International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

Analysis and Prediction of Student Performance by Using A Hybrid Optimized BFO-ALO Based Approach

Student Performance Prediction using Hybrid Approach

Sandeep Kumar[1], Bindiya Ahuja[2]
[1]Department of Computer Science and Engineering
Lingaya's Vidyapeeth
Faridabad, India
e-mail: 19phcs04w@lingayasvidyapeeth.edu.in
[2]Department of Computer Science and Engineering
Lingaya's Vidyapeeth
Faridabad, India

Abstract—Data mining offers effective solutions for a variety of industries, including education. Research in the subject of education is expanding rapidly because of thebigquantityof student data that can be utilized to uncover valuable learning behavior patterns. This research presents a method for forecasting the academic presentation of students in Portuguese as well as math subjects, and it is describing with the help of 33 attributes. Forecasting the educationalattainment of students is the most popular field of study in the modern period. Previous research has employed a variety of categorization algorithms to forecast student performance. Educational data mining is a topic that needs a lot of research to improve the precision of the classification technique and predict how well students will do in school. In this study, we made a method to predict how well a student will do that uses a mix of optimization techniques. BFO and ALO-based popular optimization techniques were applied to the data set. Python was used to process all the files and conduct a performance comparison analysis. In this study, we compared our model's performance with various existing baseline models and examined the accuracy with which the hybrid algorithm predicted the student data set. To verify the expected classification accuracy, a calculation was performed. The experiment's findings indicate that the BFO-ALO Based hybrid model, which, out of all the methods, with a 94.5 percent success rate, is the preferred choice.

Keywords-educational data mining; academic performance; BFO and ALO techniques; CNN; PCA

I. INTRODUCTION

The study of EDM, which applies data mining techniques to educational settings, is an exciting new area of research. Education data mining (EDM) programs help students learn more about education by revealing hidden patterns in student records. The field of EDM is very interested in the problem of predicting student success.

In a university context, it would be very beneficial to analyze students' learning activities and enhance their learning results if significant patterns in educational processes could be detected. The scientific community has paid a lot of attention to EDM since solving the issue of obtaining useful information from educational data is essential for improving educational institutions' teaching methods and guiding principles [1].

To capture more complex relationships between the variables that make up a dataset, the data mining community proposed the theory of Relational Association Rules (RARs). Problems in supervised classification and data mining can be efficiently addressed by using either ordinal or relational association rule mining.

Using a student's term or course grades, [2] addresses the issue of predicting whether the student will have a great or bad academic performance in a given academic field. The problem is very relevant in the field of education because it has the potential to provide valuable information to students who are at threat of weakening a particular course. Having such support available all through the semester can help students avoid falling behind in their coursework. Despite the problem's obvious practical importance, it is complex and difficult because it is highly reliant on multiple conditions, including the development, the number of teststhrough the semester, the instructors, the student's health, the student's problem, the student's relationship condition, and their instructor support emergencies.

Due to the presence of data that can be recorded throughout the online learning process, it is possible to obtain a detailed understanding of how students learn and use that knowledge to

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

make predictions about their academic performance. The bulk of current studies use the information on students' learning habits gleaned from the associated learning management structure, but they often overlook other factors that could be as important to students' overall success in school. In a review study, it was shown that students' Internet use has a significant influence on their grades. Another obtained data from the university network also logs on the frequency and duration of Internet usage inthe student's group. An individual's online conduct can be classified into a unique pattern, and a student's static information [3] both contribute to the prediction of their academic accomplishment. Students must do well in school because of the many factors that might affect their grades. Predicting a student's academic performance is a major research question in a variety of disciplines that has persisted for decades [4][5]. Predictive models can help educators identify pupils who are struggling and provide targeted instruction to increase their performance [6][7]. Different people have different ways of talking about how well students do in school, and measurable assessment is an important part of today's educational institutions. SPP (Student Performance Prediction) makes sense to help all participants in the learning system [8]. Learners could use SPP to help them choose relevant programmes or activities and develop educational strategies. SPP can help instructors adjust instructional content and teaching methods based on student ability, as well as identify at-risk students. SPP can support educational administrators in changing their behaviour and upgrading their curriculum programmes[9][10].

As a demonstration of the concept, the research presented in this paper focuses on analyzing the current educational achievement prediction model and clearly define the problem so that a classifier S PARAR based on RAR mining and CNN can be used to predict the academic performance of students. This proves that the optimised assignment can be trained faster than with traditional models. The rest of the paper is organised as follows: In the second part, you can see how the research was done. The relational association rule-based model is described in the third section as a hybrid BFO-ALO model. The experimental findings are described in the fourth section. In the fifth section, the comparative performance evaluation of the proposed BFO-ALO model is explained. Finally, and most crucially, Section 5 concludes this work and provides possible areas for further research.

