers to those circumstances which are neither present nor discovered positively, i.e., they are discovered negatively. The influential statistical evaluative measures named accuracy, sensitivity, specificity, F1-score, productively predicted values, and non-productively predicted values which are based on above four values.

6.1.1 Sensitivity (Recall)

The sensitivity for any outcome refers to its ability of classifying samples which depict a particular condition. It measures the extent to which truly productive values are identified. Sensitivity is also termed as recall, hit rate, or truly productive rate. It is equivalent to the proportion of truly productive from the set of all those having the conditions (truly productives and falsely non-productives).

Sensitivity = Truly_productive / (Truly_productive + Falsely_non_productive) = Probability of correctly predicting the values.

6.1.2 Specificity

Specificity identifies how perfectly prediction system picks truly non-productives. It is also termed as selectivity or truly non-productive rate, which refers to percentage of truly non-productive values from the bunch of all those values which do not possess any circumstance. Sensitivity and specificity terms were propounded by US-based American bio-statistics J. Yerushalmy in year 1947.

Specificity = Truly_non_productive / (Falsely_productive + Truly_non_productive).

6.1.3 Productively_Predicted Values and Non_Productively_Predicted Values

Non-Productively_Predicted Values = Truly_non_productive / (Falsely_non_productive + Truly_non_productive).

6.1.4 Accuracy

It measures the fraction of flawlessly predicted values to the total count of values against whom the predictions are made. It refers to the flawless rate of predictions.

For a particular dataset, accuracy is quantity of correct or precise predictive count over total count for samples of complete data set.

Accuracy = Truly_productive + Truly_non_productive / (Truly_productive + Falsely_productive + Truly_non_productive + Falsely_non_productive).

6.1.5 F1-Score

The F1-score (F score/F measure) is an evaluation of the model's ability to predict the values which match accurately. It refers to the balance between the precision and recall values as it depicts the weighted harmonic mean of precision and recall. Our desire for a good prediction model is to have higher F1-score which can be achieved by higher values of both precision and recall. F1-score emphasizes on improving both precision and recall for better prediction results.

It is calculated as 2*((precision*recall)/ (precision + recall)).

7 Research Implications and Explanation of Findings

From the results in Table 3 and their diagrammatic representation in Fig. 8, one can conclude that the proposed model is highly accurate and efficient with enhanced predictive power. This research study is able to empirically establish and evaluate a prediction model with high predictive efficiency. The proposed model is highly accurate with 98.07% accuracy level with 0.02 value of standard deviation, which ascertains the acceptability of the proposed model in this research work. Three other machine learning models Decision Tree, Support Vector Machine, and Random Forest model were also deployed on the same dataset and the results are shown in Tables 3,4 and Fig. 8. We have been able to obtain the highest value of the most important parameter accuracy with least value of standard deviation using the proposed model as contrary to the other models which are generating 72.83% (with 6.09 value of

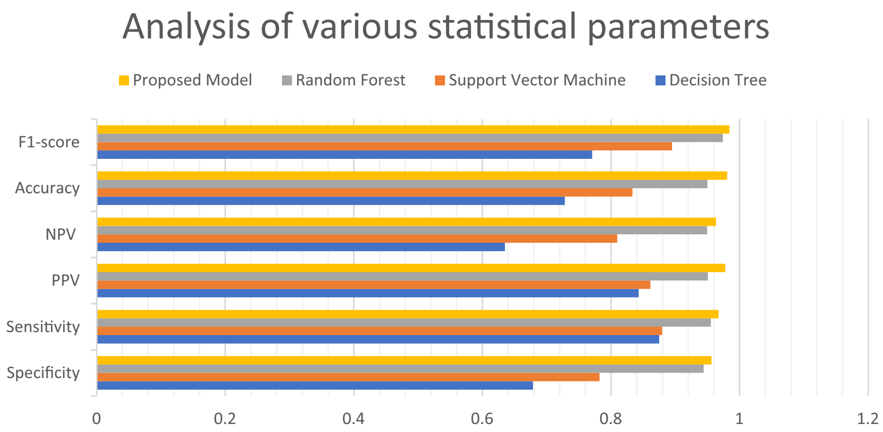


Fig. 8 Comparison of parameters using various models

Table 4 Empirical results

Base models	Specificity (%)	Sensitivity (%)	PPV (%)	NPV (%)	Accuracy (%)±*SD*	F1-score (%)
Decision Tree	67.87	87.50	84.30	63.50	72.83 ±6.09	77.08
Support Vector Machine	78.22	87.96	86.13	80.98	83.33 ±4.62	89.49
Random Forest	94.41	95.54	95.07	94.94	95.00 ±1.97	97.40
Proposed model (2-step optimization)	95.63	96.74	97.77	96.34	98.07 ±0.02	98.43

