# An Effective Data Mining Method for Determining Higher Education Students' Satisfaction with Online Learning During the COVID-19

Harshini Reddy Dasari[1], Dr. G. N. R. Prasad[2]
[1]*MCA Student, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India*
[2]*Assistant Professor, Department of MCA, Chaitanya Bharathi Institute of Technology (A), Gandipet, Hyderabad, Telangana State, India*

***Abstract*: All educational organizations strive to improve the overall quality of education by raising students' academic performance. In this regard, Educational Data Mining (EDM) is a rapidly growing research field that employs the the core ideas of data mining (DM) to assist academic institutions in determining useful details about how satisfied students are with the online learning experience (SSL) (OL) during the COVID-19 lock-down. To provide the optimum educational environments, several approaches have been explored using EDM to anticipate students' behavior. As a result, Feature Selection (FS) is often used to discover the most relevant subset of characteristics with the lowest cardinality. Because the FS process has a substantial impact on the predicted accuracy of a satisfaction model, the usefulness of the SSL model in conjunction with FS approaches is thoroughly investigated in this research. In this regard, a dataset of student evaluations of OL courses was initially gathered online through a questionnaire. The performance of wrapper FS approaches in DM and classification algorithms was evaluated in terms of fitness values using this datasets. Finally, the goodness of subsets with various cardinalities is assessed in terms of prediction accuracy and the number of chosen features through evaluating the performance of 11 wrapper-based FS algorithms in addition to Support Vector Machine (SVM) and k-Nearest Neighbor (k-NN) as baseline classifiers. The studies indicated the optimum dimensionality of the feature subset as well as the best technique. The current study's results clearly corroborate the well-known association between the presence of a small number of characteristics and an improvement in prediction accuracy. The relevance of FS for high-accuracy SSL prediction is outstanding, as the necessary collection of traits may effectively aid in the development of constructive instructional initiatives. On the used real-time dataset, our work offers a feature size reduction of up to 80% as well as up to 100% classification accuracy.**

***Keywords – Classification, COVID-19, educational data mining (EDM), feature selection (FS), machine learning (ML).***

## 1. INTRODUCTION

The onset of the COVID-19 outbreak has caused widespread public-health concern. As a result of these emergency conditions, several governments have opted to implement lock-downs in order to reduce social interaction and minimise transmission [1], [2]. The COVID-19 has had a significant impact on Higher Education Institutions (HEIs). Many unorthodox educational methods have been presented to ensure the continuation of the educational process in light of the effects of this pandemic and the necessity for alternative remedies. Learning in an asynchronous or synchronous environment using various devices, such as PCs and mobile phones, is referred to as online learning (OL) and so on with an Internet connection, was one of the answers. Using these platforms, students may study and interact with professors and other classmates from anywhere [3]. OL has grown in popularity in the last decade because it allows for more flexibility in time and place, faster study, greater accessibility, more active access to a wider variety and greater amount of information, and lower monetary charges [4]. The most noticeable part of the transition was the use of online platforms and courses. However, we continue to meet a variety of roadblocks and obstacles. Despite the fact that strong digital platforms and infrastructure are necessary for the delivery of online courses and the collecting of data; worldwide

student learning is hampered by inadequate Internet connections. New technology is required for both students and instructors to engage smoothly with self-directed and dynamic instruction. In order to guarantee the general caliber of virtual learning about academic achievements, a credible evaluation system was required. In the era of epidemics, quality is assessed based on the accomplishment of learning objectives and the growth of social and emotional aspects [5, 6]. Because of this, a tool is needed to analyze the entire learning process as well as the functions and interactions of teachers, students, and instructional materials in post-digital learning environments. The accomplishment of learning objectives and the improvement of Student Satisfaction Level (SSL) by colleges and universities are important since these traits suggest, however indirectly, the effectiveness of those institutions' OL systems [7]. Throughout the research period, SSL has been a part of the relative degree of experiences and perceived performance related to educational services. Students' opinions of their educational experiences, resources, and facilities play a role in SSL evaluations [8]. According to [9], SSL can only be completed if there is no discrepancy between what is introduced by the service provider and what is expected.