II. Literature Review

Researchers would examine the research that has been done on the topic of relational association rule-based deep learning models for student academic performance prediction with several different approaches. In addition to that, this chapter provides an analysis of multiple research studies for convolutional neural networks (CNN) and naive bayes (NB) that were written through a wide range of authors.

Padma et al. (2022) [11] discussed that data mining methods are used with a range of businesses to evaluate the vast amounts of available data and information used in making important business decisions. Data mining has several applications in the field of education, including the evaluation of students' academic performance, attendance, instructor views, online activity, hazing, and stress. The findings will help teachers and mentors break down barriers between classes. The failure rate on the following semester's exams may be reduced thanks to this study's identification of students who need further instruction.

Li et al. (2022) [12] proposed that an all-inclusive model for predicting athletic performance on the basis of training is developed and deployed using a generative adversarial neural network. One paper suggests a multi - generational oppositional image enhancement algorithm that uses sampling from multiple restoration granules to fix the problems with feedforward neural network-based image restoration. These problems include a gradient that goes away, training that isn't stable, different local repair results, and a long training time. The data is input into a deep convolutional deep neural networks that can predict how well educators will do in future discussions. A temporal attention mechanism is included in the generative adversarial deep neural network's output, which gives different aspects of weekly student behaviour more priority.

Yağcı et al. (2022) [13] suggested that by using educational data mining, researchers can now probe the hidden connections in student records and predict how they will do in the classroom. This research presents an innovative use of machine learning by utilising intermediate exam scores to forecast students' final test scores. In order to make these projections, researchers utilised only the data collected from midterm exams, the departments, and the teachers.

Shreem et al. (2022)[14] proposed that educational institutions such as colleges, high schools, and vocational schools rely heavily on student performance prediction algorithms to boost their teaching and learning effectiveness. Examination centers, online courses, registration offices, and e-learning systems are all potential locations for such massive amounts of information. To extract useful insights from educational data, complexity is essential, necessitating a decrease in data dimensionality. In this study, the authors suggest using an EBGA to choose characteristics to be used as wrappers. To address this issue, researchers propose a novel hybrid selection mechanism that incorporates elements of both the k-means algorithm and the electromagnetic-like mechanism (EM) technique. In K-means clusters the data sets are into the manageable chunks, whereas EM rates each solution based on its "total force" (TF). According to the results, implementing the suggested technique has the potential to significantly improve the binary genetic algorithm's overall performance. Additionally, there is an increase of 1% to 11% across the board for classifier performance.

Najm et al. (2022) [15] said that the best way for universities to judge the short-term and long-term educational achievement of their students is to use solid data. These programmes use information from the past to help plan for the future. To find the hidden patterns in the data, you need to use certain tools and methods. Clustering with function approximation was the most accurate, with a 96.76% accuracy rate for predicting student performance and a 96.12% accuracy rate for predicting student grades.

Alachiotis et al. (2022)[16] suggested that in recent years, Educational Data Mining has matured into a robust strategy for uncovering latent relationships within educational datasets and predicting the outcomes of individuals' academic endeavors. Children's academic history, social behavior, and extracurricular pursuits are all data points that can be analyzed. Customized instruction, effective participation in student behavior, and a transparent link between students, instruction, and data are all outcomes. To predict student performance in the classroom, researchers use a variety of supervised methods in the current investigation. Researchers show that a voting generalization technique involving the top three classifiers may outperform the results of a single finely-tuned learning algorithm when the default settings of the learning algorithms are used.

Trakunphutthirak et al. (2022) [17] proposed that statistics gleaned from a school's learning management system (LMS) are typically all that professors in colleges and universities use to judge their students' performance in class. The study deviates from the previous literature by adding a new data set to the analysis, which increases our knowledge of the factors that affect students' academic performance. There are four main goals of such an investigation. The first is to fill in the blanks in the educational literature using Internet use log files and learning management system (LMS) data for educational data mining through machine learning methods. Second, machine learning techniques were applied to student LMS data and Internet activity log files to identify at-risk students with combining these datasets with demographic information. And last, the demographic details help lecturers make sense of the outlook. To further refine the forecasts, the study used other types of Internet usage information that were separated based on the nature of the usage data and the characteristics of the online surfing data.