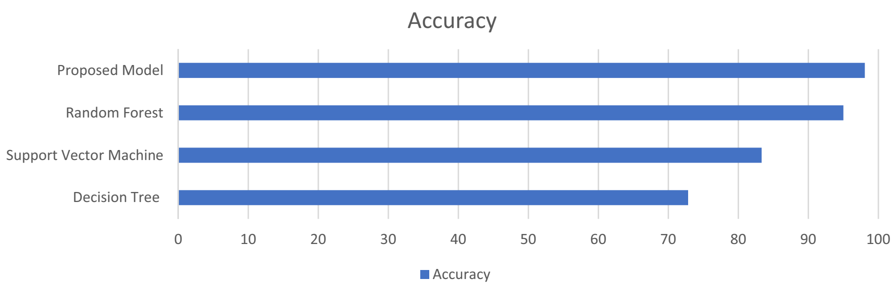


Fig. 9 Comparison of accuracy using various models

standard_deviation), 83.33% (with 4.62 value of standard_deviation) and 95% (with 1.97 value of standard_deviation) as depicted using bar graph in Fig. 9. The reported results are the average test accuracies for various methods with their corresponding standard_deviation values with fivefold cross-validation method for student dataset. The most perfect values are obtained with higher accuracy values and lower standard deviation values. The same is obtained using our proposed model. The two-tailed paired *t* test is applied with 5% significance level. The specificity and the sensitivity values are also obtained highest with the proposed model as 95.63% and 96.74%, respectively. The other models on the same dataset are able to obtain comparatively lesser values for these parameters also as is depicted in Fig. 8. complete paradigm shift has been observed. The students ranging from class nursery grade to higher educational institutions are able to continue their academics without interruptions. This all is because of the e-classroom-based educational system. The students shifted completely to the online system of education without latest ICT-based tools and technologies. The academia domain has completely drifted to online mode. The traditional mode of offline education is not possible during the pandemics. Some of the students are able to cope up with the new transformed mode while few are still hesitant to adopt the new technology. But e-classroom mode is made compulsory by the current educational system. It is the only means to continue education using pandemics. It compels students to drift to the new mode even if the young learners are not very much comfortable in technical know-how of the new ICT-based tools. They are compelled to adopt to new mode of education digitally which is affecting their health adversely. The novel model which we have proposed using a computationally strong XGBoost prediction model is highly accurate and efficient with promising values for sensitivity, specificity, F1-score, and various other parameters. Our study is not limited to only a particular age group depicting only school goers or college goers, rather it is a mix of subjects. Their ages are ranging from a 7-year school going kid to a 59-year-old person for whom age is no bar to study out of a group of 1182 subjects. Thus, there are minimum chances of result biasness. But out majority of age group involves teenagers as is evident from Fig. 10.

If we see we can see the whole subject population and check how many of them developed health issues during and after the e-classroom education, then from Fig. 11, we conclude that more than 13% of the young budding learners (160) are affected by remarkable amount of health issue—be it physical or mental. This number would have been so lesser if offline mode of education is followed.

Further, if we notice from Fig. 12, the changes in body weights of the subjects, then around 37% reported increment in weights, which is a huge percentage. This data not only reflects the health of these kids, rather it reflects the health status of the “future” of any nation. There is only a very small percentage near about 18% whose weights are decreased. Around 45% student population is there which does not report any change in the body weights. The credit goes mainly to the couch confined kids.

8 Comparison with Other Studies

A comparative analysis of the proposed model with other existing models is made using Fig. 13. It can be clearly concluded that the proposed model is highly accurate and efficient with enhanced predictive power as compared to the other models. The