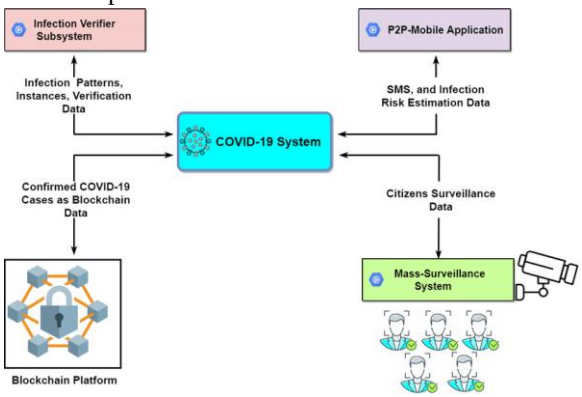


Fig.1: Example figure

In this light, it is worth noting that Educational The educational research process stands to benefit greatly from the application of data mining (EDM) [10], [11]. Therefore, in order to provide conclusions that can be understood, the collected data needs to be appropriately structured and analyzed. The selection of an appropriate strategy for the analysis is also crucial to the effectiveness of EDM approaches. One of the most effective and important data analytics technologies is feature selection (FS). Highdimensional data may have undesirable repercussions in applied models. Two examples include extending the training period with improved features and model processing [12]. In machine learning (ML) and data mining (DM), FS is crucial, especially when dealing with high-dimensional datasets that contain characteristics that are redundant, noisy, and useless. In order to provide strong prediction results, FS aims to select a subset of variables from the inputs that can more accurately represent the data while reducing the impact of noise and extraneous characteristics. The selection of feature subsets is an important problem in knowledge discovery, the acceleration of DM approaches, and performance optimization [13], [14]. FS has been shown to be a successful and efficient data preparation method for preparing high-dimensional data in numerous DM and ML scenarios. FS goals could be to create more precise models, speed up data mining, and show data in an intelligible manner [15].

## 2. LITERATURE REVIEW

Economic and social consequences of human mobility restrictions under COVID-19:

Several national governments have imposed lockdown restrictions in response to the coronavirus disease 2019 (COVID-19) pandemic in order to lower infection rates. We study how lockdown methods influence the economic situations of people and local governments by conducting a comprehensive analysis on near-real-time Italian mobility data given by Facebook. The shift in mobility is modelled as an external shock, akin to a natural catastrophe. There are two ways in which mobility constraints impact Italian individuals. First, we find that lockdown has a greater effect in localities with more budgetary capability. Second, we find evidence of a segregation impact, since mobility contraction is larger in towns with more inequality and persons with lower per capita income. Our findings indicate both the societal costs of lockdown and an unparalleled level of difficulty: On the one hand, the crisis reduces fiscal income for both national and local governments; on the other hand, a considerable fiscal effort is required to support the most vulnerable persons and to offset the rise in poverty and inequality caused by the lockdown.

Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China:

We estimate the effect of human movement limitations, namely the lockdown of Wuhan on January 23, 2020, on the containment and delay of the Novel Coronavirus's transmission (2019-nCoV). We use difference-in-differences (DID) estimates to separate the lockdown impact from other confounding variables such as the fear effect, virus effect, and Spring Festival effect. Wuhan's lockdown lowered inflows to Wuhan by 76.98%, outflows from Wuhan by 56.31%, and movements within Wuhan by 55.91%. We also assess the dynamic impact of up to 22 delayed population arrivals from Wuhan and other Hubei cities – the heart of the 2019-nCoV epidemic – on new infection cases in the destination cities. We also show that improved social distancing strategies in 98 Chinese cities outside Hubei province were successful in lowering the influence of population inflows from Hubei province epicentre cities on the spread of 2019-nCoV in destination cities. We discover that if Wuhan had not been closed down on January 23, 2020, the number of COVID-19 cases in the 347 Chinese cities outside Hubei province would be 105.27% greater. Our results are significant to worldwide pandemic control efforts.

How many ways can we define online learning? A systematic literature review of definitions of online learning (1988–2018):

For more than two decades, online learning as a concept and a buzzword has been a focus of educational study. We give findings from a systematic literature review for definitions of online learning in this work since the idea of online learning, although often described, has a variety of meanings associated to it. Authors and intellectuals use the phrase to refer to highly different, if not opposing, ideas. We did a comprehensive review of the literature from 1988 to 2018 to explore the quantity and substance of definitions of online learning. We gathered 46 definitions from 37 sources and performed a content analysis on these definition sets. The content analysis of the gathered definitions resulted in a knowledge of the key factors for defining online learning, as well as the ambiguity around the terminology and synonyms for online learning. The history of the definition of online learning has also been traced to the progress of technology over the previous three decades.