Poudyal et al. (2022) [18] discussed that the potential for data mining to be used in the evaluation of educational data and the improvement of teaching and learning is expanding. A mountain of information is produced as a result of academic institutions' use of e-learning platforms and the proliferation of online and distance education programmes. This information could be useful for teachers in analysing and comprehending their student's learning habits. Because the data collected is unprocessed raw data, educational data mining is necessary to make inferences about students, such as their academic achievement.Despite the widespread application of standard machine learning to the task of predicting academic achievement in the classroom, convolutional neural network (CNN) development in this context has received very little attention. By combining two separate 2D CNN models, researchers were able to develop a hybrid model for predicting student performance. To assess the efficacy of our hybrid model, researchers transformed the 1D data in our sample into 2D image data. We compared our model's results to those of other common benchmarks. Researchers associated the performance of our model to that of other popular methods, and found that it was more accurate than k-nearest neighbour analysis, naive Bayes, decision trees, and logistic regression[19].

Surenthiran et al. (2021)[20]suggested an artificial intelligence (AI) methods for the prediction of academic performance. The neural network was utilized to analyze the pupils' performance. To classify the likelihood of dominating the tournament, a multiplayer perceptron neural network was constructed using a backward propagating method. The categorization performance rates were quite high, averaging 74.98 percent across all programs, illustrating the predictability of educational success provided by the determinants.

Hussain et al. (2021)[21]discussed that the use of computational intelligence in the creation of a forecasting model. Deep belief neural networks with atomic searching development optimization were used to categorise the participants in the study. Convergence of thought achieved by using several Boltzmann model powers with constrained behaviour. To begin, researchers used a collection of freely available educational materials to build the model displayed here. As a first step, the data was cleaned and organised into relevant categories. It is more difficult to adjust the second-stage learning rate variables in a Deep Belief Neural Network (DBNN). The ASO optimizer did the math on the user's basic school experience automatically. Evaluation measures were tested using the proposed research, and the outcomes were better than those obtained using older methods. A false-positive rate of less than 20% is achieved, increasing the conceptual model's accuracy to 90%.

Huang et al. (2021) [22] talked about how using deep learning in computer vision applications with a lot of data was becoming more common. The authors mostly used evolutionary algorithm and regression analysis to sort the information. When the parameter k equals 3, the learning

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

approach records a roughly comparable standard performance of 1.61 MAE and a loss of 4.7. The coefficient of determination, on the other hand, gives a loss of 6.77 and an MAE of 1.97. Multiple linear regression wasn't as good as the deep neural network.

Tsiakmaki et al. (2021) [23]looks at the potential of using AI to calculate the chances of academic achievement. The authors came up with two ways to predict student performance. One used feature-rich Support Vector Machine (SVM) training, and the other used artificial neural networks (ANN) training. Scientists used a component-loaded SVM to create a sized particles classification strategy. T The research and sensitivity findings indicates that this hybrid approach is helpful because it uses data from two significant levels in Portugal.

Bujang et al. (2021)[24]machine learning was used to provide a forecast of academic success. The proposed method relied on the student's own initiative to study. To improve prediction of student performance, the fuzzy approach was used. In making the most of a limited set of behavior data containing a massive amount of varied information, effective learning reduces the cost and time of behavior operation.

Ojajuni et al. (2021) [25] comprehensively examined the use of machine learning algorithms to predict long-term academic success early on in the learning process, with the goal of improving prediction accuracy. Synchronous generators would be the primary research object for the duration of the project. Using data from 1282 real students' course marks, the authors of this study evaluated the efficacy of six popular neural network models. Next, the authors devised a method for categorising their findings. Based on the collected data, it seems that integrating the conceptual model with random forest leads to a dramatic boost, with an F-measure of 99.5% being achievable.

Czibula Gabriela et al. (2019)[26]using self-driving computational methods like Logistic Regression (LR) and Support Vector Machine" (SVM). Sequential minimum optimization was shown to be more effective than regression analysis after comparing the results of multiple trials using various methods. The study's findings might also help educators better foresee and evaluate students' future behavior and productivity. The goal was to find ways to predict students' results and provide adequate effective tactics for prioritizing issues like teacher effectiveness and educational incentives that have the greatest potential to reduce student dropout rates.

CriveiGabriela et al. (2020) [27] increasing attention in the field of Educational Data Mining has focused on the challenge of predicting student performance, and this work contributed to that body of research. More and more research shows that both carefully monitored and unsupervised learning methods can help teachers and students learn more about how education works and improve the way they teach and learn. Going to predict a student's final grade in a people oriented on their semester grades is hard because it depends on many things, such as the area of study, the number of tests administered, and how each teacher grades. Researchers also come up with a new classification model called S PRAR to predict how well a student will do in a certain subject (RARs). RARs are better than traditional association rules when it comes to identifying multiple links between data characteristics. In the experiments, there are three real academic data sets.