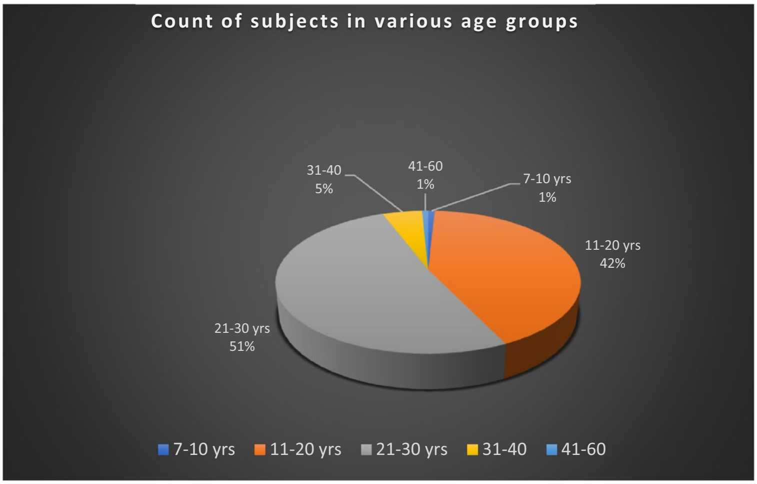


Fig. 10 Distribution of students’ age groups

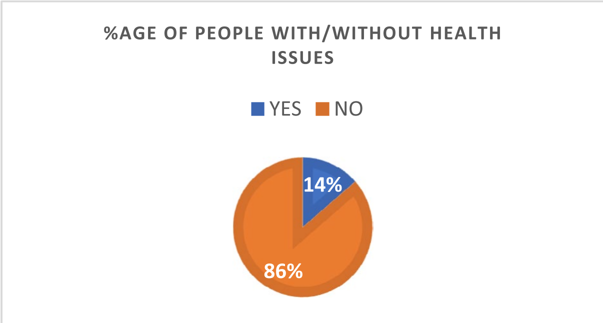


Fig. 11 Health issues during lockdown

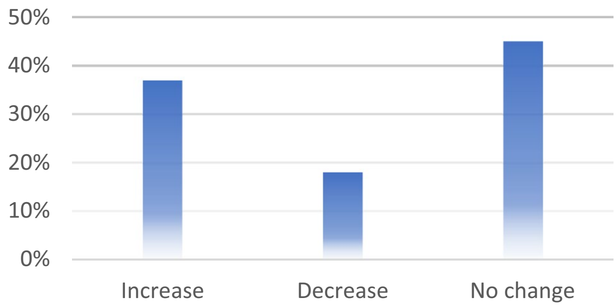


Fig. 12 Weight changes during pandemics

proposed model is highly accurate with 98.07% accuracy level and with 0.02 value of standard deviation, which ascertains the acceptability of the proposed model in this research work. Three other machine learning models Decision Tree, Support Vector Machine, and Random Forest model were also deployed on the same dataset and the results are shown in Tables 3, 4 and Fig. 13. We have been able to obtain the highest value of the most important parameter accuracy with least value of standard deviation using the proposed model as contrary to the other models which are generating 72.83% (with 6.09 value of standard_deviation), 83.33% (with 4.62 value of standard_deviation), and 95% (with 1.97 value of standard_deviation) as depicted using bar graph in Fig. 13. The reported results are the average test accuracies for various methods with their corresponding standard_deviation values with fivefold cross-validation method for student dataset. The most perfect values are obtained with higher accuracy values and lower standard deviation values. The same is obtained using our proposed model. The two-tailed paired *t* test is applied with 5% significance level. The other models on the same dataset are able to obtain comparatively lesser values for these parameters also as is depicted in Fig. 13.

As we can conclude from Fig. 13 that if we use other machine models like Random Forest, Support vector machine, and Decision tree, then the results obtained for various parameters are not so promising. As is clearly evident that the accuracy is highest in case of the proposed model. If we use Random Forest, then it is 95%