Exploring the role of multimedia in enhancing social presence in an asynchronous online course:

The demand for online education is increasing, as is worry over the quality of online education. One of the primary drawbacks of online education is the social isolation. To reduce this sense of isolation, previous study suggests concentrating on measures that improve social presence in an online classroom. However, there are several barriers to developing an online learning experience in which learners have a strong sense of social presence. This might be due to the ambiguity of the social presence concept, and most previous research has assessed social presence using self-report questionnaires. The goal of this research was to look at how social presence develops in a multimodal discussion forum and how multimedia aids in the development of social presence in an asynchronous online course. Furthermore, the goal was to learn about the perspectives of students and the teacher on utilising multimedia in an online course for diverse objectives. This research investigated the influence of multimedia in boosting social presence in an online course using a mixed method exploratory case study technique. To analyse the usage of multimedia and the pattern of growth of social presence, the research used three distinct frameworks: social constructivism, community of inquiry, and social network analysis. The research revealed how multimedia might improve social presence in an online learning community in a variety of ways. The results revealed that, although certain multimodal technologies, such as VoiceThread, enhanced the quantity of engagement, it did not result in an increase in social presence. Furthermore, the research demonstrated that the Rourke et al. (2001) social presence coding procedure was unable to capture some social presence markers in a multimodal discussion forum. Several hypotheses were created in this research to better understand a student's popularity inside a learning network.

The possible immunological pathways for the variable immunopathogenesis of COVID-19 infections among healthy adults, elderly and children:

COVID-19, a novel Coronavirus identified in December 2019 in Wuhan, China, triggered a global outbreak. It is still unknown what causes this viral infection in humans or the precise tactics of host immune response in battling this unprecedented danger to humans. However, the morbidity and fatality rates of COVID-19 infections range from

asymptomatic and moderate to lethal and severe. Surprisingly, youngsters were shown to be immune to severe or fatal critical infections, but the elderly and immunocompromised adults are the most severely impacted by this virus. It is crucial to reveal the probable viral and host interactions that result in such diverse clinical outcomes in COVID-19 individuals.

## 3. METHODOLOGY

The most noticeable part of the transition was the use of online platforms and courses. However, we continue to meet a variety of roadblocks and obstacles. Despite the fact that solid digital infrastructure and platforms are essential for online course delivery and data collection engagement, poor Internet connection impairs student learning globally. New technology is required for both students and instructors to engage smoothly with active and self-directed learning. To assure the overall quality of online education in terms of learning outcomes, a credible evaluation system was required. Quality is judged in the epidemic age by the attainment of learning goals and the development of emotional and social dimensions. As a result, a tool to analyse the learning process as a whole, as well as the roles and interactions of instructors, learners, and teaching materials in post-digital learning contexts, is required. The capacity of universities and colleges to accomplish learning goals and enhance Student Satisfaction Level (SSL) is significant since these metrics indicate the efficacy of such institutions' OL systems in an indirect way. SSL is a component of the relative level of experiences and perceived performance connected to educational services throughout the research period. SSL is decided in part by how students assess their educational experiences, services, and facilities. SSL can only be accomplished when there is no gap between what is anticipated and what is introduced by the service provider, according to.

Disadvantages:

1. However, we continue to meet numerous roadblocks and obstacles.
2. The capacity of universities and colleges to accomplish learning goals and enhance Student Satisfaction Level (SSL) is significant since these characteristics indicate the efficacy of such institutions' OL systems in an indirect way.

PROPOSED SYSTEM:

An ML model was created in this research to get the maximum accuracy outcomes during testing. The suggested model was built using two ML Classifiers, k-NN and SVM. Historically, the most extensively used classification systems have been k-NN and SVM. Furthermore, this research employed random forest, decision tree, voting classifier, transfer learning using CNN integrated with LSTM layers, and some other feature selection methods like ABC, Whale optimization, sailfish optimization, and others.

Advantages:

1. The current study's results clearly corroborate the well-known association between the presence of a small number of characteristics and an improvement in prediction accuracy.
2. The relevance of FS for high-accuracy SSL prediction is amazing, as the appropriate collection of traits may effectively aid in the development of constructive instructional tactics.
3. On the used real-time dataset, our work offers a feature size reduction of up to 80% as well as up to 100% classification accuracy.