TABLE I. COMPARISON OF VARIOUS TECHNIQUES OF ACADEMIC PERFORMANCE PREDICTION

Author(s)	Methodology	Results
PADMA et al. (2022) [11]	EDM, DT	The findings will help teachers and mentors break through barriers between classes. The failure rate on the following semester's exams may be reduced thanks to this study's identification of pupils who need further instruction.
Li et al. (2022) [12]	DBNN and ASO	The collected attributes are used to forecast how well pupils will perform on subsequent tasks. Temporal attention is added to the GAN's output to prioritize weekly student behavior.
Yağcı et al. (2022) [13]	EDM, LA	The findings demonstrate that the proposed model does a good job of classifying between 70% and 75% of the time. The estimates were based only on information from midterm, departments, and teachers..
Shreem et al. (2022) [14]	SVM, LDA, DT, K-NN	The results showed that the method that was suggested might make the binary genetic algorithm work better. Also, all of the results from the classifier are improved from between 1% and 11%.
Najm et al. (2022) [15]	K-means clustering	Clustering with expectation maximization showed the highest accuracy, predicting student performance at 96.76% and student grades at 96.12%, relative to all other approaches.
Alachiotis et al. (2022)[16]	ML	Researchers show that a voting generalization technique involving the top three classifiers may outperform the results of a single finely-tuned learning algorithm when the default settings of the learning algorithms are used.
Trakunphutthirak et al. (2022)[17]	LMS	The conceptual model's degree of correctness increases to 92%.
Poudyal et al. (2022)[18][19]	Hybrid CNN	Our hybrid model surpassed all previously published models with an 88% accuracy rate.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
Article Received: 09 October 2022 Revised: 16 November 2022 Accepted: 22 December 2022

Author(s)	Methodology	Results
Silva et al. (2021) [20]	Neural network and AI	High categorization performance rates (average of all programs: 74.98%) proved the value of the factors in predicting degree completion.
Surenthiran et al. (2021) [21]	DBNN and ASO	The false-positive rate drops to 20% and the accuracy of the conceptual model improves to 90%.
Hussain et al. (2021) [22]	DCNN	The accuracy of the conceptual model improves to 92%.
Huang et al. (2021) [23]	SVM and ANN	Student data from two Portuguese primary levels were used in the study and elimination analysis, which both confirmed the efficacy of this hybrid approach.
Tsiakmaki et al. (2021) [24]	EDM, auto ML, LA	The correctness of the conceptual model rises to 91%.
Bujang et al. (2021) [25]	KNN, DT, LR, RF, NB, and SVM	Based on the data, it seems that a maximum F-measure of 99.5% is achieved when the conceptual model is integrated using a random forest.
Ojajuni et al. (2021) [26]	SVM, Extreme Gradient Boosting, and LR	The results demonstrate that Extreme Gradient Boosting can accurately forecast academic performance 97.12% of the time.

III. PROPOSED METHODOLOGY

The Paulo Cortez, University of Minho dataset [30] was used in this study. This article looks at how well two Portuguese high schools do in school. The information was gathered from student reports and teacher surveys. It includes students' academic performance, their demographic and social characteristics, and their school experiences. There are two sets of data about performance in two different areas: “mathematics (mat)” and the “Portuguese language (por)”. In [36] and [37], binary/five-level classification and regression tasks were used to model the two datasets. Important: There is a strong link between the wanted quality G3 and qualities G2 and G1. This is because G3 is the last grade of the year and is given in the “third” period, while G1 and G2 are given in the “first” and “second” periods, respectively. If you don't know what G2 and G1 are, it's harder to guess what G3 will be, but that prediction is much more valuable. We have used set of data in python tool to figure out how students would do. Figure1 shows a flow chart of the work that will be done.

A. Data Set

The data collection we used to make more accurate predictions is presented in figure 2.

B. Data Pre-processing

The python tool was given student.csv data to predict their performance. There are 33 qualities in total. This thirty-three qualities were applied to data preparation in order to improve prediction. In the data pre - processing stage, we used the Random Forest classifier from the attribute selection technique for improved classification. After verifying and managing the dataset's completeness, the following step is to normalise the data. The separation of the dataset into the train dataset and the test dataset is another important step in the data pre-processing procedure. Seventy percent of the data set is used to train the model, and thirty percent is used for evaluation, i.e., to forecast the results. If the same datasets were used for training and testing, the evaluation of the model's performance would not be neutral.