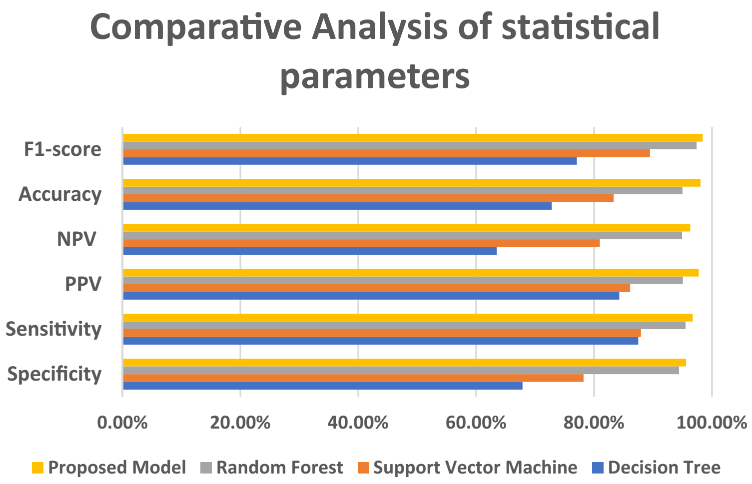


Fig. 13 Comparative analysis of various models with proposed model

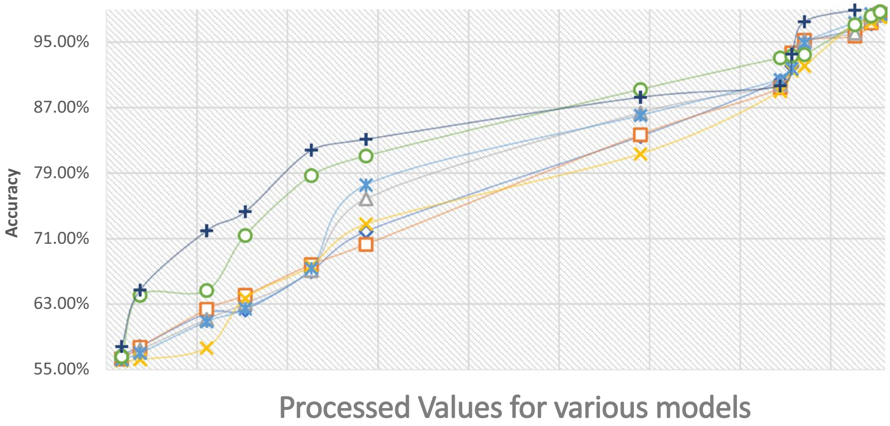


Fig. 14 Accuracy variation for various models

which is also a fair value but much lesser than that obtained using the proposed model.

The Fig. 14 gives a clear-cut depiction of how the accuracy values are being obtained for various models, out of which that one obtained for our proposed model is the highest. If we investigate other parameters like specificity, sensitivity, F1-score after using Random Forest model, then the values are lesser than the proposed model. The same concept is also applicable for other machine learning models. The precision, recall, and F1-score have attained highest values as compared to other models, which ascertains the high predictive performance of the proposed model in this research work. Now, if we concentrate on Fig. 15, then it

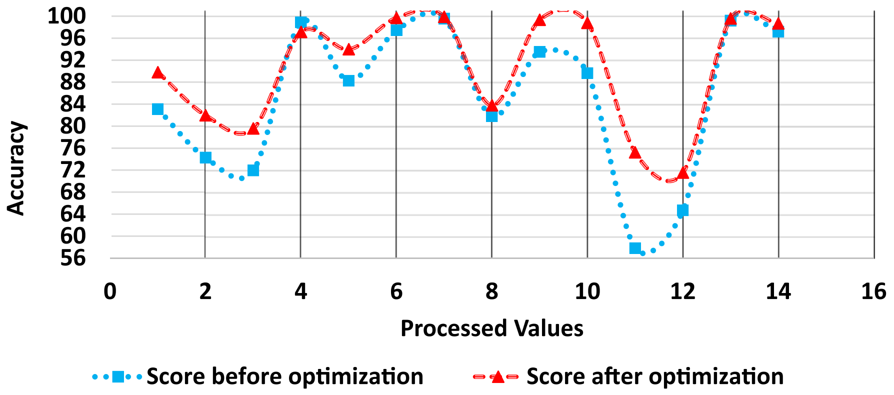


Fig. 15 Accuracy before and after optimization

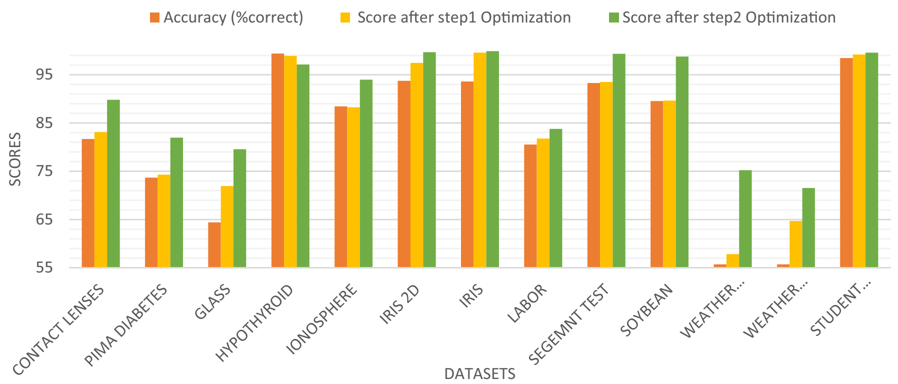


Fig. 16 Accuracy after step 1, step 2 for various datasets

can be concluded that using a two-step optimization, we are able to obtain significantly higher values of accuracy. Since accuracy plays the most important role as a parameter, it can be clearly seen from figure that elevated accuracies at the end of first step of optimization are elevated further using second step of optimization. This is the reason behind proposing such a two-step-based optimization tool.

As is shown in Fig. 16, a comparative analysis of the proposed model is made for additional datasets for accuracy score, optimized scores after step 1 and step 2 of optimizations. Figure 16 shows that the last dataset is the same COVID-19 dataset which we are using in this research work. Some additional datasets are used and are tried to be optimized using the proposed algorithm at two steps. These datasets are mentioned in the figure above. As it can be concluded that the step 1 is generating better accuracy results which are improvized further at step 2 for every dataset. This empirical study also ascertains the overall enhanced predictive accuracies generated using the proposed model. As is indicative from Fig. 16 that for every dataset, the accuracies are elevated from step 1 to step 2 of the optimization process. The time complexity obtained is also of the order of "n".