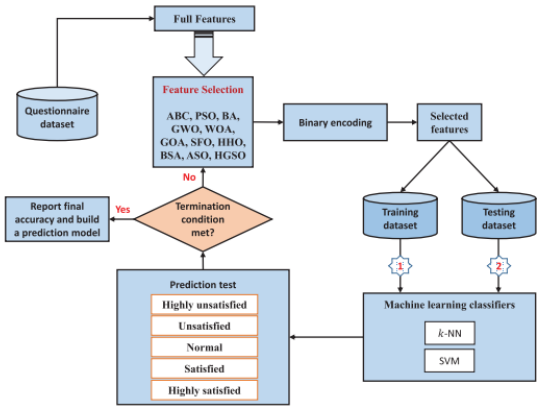


Fig.2: System architecture

MODULES:

To carry out the aforementioned project, we created the modules listed below.

- Data exploration: we will put data into the system using this module.
- Processing: we will read data for processing using this module.
- Using this module, data will be separated into train and test groups.
- Model generation: Create k-NN, SVM, random forest, decision tree, voting classifier, transfer learning using CNN integrated with LSTM layers

ABC, Whale optimization, and sailfish optimization.
- Calculated algorithm accuracy.
- User signup and login: Using this module will result in registration and login. User input: Using this module will result in predicted input.
- Prediction: final predicted shown

## 4. IMPLEMENTATION

ALGORITHMS:

K-NN: The k-nearest neighbours method, often known as KNN or k-NN, is a non-parametric, supervised learning classifier that employs proximity to classify or predict the grouping of a single data point.

SVM: The SVM algorithm's purpose is to find the optimum line or decision boundary for categorising n-dimensional space so that we may simply place fresh data points in the proper category in the future. A hyperplane is the optimal choice boundary.

Random forest: Random forest is a kind of Supervised Machine Learning Algorithm that is often used in classification and regression issues. It constructs decision trees from several samples and uses their majority vote for classification and average for regression.

Decision tree: A decision tree is a graph that illustrates every potential outcome for a given input using a branching mechanism. Decision trees may be hand-drawn or generated using a graphics application or specialist software. When a group has to make a decision, decision trees may help concentrate the debate.

Voting classifier: A voting classifier is a machine learning estimator that trains many base models or estimators and predicts by aggregating their results. Aggregating criteria may be coupled voting decisions for each estimator output.

Transfer learning with CNN embedded with LSTM layers: The practise of adapting previously learned information to new settings is known as transfer of learning. Examples of learning transfer: In class, a student learns to solve polynomial equations and then applies that knowledge to comparable problems for homework. In class, a teacher explains numerous psychological diseases.

The basic premise of transfer learning is simple: Transfer the information of a model trained on a big dataset to a smaller dataset. For object recognition, we freeze the network's early convolutional layers and just train the final few levels that make a prediction.

LSTM Recurrent Neural Networks are a suitable option for time series prediction tasks, however the technique is predicated on having enough training and testing data from the same distribution.

ABC: The artificial bee colony (ABC) is a swarm intelligence-based stochastic search approach that simulates the behaviour of honey bee swarms hunting for food. The ABC algorithm divides bees in a colony into three categories: employed bees (forager bees), spectator bees (observation bees), and scouts. There is only one hired bee for each food source. That is, the number of bees employed equals the number of food sources.

Whale optimization: The Whale Optimization Algorithm (WOA) is a novel optimization approach for problem solving. This algorithm comprises three operators that imitate the humpback whale's hunt for prey, surrounding prey, and bubble-net foraging behaviour.

Sailfish optimization: The SFO is a metaheuristic algorithm based on population. The sailfish are supposed to be candidate solutions in this technique, and the issue variables are the location of the sailfish in the search space. As a result, the population of the solution space is produced at random.