C. Feature Extraction

After the first stage, data preparation, is finished, the process of feature extraction can begin. Feature extraction is the procedure in which a large amount of raw data is distilled down to more manageable sets for processing, as defined via the Oxford English Dictionary. The term "feature extraction" is used to describe methods of choosing and/or combining variables into features, which drastically reduces the amount of data that must be processed while still providing an accurate characterization of the original data set. Principal Component Analysis (PCA) methods were used to extract features for this investigation. By focusing on the relationships between a small numbers of variables, Principal Component Analysis (PCA) can reveal the most significant shifts in the underlying data and thereby reduce the dimensionality of the whole dataset. As part of the process of predicting pupils' academic performance, principal component analysis is used to simplify the underlying data[27][28][29][30][31]. On the described data set, two experiments were conducted to extract features. The data set will be subjected to a PCA analysis in the first experiment to better assess its complexity. The initial attribute space is converted into a two-dimensional space using the PCA, where the two independent principle components indicate the highest variance in the data. There are so many attributes inside the datasets, and PCA combines these attributes into features such as school, sex, age, family, and grade. Therefore, with the help of these features, we can easily classify data, like the school feature can be used to classify how a family can pick a school for their child, and the family feature can be used to classify family size, family status, family relationships with students or children, etc.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

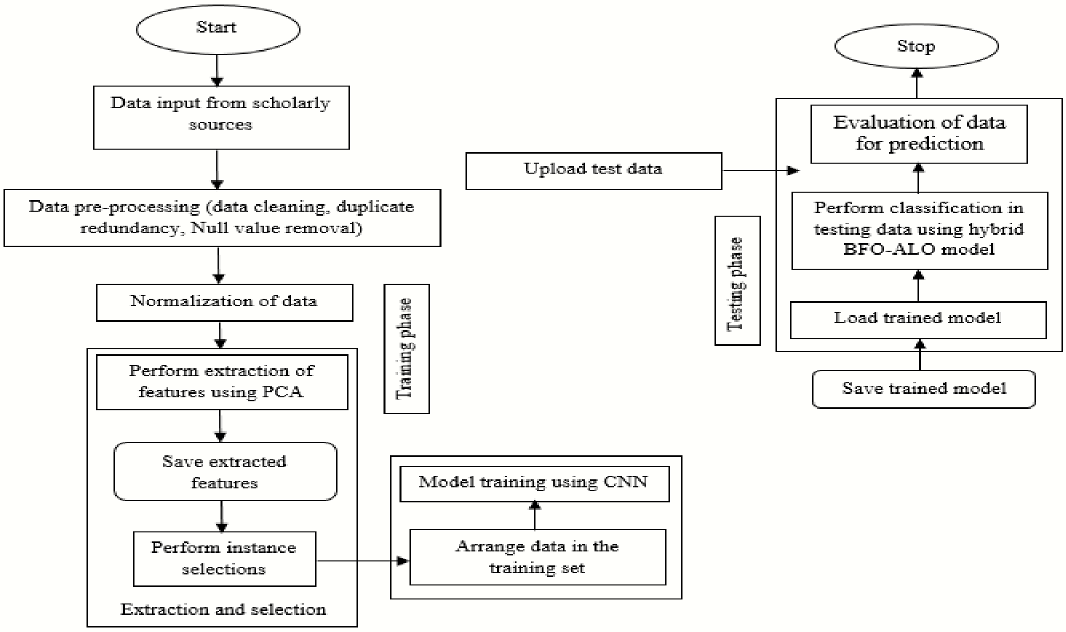


Figure 1. Flow chart of proposed work.

Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:
1 school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2 sex - student's sex (binary: 'F' - female or 'M' - male)
3 age - student's age (numeric: from 15 to 22)
4 address - student's home address type (binary: 'U' - urban or 'R' - rural)
5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education)
8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€“ 5th to 9th grade, 3 â€“ secondary education or 4 â€“ higher education)
9 Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
10 Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
11 reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup - extra educational support (binary: yes or no)
17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
20 nursery - attended nursery school (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no)
22 internet - Internet access at home (binary: yes or no)
23 romantic - with a romantic relationship (binary: yes or no)
24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

these grades are related with the course subject, Math or Portuguese:
31 G1 - first period grade (numeric: from 0 to 20)
31 G2 - second period grade (numeric: from 0 to 20)
32 G3 - final grade (numeric: from 0 to 20, output target)

Figure 2. Attributes Information

D. Instance Selection

As soon as we have finished the feature extraction process, researchers can go on to the next phase, which is selecting instances. By keeping the underlying distribution intact, instance selection ensures that the selected data accurately represents the properties of the whole sample. Instance selection is a straightforward method for removing less reliable characteristics from data sets. Thus, just the five most significant characteristics are passed on to the model training process, and any extraneous ones, such family distribution or alcohol intake in the family, are discarded. Access to relevant instances is essential for learning because it facilitates categorization, and classification is a key component of learning. Adding unrelated events to a simulation can lower its accuracy since they don't add anything to the picture. It is the duty of feature extraction approaches to assess the relevance of extracted characteristics. Therefore, when characteristics are

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
Article Received: 09 October 2022 Revised: 16 November 2022 Accepted: 22 December 2022

extracted from it, selecting instances is the next step in optimising it. In this research, researchers employ the Gini index to pick instances, as will be shown in the next section. Wasteful information is omitted but key instances are preserved using this approach.