9 Conclusion and Future Work

The model proposed in this research work uses advanced and computationally strong machine learning algorithm XGBoost to predict the impact of enhanced students' computer interactions on their physical and mental health. There is an inevitable need to have such a pioneering tool which can provide an insight to any country's policy makers to combat with the above said condition. The work considers a humongous dataset with varied age groups of students with more than 10 attributes. The proposed model is highly efficient in making predictions with 98.07% accuracy level, 0.02 value for standard deviation, and 98.43% F1-score. Strong empirical results for other parameters also like specificity (95.63%) and sensitivity (96.74%) ascertain the enhanced predictive power generated using the proposed model. The time complexity obtained is also of the order of "n". Till now, none of the researchers have proposed any such pioneering tool for parents, doctors, and academicians using advanced machine learning algorithms.

In this study, the side effects of e-education, which marked a complete paradigm shift in academia during current pandemics, on the physical and mental health of students are analyzed in great depth. The students belonged to various age groups ranging from a 7-year-old kid to a 59-year-old student who is pursuing education with job. It is predicted that the e-education led a great impact on the health of students. There are gripped under various types of ailments following compulsory e-classroom mode. The bitter fact is that it is the only way through which they can continue their education. Without this educational mode, the students will not be able cope up with their academic syllabus. This may result in the zero-year policy which would be a great loss for the students of any nation. But it has been observed that because of the long stretches of sitting for e-classroom education purpose, the students are getting so many health ailments. In some of the countries, even the government norms regarding the duration of e-classroom education is also violated. The only purpose of the schools, parents, teachers, and various other academic institutions is to help students cater with their educational needs during the pandemics when the students are home restricted. But over-burdening the young learners, who are the future of any nation which is affecting their health in both physical and mental way, is also not a wise decision. Thus, there is an urge to generate such types of methodologies during the future pandemics, which should not re-occur by the grace of God, that the students can continue their e-education as well as their health is also taken care of. One solution can be to reduce the hours of compulsory e-educational mode and introduce certain compulsory physical activities sessions like yoga or various other body stretching exercises. Academia should also plan to the duration of such types of activities so that the classroom education is compensated with the physical exercise education as well. Otherwise, the healthy future of these young learners will be diminished. One limitation of this current research work is we are unable to propose any tool which can reduce the physical and mental stress of the students during the compulsory e-classroom education during the time of pandemics. Researchers in consultation with the medical practitioners can further work in this area and develop such tools which can compensate the long stretches in which students are table, chair or couch confined and some compulsory physical education sessions like yoga, etc., should be involved with the academic sessions.

To summarize, the main strengths of the proposed model and work are as follows:

1. A novel two-step optimization-based tool to predict the side effects of excessive usage of ICTs on students.
2. Highly accurate and efficient two-step-based model optimized model with strong empirical support.
3. Usage of a humongous data set to obtain unbiased results.
4. Proposal and usage of an enhanced variant of a highly computationally strong machine learning model to obtain highly accurate results by the end of step 1.
5. Proposal and usage of a hybridized bio-inspired optimization algorithm to improve the results further by the end of step 2.
6. First work of its type which can act as an insight to deal with future pandemics.
7. Would help countries worldwide to take preventive measures to help us combat with the future pandemics.
8. Would act as an insight for the policy makers to define and limit e-education duration for young blooming kids to maintain a perfect balance between their health and education.

As already mentioned above, the proposed model bears certain shortcomings also which can be improved in the future and are summarized as follows:

1. Inclusion of additional attributes for the robustness check of the model. Currently, we have used only one attribute recursive feature elimination (RFE) for this purpose.
2. We have just been able to propose a tool to predict the aftereffects of compulsory e-classroom education during the time of pandemics.
3. But we are unable to propose any tool which can reduce the physical and mental stress of the students during the compulsory e-classroom education during the time of pandemics.
4. Researchers in consultation with the medical practitioners can further work in this area and develop such tools which can compensate the long stretches in which students are table, chair or couch confined and some compulsory physical education sessions like yoga, etc., should be involved with the academic sessions.

Funding This work is not funded by any funding agencies.

Data availability Data available on demand.

Declaration

Ethical standards No humans and/or animals are involved in this study.