## 5. EXPERIMENTAL RESULTS

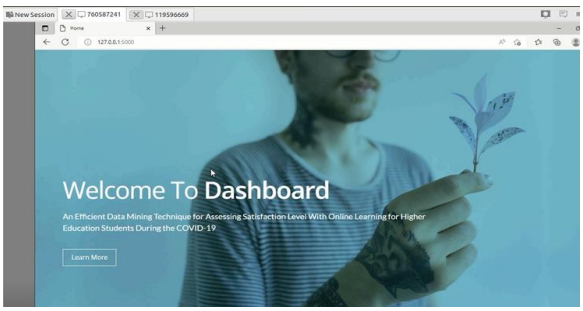


Fig.3: Home screen

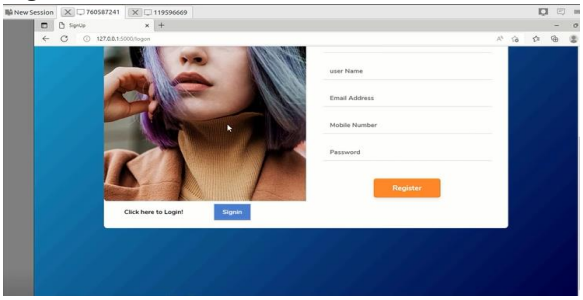


Fig.4: User signup

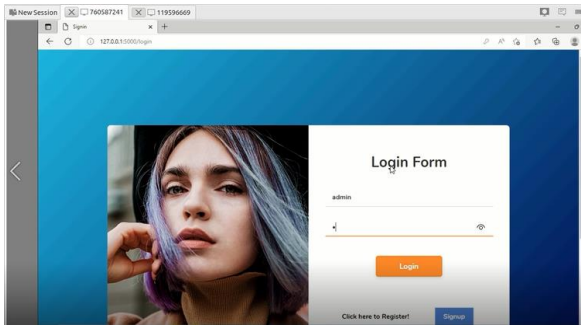

Fig.5: User signin

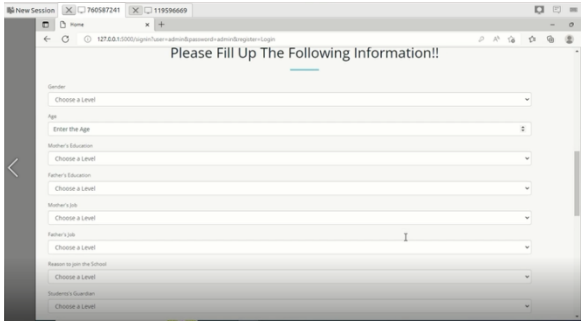

Fig.6: Main screen

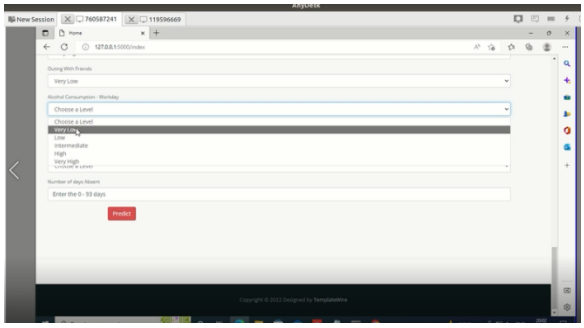
Fig.7: User input

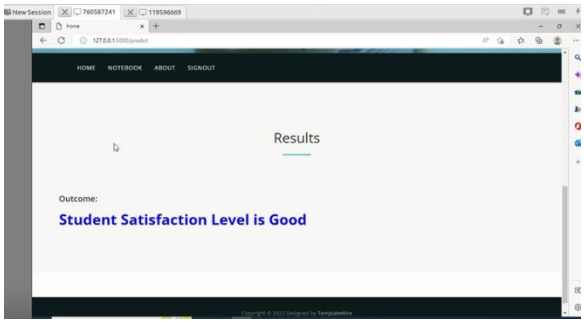

Fig.8: Prediction result

## 6. CONCLUSION

An SSL prediction model was suggested in this article to enhance the educational process during COVID 19 and overcome obstacles hindering OL advancement. Our model is made up of four parts: data preparation, FS, ML classifiers, and ML model evaluation. The data was gathered using a questionnaire created specifically to determine how OL affects children. The present research included certain common SSL evaluation criteria, such as faculty member duty (online), online teaching and lectures, assessment methods, and E-Tests. Eleven wrapper-based FS algorithms were used to choose the optimal collection of features. In addition, to detect differences, two ML classifiers, k-NN and SVM, were applied to all characteristics and the chosen ones. The results showed that utilising just the chosen features increased overall accuracy by 2% and 8% for k-NN and SVM, respectively, when compared to using all features; applying FS methods enhanced overall mean accuracy by 100% for k-NN and SVM. In terms of exploration and exploitation skills, the SFO algorithm with k-NN and SVM performs the best (fitness). It only determined four characteristics. We find that four characteristics (rather than the original 20) influenced SSL and are adequate to predict SSL using OL with 100% accuracy. "The lectures are presented in an attractive style," "the teaching method in this course encourages me to participate actively during the classes," "the quiz has been prepared in degrees," and "students trained on how to solve exams online by designing an experimental quiz" are the minimal, yet critical, selected features. This might assist HEIs in predicting SSL early on and presenting the diagnosis and treatment to prevent hiccups in the educational process and obtain the most important potential results during acute crises such as the COVID-19.