- **Gini Index (GI)**

Distributions of data points can be measured and analysed with the use of the GI method [32][33]. This method is used to pick features while simultaneously assessing the validity of characteristics that are associated to classes. A characteristic's purity is measured through how well it can differentiate between different types of data. This method of picking features evaluates whether or not the application of a certain feature produces any solutes before making a final decision. The following equation is used to get the GI for a given t1:

$$GI(t1k) = \Sigma p(t1k|c1l)2nl = 1p(c1l|t1k)2 \quad (1)$$

In the above equation (1), n is the count of classes, $(t1k|c1l)$ is the term $t1k$ possibility provided by the class$c1l$, $(c1l|t1k)$ is the class $c1l$ possibility provided by the label$t1k$.

E. Training Process

After the completion of instance selection process, next step is performed that is training process. As a first step, a training dataset is used to fine-tune the model's parameters. With the help of the training data set, the model is trained using a three-layer CNN model and 200 training iterations. The input vector and its matching output vector make up the training data set, and the response key is sometimes referred to as the objective. For classification purposes, the input from the trained CNN model is now fed into the hybrid BFO-ALO model. The suggested BFO-ALO model for forecasting student performance was tested in experiments to determine its training accuracy, sensitivity, specificity, and F-measure. It specifies a binary classification predicament of '"Satisfactory,' 'Good,' 'Poor,' 'Excellent,' and 'Failure."'[34] We see that the main goal of the problem of predicting whether a student will do better or not is to lower the true positive rate (TPR) or recall (sensitivity) or raise the true negative rate (specificity) (i.e. to maximise the number of students that were correctly classified as failure).

F. Test Process

Once the optimised model has been developed via the process of training, the last step, testing, is performed using a hybrid BFO-ALO model. For hybrid classification, we've employed the BFO and ALO optimization algorithms, which are described in detail below. Unknown test data is currently being provided and analysed in order to complete the forecast. To identify student performances, the trained model is loaded, and unknown data is created based on information travelling between nodes.

- **BFO**

Passino [27] has presented the novel evolutionary computing method known as "bacterial foraging optimization" (BFO). It draws inspiration from the pattern of bacterial foraging behaviour. Chemotaxis is a process that bacteria use to go toward locations that are nutrient-rich. Bacteria are known to move through rotating whip-like flagella that are propelled in a reversible motor encased in the cell wall. 8–10 flagella are dispersed at random on an E. coli cell body. Run is the term for the movement of a cell along a trajectory when all of its flagella rotate counter clockwise. The pulling forces exerted by the flagella as they rotate in a clockwise direction lead the bacteria to tumble. Chemotaxis, swarming, reproduction, and elimination-dispersal are the four sequential mechanisms that make up the majority of the bacterial foraging system. In the study of [35][36][37][38][39][40], each of these procedures is briefly described.

- **ALO**

An innovative, naturally inspired algorithm known as the Antlion Optimizer (ALO), which was first proposed as a result of Seyedali Mirjalili in 2015[41], imitates the natural hunting behaviour of antlions. ALO's five main methods of collecting food include capturing, entrapping ants, catching prey, and repairing taps. According to [39][40], there are six operators in the ALO algorithm, and they are as follows: Trapping in Antlion's Pits, Random Walks of Ants, Sliding Ants towards Antlion, Building Trap, Elitism, Sliding Ants towards Antlion.

The working steps of the BFO-ALO hybrid model is shown below and figure is shown in figure 3.

Algorithm steps of BFO-ALO model:

Step 1: After the retrieved features have been extracted and the model has been trained, the tested dataset is delivered to the BFO-ALO model for classification. Within the hybrid model, there are several processes or stages that must be encountered during classification.

Step 2: Mental health of the student, psychological health of the student, and academic performance of the student were derived from datasets. These three parameters have several attributes that help in data classification and the generation of relevant results. Internally, these parameters now confront several BFO-ALO phases.

Step 3: Now the evaluated dataset will undergo swarm initialization. Here, the BFO model operates independently from the swarm stage's inception until its conclusion. Here, the textual information about swarm that is included inside the dataset is analysed. Swarms are complex and stable spatio-temporal patterns of E.coli bacteria. This bacterium is used to assess the physical fitness of students.