References

1. Abdous, M. (2019). Influence of satisfaction and preparedness on online students' feelings of anxiety. The Internet and Higher Education, 41,34–44.https://doi.org/10.1016/j.iheduc.2019.01.001
2. Agarwal, N., Jain, A., Gupta, A., & Tayal, D. K. (2021, November). Applying XGBoost Machine Learning Model to Succor Astronomers Detect Exoplanets in Distant Galaxies. In International Conference on Artificial Intelligence and Speech
3. Agarwal, N., Tayal, D.K.: FFT based ensembled model to predict ranks of higher educational institutions. Multimedia Tools and Applications **81**(23), 34129–34162 (2022)
4. Agarwal, N. and Tayal, D.K. (2023) A new model based on the extended COPRAS method for improving performance during the accreditation process of Indian Higher Educational Institutions. Computer Applications in Engineering Education.
5. Al-Asadi, M.A., Tasdemír, S.: Empirical comparisons for combining balancing and feature selection strategies for characterizing football players using FIFA video game system. IEEE Access **9**, 149266–149286 (2021)
6. Al-Asadi, M.A., Tasdemır, S.: Predict the value of football players using FIFA video game data and machine learning techniques. IEEE Access **10**, 22631–22645 (2022)
7. Alhadreti, O.: Assessing academics' perceptions of blackboard usability using SUS and CSUQ: a case study during the COVID-19 pandemic. International Journal of Human-Computer Interaction **37**(11), 1003–1015 (2021). https://doi.org/10.1080/10447318.2020.1861766
8. Alonso, F., Manrique, D., Martinez, L., & Vines, J. M. (2011). How blended learning reduce sunder achievement in higher education: An experience in teaching computer sciences. IEEE Transactions on Education, 54(3), 471–478. https://doi.org/10.1109/TE.2010.2083665
9. Anthony, B., Kamaludin, A., Romli, A., Raffei, A.F.M., Nincarean ALEhPhon, D., Abdullah, A., Ming, G.L., Shukor, N.A., Nordin, M.S., & Baba, S. (2019). Exploring the role of blended learning for teaching and learning effect iveness in institution so fhigher learning: An empirical investigation. Education and Information Technologies, 24(6), 3433–3466. https://doi.org/10.1007/s10639-019-09941-z
10. Banik, D., Bhattacharjee, D.: Mitigating data imbalance issues in medical image analysis. In: Rana, D.P., Mehta, R.G. (eds.) Data Preprocessing, Active Learning, and Cost Perceptive Approaches for Resolving Data Imbalance, pp. 66–89. IGI Global (2021)
11. Beaunoyer, E., Dupéré, S., & Guitton, M.J. (2020). COVID19 and digital in equalities: Reciprocal impacts and mitigation strategies. Computers in Human Behavior, 111, 106424. https://doi.org/10.1016/j.chb.2020.106424.
12. Bhagat, R.S., Krishnan, B., Nelson, T.A., Leonard, K.M., Ford, D.L., Billing, T.K.: Organizational stress, psychological strain, and work outcomes in six national contexts. IEEE Eng. Manage. Rev. **38**(4), 39–57 (2010)
13. Biner, P. M., Welsh, K. D., Barone, N. M., Summers, M., & Dean, R. S. (1997). The impact of remote–site group size on student satisfaction and relative performance in interactive tele courses. Int. J. Phytoremediation, 11(1), 23–33. https://doi.org/10.1080/08923649709526949
14. Brooks, S., Longstreet, P., &Califf, C. (2017). Social media induced techno stress and its impact on internet addiction: Adistraction-conflict theory perspective. AIS Trans. Hum.-Comput. Interaction, 9(2), 99–122.https://doi.org/10.17705/1thci.00091
15. Cabero-Almenara, J., Fernández-Batanero, J.M., & Barroso Osuna, J. (2019). Adoption of augmented reality technology by university students. Heliyon, 5(5), e01597. https://doi.org/10.1016/j.heliyon.2019.e01597