## 7. FUTURE SCOPE

In the future, since the Random Forest (RF) model can completely suit the input-output relationship with unbounded high complexity, it may be attempted on the suggested real-time dataset with the 11 FS approaches.

## REFERENCE

[1] G. Bonaccorsi, F. Pierri, M. Cinelli, A. Flori, A. Galeazzi, F. Porcelli, A. L. Schmidt, C. M. Valensise, A. Scala, W. Quattrociocchi, and F. Pammolli, ‘‘Economic and social consequences of human mobility restrictions under COVID-19,’’ Proc. Nat. Acad. Sci. USA, vol. 117, no. 27, pp. 15530–15535, Jul. 2020.

[2] H. Fang, L. Wang, and Y. Yang, ‘‘Human mobility restrictions and the spread of the novel coronavirus (2019-nCoV) in China,’’ J. Public Econ., vol. 191, Nov. 2020, Art. no. 104272.
[3] V. Singh and A. Thurman, ‘‘How many ways can we define online learning? A systematic literature review of definitions of online learning (1988–2018),’’ Amer. J. Distance Educ., vol. 33, no. 4, pp. 289–306, Oct. 2019.
[4] C. Khurana, ‘‘Exploring the role of multimedia in enhancing social presence in an asynchronous online course,’’ Ph.D. dissertation, Rutgers Univ.-Graduate School, New Brunswick, NJ, USA, 2016.
[5] M. A. Peters, H. Wang, M. O. Ogunniran, Y. Huang, B. Green, J. O. Chunga, E. A. Quainoo, Z. Ren, S. Hollings, C. Mou, S. W. Khomera, M. Zhang, S. Zhou, A. Laimeche, W. Zheng, R. Xu, L. Jackson, and S. Hayes, ‘‘China’s internationalized higher education during COVID-19: Collective student autoethnography,’’ Postdigital Sci. Educ., vol. 2, no. 3, pp. 968–988, Oct. 2020.
[6] A. S. Abdulamir and R. R. Hafidh, ‘‘The possible immunological pathways for the variable immunopathogenesis of COVID-19 infections among healthy adults, elderly and children,’’ Electron. J. Gen. Med., vol. 17, no. 4, p. em202, Mar. 2020.
[7] A. Rasouli, Z. Rahbania, and M. Attaran, ‘‘Students’ readiness for e-learning application in higher education,’’ Malaysian Online J. Educ. Technol., vol. 4, no. 3, pp. 51–64, 2016.
[8] S. Raime, M. F. Shamsudin, R. A. Hashim, and N. A. Rahman, ‘‘Students’ self-motivation and online learning students’ satisfaction among unitar college students,’’ Asian J. Res. Educ. Social Sci., vol. 2, no. 3, pp. 62–71, 2020.
[9] S. Wilkins and M. S. Balakrishnan, ‘‘Assessing Student satisfaction in transnational higher education,’’ Int. J. Educ. Manage., vol. 27, no. 2, pp. 143–156, Feb. 2013.
[10] A. Dutt, M. A. Ismail, and T. Herawan, ‘‘A systematic review on educational data mining,’’ IEEE Access, vol. 5, pp. 15991–16005, 2017.
[11] N. Kapasia, P. Paul, A. Roy, J. Saha, A. Zaveri, R. Mallick, B. Barman, P. Das, and P. Chouhan, ‘‘Impact of lockdown on learning status of undergraduate and postgraduate students during COVID-19 pandemic in West Bengal, India,’’ Children Youth Services Rev., vol. 116, Sep. 2020, Art. no. 105194.
[12] M. Canayaz, ‘‘MH-COVIDNet: Diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images,’’ Biomed. Signal Process. Control, vol. 64, Feb. 2021, Art. no. 102257.
[13] Q. Al-Tashi, H. Rais, and S. Jadid, ‘‘Feature selection method based on grey wolf optimization for coronary artery disease classification,’’ in Proc. Int. Conf. Reliable Inf. Commun. Technol. Springer, 2018, pp. 257–266.
[14] A. I. Hammouri, M. Mafarja, M. A. Al-Betar, M. A. Awadallah, and I. Abu-Doush, ‘‘An improved dragonfly algorithm for feature selection,’’ Knowl.-Based Syst., vol. 203, Sep. 2020, Art. no. 106131.
[15] Q. Al-Tashi, H. M. Rais, S. J. Abdulkadir, S. Mirjalili, and H. Alhussian, ‘‘A review of grey wolf optimizer-based feature selection methods for classification,’’ in Evolutionary Machine Learning Techniques, 2020, pp. 273–286.