Step 4: Here now fitness value of students are analysing with parameters of mental and psychological health.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
Article Received: 09 October 2022 Revised: 16 November 2022 Accepted: 22 December 2022

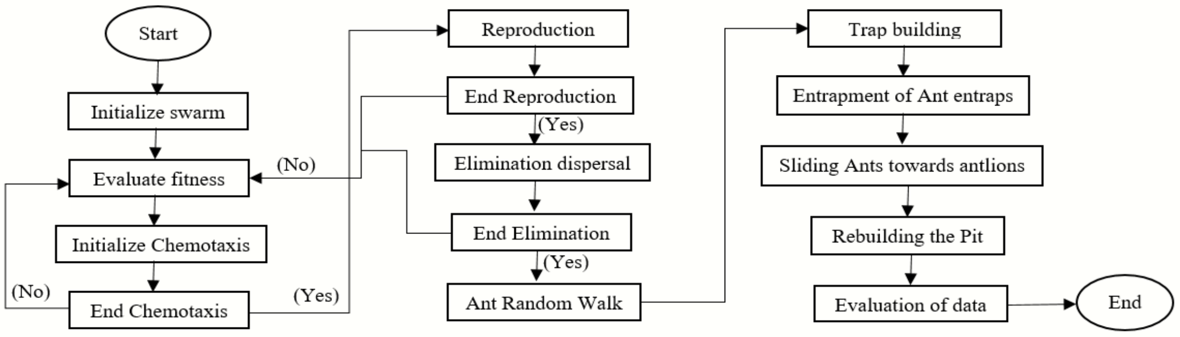


Figure 3. Working of BFO-ALO Hybrid Model

Step 5: Chemotaxis is initiated at this stage. Chemotaxis is the migration of a cell along a gradient of increasing chemical concentration. Now that chemotaxis has been effectively terminated, the process will restart at the assessment of fitness values if any cell is too weak to migrate toward the gradient. This cell mobility is advantageous for reproduction.

Step 6: In the reproduction stage, chemotaxis cell reproduction is occurring, and instead of reproducing sexually, the healthier bacteria perform asexual division, producing two germs that are implanted in the same area as the less healthy bacteria. This method preserves the size of the swarm. Now the reproduction phase of chemotaxis is complete, however if a cell is weakened during reproduction, the process will begin again at the fitness value assessment step. These chemotaxis cells and swarm bacteria are essential for analysing the mental and psychological behaviour of students, which helps in assessing their academic performance.

Step 7: In this step, chemotaxis cells and swarm bacteria are weak and will be destroyed; those who are perfect for the subsequent procedure and will be delivered to the ALO process stages. In the ALO stages, academic data on students will be analysed based on their mental and psychological behaviour, which was acquired in the BFO stages.

Step 8: This stage is an ant random walk, and it is used to generate the parameters from the dataset. The parameters are mental health of the student, psychological health of the student, and academic performance of the student.

Step 9: these are the most important stages for the prediction of student academic performance and the stages are trap building, entrapment of ant entrap, sliding ant towards antlions, and rebuilding the pit. Here dataset is analysed with the help of different parameter’s attributes like family support, family size, student with school relations, student’s personal relations with other student, school environment for students etc.

Step 10: This is the final stage of this hybrid model, and when step 9 has been successfully completed and implemented, the model will produce a number of results, including model accuracy and different distribution graphs of student academic performance prediction, such as distribution graphs for grade as a result of family support, grade as a result of family size, grade as a result of student's relationships, and grade as a result of student and school environment, among others.

IV. EXPERIMENTAL ANALYSIS AND RESULTS

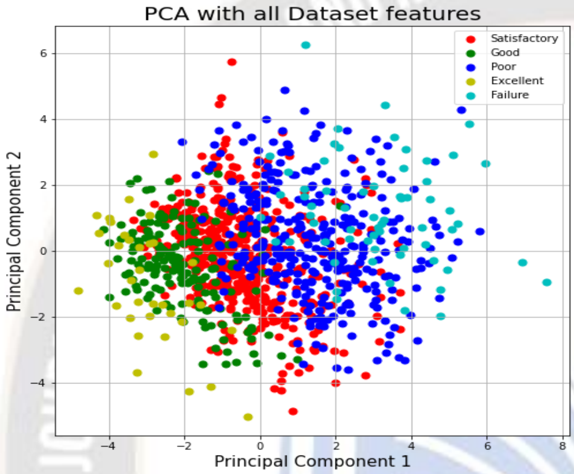
Figure 4 depicts the uploading data in which the training data that will be processed is uploaded. It displays the properties on which the data correlations are evaluated, which is useful for evaluating the planned work during the testing phase for classifications. In order to evaluate the significance of the characteristics, we computed the Pearson correlation coefficients between the attributes age, medu, fedu, travel time, study time, failures, community, relatives, free time, going out, Dalc, Walc, health, G1, G2, and G3 and the final examination grade (Final Grade). A number of 0 indicates that there are no linear links between the variables, while a Pearson correlation coefficient of 1 or 1 implies a linear monotonic relationship between the two variables under comparison. We found that the attributes are moderately well connected with the final examination grade, with the project score of 1.0 (attribute age) having the highest association. For attribute study time, the lowest correlation of -0.007 was found.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