16. Cao, X., Masood, A., Luqman, A., Ali, A.: Excessive use of mobile social networking sites and poor academic performance: antecedents and consequences from stressor-strain-out come perspective. Comput. Hum. Behav. **85**, 163–174 (2018). https://doi.org/10.1016/j.chb.2018.03.023
17. Chatterjee, S., Maity, S., Bhattacharjee, M., et al.: Variational autoencoder based imbalanced COVID-19 detection using chest X-ray images. New Gener. Comput. (2022). https://doi.org/10.1007/s00354-022-00194-y
18. Chawla, N.V., Bowyer, K.W., Hall, L.O., Philip Kegelmeyer, W.: Smote synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
19. Chin, C. (2020). Learning must n't stop with Covid19. The Star Online. https://www.thestar.com.my/news/education/2020/03/29/learning-mustnt-stop-with-covid-19
20. Chiu, C. M., & Wang, E. T. G. (2008). Understanding Web-based learning continuance intention: The role of subjective task value. Information and Management, 45(3), 194–201.
21. Etherington, C. (2017). Selfmotivation is essential to elearning. Elearning Inside. https://news.elearninginside.com/self-motivation-essential-elearning/
22. Fozdar, B.I., & Kumar, L.S. (2007). Mobile learning and student retention. International Review of Research in Open and Distance Learning, 8(2), 1–18. https://files.eric.ed.gov/fulltext/EJ800952.pdf
23. Garrison, D.R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. The Internet and Higher Education, 7(2), 95–105. https://doi.org/10.1016/j.iheduc.2004.02.001
24. Güğerçin, U. (2020). Does tech no-stress justify cybers lacking? An empirical study based on the neutralisation theory. Behaviour & Information Technology, 39(7), 824–836. https://doi.org/10.1080/0144929X.2019.1617350.
25. Gupta, A., Sharma, S., Goyal, S., Rashid, M. (2020). Novel XGBoost Tuned Machine Learning Model for Software Bug Prediction, 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 376–380. https://doi.org/10.1109/ICIEM48762.2020.9160152.
26. Hung, W.H., Chen, K., Lin, C.P.: Does the proactive personality mitigate the adverse effect of technostress on productivity in the mobile environment? Telematics Inform. **32**(1), 143–157 (2015)
27. Johnson, D.A., & Christensen, J. (2011). A comparison of simplified-visually rich and traditional presentation styles. Teaching of Psychology, 38(4), 293–297. https://doi.org/10.1177/0098628311421333
28. Kapasia, N., Paul, P., Roy, A., Saha, J., Zaveri, A., Mallick, R., & Chouhan, P. (2020). Impact of lockdown on learning status of undergraduate and postgraduate students during COVID-19 pandemic in West Bengal, India. Children and youth services review, 116, 105194.
29. Lee, D.Y., Ryu, H.: Learner acceptance of a multimedia-based learning system. Int. J. Hum.-Comput. Interaction **29**(6), 419–437 (2013). https://doi.org/10.1080/10447318.2012.715278
30. Lee, D.Y., Shin, D.-H.: Effects of spatial ability and richness of motion cue on learning in mechanically complex domain. Comput. Hum. Behav. **27**(5), 1665–1674 (2011). https://doi.org/10.1016/j.chb.2011.02.005
31. Iivari, N., Sharma, S., & Ventä-Olkkonen, L. (2020). Digital transformation of everyday life – How COVID-19 pandemic transformed the basic education of the young generation and why information management research should care? Int. J. Inform. Manag., 55, 102183. https://doi.org/10.1016/j.ijinfomgt.2020.102183
32. Li, L.-Y. (2019). Effect of prior knowledge on attitudes, behavior, and learning performance in video lecture viewing. Int. J. Hum.–Comput. Interaction, 35(4–5), 415–426. https://doi.org/10.1080/10447318.2018.1543086
33. Lwoga, E.T., Komba, M.: Antecedents of continued usage intentions of web-based learning management system in Tanzania. Educ. Train. **57**(7), 738–756 (2015). https://doi.org/10.1108/ET-02-2014-0014
34. New Straits Times. (2020). Online classes lack student-teacher engage- ment: Study. New Straits Times. https://www.nst.com.my/world/region/2020/05/589963/online-classes-lack-student-teacher- engagement-study.
35. Nguyen, Q.N., Ta, A., Prybutok, V.: An integrated model of voice-user interface continuance intention: the gender effect. International Journal of Human-Computer Interaction **35**(15), 1362–1377 (2019). https://doi.org/10.1080/10447318.2018.1525023
36. O'Callaghan, F.V., Neumann, D.L., Jones, L., Creed, P.A.: The use of lecture recordings in higher education: a review of institutional, student, and lecturer issues. Educ. Inf. Technol. **22**(1), 399–415 (2017). https://doi.org/10.1007/s10639-015-9451-z

37. P., N. P., Rajani, M., Georg, G., Lynnea, E., & Raghu, R. (2018). Towards an inclusive digital literacy framework for digital India. Education + Training, 60(6), 516–528. https://doi.org/10.1108/ET-03-2018-0061
38. Pal, D., Patra, S.: University students' perception of video-based learning in times of COVID-19: A TAM/TTF perspective. International Journal of Human-Computer Interaction **37**(10), 903–921 (2021). https://doi.org/10.1080/10447318.2020.1848164
39. Pal, D., & Vanijja, V. (2020). Perceived usability evaluation of Microsoft Teams as an online learning platform during COVID-19 using system usability scale and technology acceptance model in India. Children and Youth Services Review, 119, 105535. https://doi.org/10.1016/j.childyouth.2020.105535
40. Park, C., Kim, D., Cho, S., Han, H.-J.: Adoption of multimedia technology for learning and gender difference. Comput. Hum. Behav. **92**, 288–296 (2019). https://doi.org/10.1016/j.chb.2018.11.029
41. Saade, R. G., Kira, D., Mak, T., & Nebebe, F. (2017). Anxiety and performance in online learning. Informing science and information technology education conference (Vietnam: Informing Science Institute) (pp. 147–157).
42. Sarstedt, M., Henseler, J., Ringle, C.M.: Multigroup analysis in partial least squares (PLS) path modeling: Alternative methods and empirical results. Adv Int Mark **2011**(22), 195–218 (2011)
43. Shi, C., Yu, L., Wang, N., Cheng, B., & Cao, X. (2020). Effects of social media overload on academic performance: A stressor–strain–outcome perspective. Asian Journal of Communication, 30(2), 179–197. https://doi.org/10.1080/01292986.2020.1748073
44. Shen, R., Wang, M., Gao, W., Novak, D., Tang, L.: Mobile learning in a large, blended computer science classroom: system function, pedagogies, and their impact on learning. IEEE Trans. Educ. **52**(4), 538–546 (2009). https://doi.org/10.1109/TE.2008.930794
45. Shu, Q., Tu, Q., Wang, K.: The impact of computer self-efficacy and technology dependence on computer-related technostress: a social cognitive theory perspective. Int. J. Hum.- Comput. Interaction **27**(10), 923–939 (2011). https://doi.org/10.1080/10447318.2011.555313
46. Tamm, S. (2019). Disadvantages of e-learning. E-Student.Org. https://estudent.org/disadvantages-of-e-learning/
47. Ulrich, F., Helms, N.H., Frandsen, U.P., Rafn, A.V.: Learning effectiveness of 360° video: Experiences from a controlled experiment in healthcare education. Interact. Learn. Environ. **26**(1), 1–14 (2019). https://doi.org/10.1080/10494820.2019.1579234
48. NESCO. (2020). Quality education. In COVID-19 educational disruption and response. https://en.unesco.org/news/covid-19-educational-disruption-and-response
49. UNESCO. (n.d.). No title. COVID-19 Educational Disruption and Response. Retrieved June 30, 2020, from https://en.unesco.org/news/covid-19-educational-disruption-and-response.
50. Xu, D., Huang, W. W., Wang, H., & Heales, J. (2014). Enhancing e-learning effectiveness using an intelligent agent-supported personlized personalized virtual learning environment: An empirical investigation. Information and Management, 51(4), 430–440. https://doi.org/10.1016/j.im.2014.02.009
51. Zhang, D., Zhou, L., Briggs, R.O., Nunamaker, J.F.: Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. Inform. Manag. **43**(1), 15–27 (2006). https://doi.org/10.1016/j.im.2005.01.004
52. Zheng, X., Lee, M.K.O.: Excessive use of mobile social networking sites: Negative consequences on individuals. Comput. Hum. Behav. **65**, 65–76 (2016). https://doi.org/10.1016/j.chb.2016.08
53. Zhou, J., Rau, P.-L. P., & Salvendy, G. (2014). Older adults' text entry on smartphones and tablets: Investigating effects of display size and input method on acceptance and performance. International Journal of Human–Computer Interaction, 30(9), 727–739. https://doi.org/10.1080/10447318.2014.924348
54. Zimmerman, B.J., Bandura, A., Martinez-Pons, M.: Self- motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting. Am. Educ. Res. J. **29**(3), 663–676 (1992). https://doi.org/10.3102/0002831202.9003663

Authors and Affiliations

Nidhi Agarwal[1,2] · Sachi Nandan Mohanty[3] · Shweta Sankhwar[4] · Jatindra Kumar Dash[5]

✉ Jatindra Kumar Dash
jatindrakumar.d@srmap.edu.in.com

Nidhi Agarwal
nidhiagarwal82@gmail.com

Sachi Nandan Mohanty
sachinandan09@gmail.com

Shweta Sankhwar
shweta.sank@gmail.com

1 Department of Computer Science & Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India

2 Department of Computer Science Engineering, Galgotias University, Greater Noida, India

3 Department of Computer Science Engineering, Vardhaman College of Engineering (Autonomous), Hyderabad, Telengana, India

4 Department of Computer Science, Maitreyi College (University of Delhi), Bapu Dham, Chanakyapuri, New Delhi, India

5 Department of Computer Science & Engineering, SRM University-AP, Amaravati, Andhra Pradesh, India