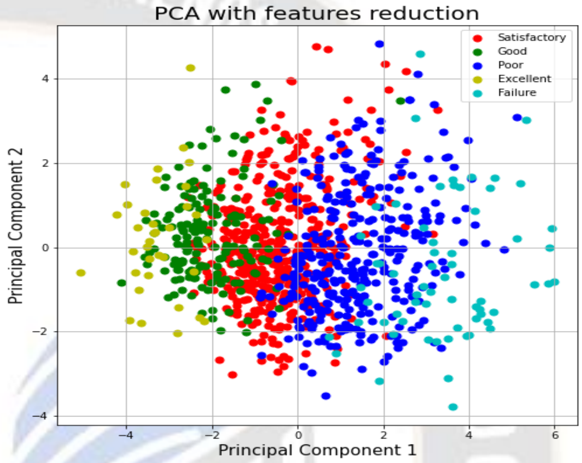
```
[2] train  = pd.read_csv('students.csv')
    train.columns

    Index(['school', 'sex', 'age', 'address', 'famsize', 'Pstatus', 'Medu', 'Fedu',
           'Mjob', 'Fjob', 'reason', 'guardian', 'traveltime', 'studytime',
           'failures', 'schoolsup', 'famsup', 'paid', 'activities', 'nursery',
           'higher', 'internet', 'romantic', 'famrel', 'freetime', 'goout', 'Dalc',
           'Walc', 'health', 'absences', 'G1', 'G2', 'G3', 'subject'],
          dtype='object')
```

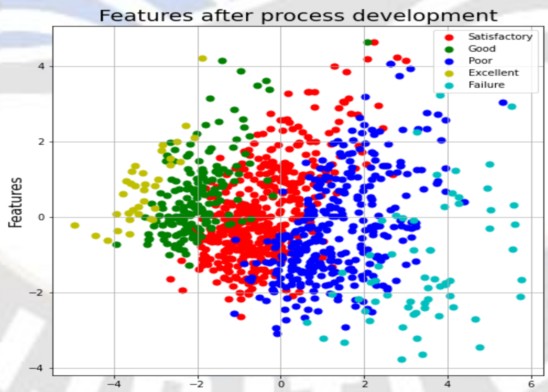
Figure 4. Uploading data



a) PCA feature of Dataset



b) PCA with feature reduction



c) PCA feature after process development

Figure 5. PCA simulation of student's final grade results equal to or more than 4.

There will be two PCA visualisations offered: one for the complete data set (the result is displayed in figure 5a), and the other for the subset of data containing the chosen attribute (the result shown in figure 5b). We want to see if there would be a better separation between the examples if the property with the least correlation to the final exam grade were removed. The implementation from [27][28] is used for the PCA analysis that was applied to the first experiment.

Figure 5c depicts the feature values in terms of correlations such as prior attempts and student credits, revealing the interdependence between an individual's previous attempts and the student credits that he or she earns on the examination. Similarly, the entropy indicates the state of the data points. The third feature displays the past tries alongside the obtained results, demonstrating that the previous attempt affects the individual's obtained outcomes.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

Earlier, it was said that the studies comprised of three distinct scenarios. Figure 5a shows the results of a second experiment in which instance selection was conducted independently of cross-validation and the prediction model was built using an uncompressed dataset.Experiment two (shown in Figure 5b) attempted to replicate experiment one by building a prediction model using a training set filtered by the instance selection method. This experiment replicates a common scenario in which the instance selection strategy is put to use, with the dataset compressed through a method then being used to build a prediction model. The final instance process, shown in figure 6c was performed with the final selected attributes. The detailed view of the test procedure for all comprised scenarios is shown in Figure 6.

Figure 6 depicts the instances selection using PCA feature extraction. In the suggested method, this phase follows the feature extraction procedure. The linear PCA will extract features in the form of eigenvalues and eigenvectors, which will be arranged in the feature vector, and this feature vector will be used for feature optimization in terms of Gini index that indicate the correlations and dependencies between data points in terms of features in the training set.

Figure 7 depicts the classification outcome, which is represented in labels. Observations indicate that the proposed method may classify an individual's test results as 'Satisfactory,' 'Good,' 'Poor,' 'Excellent,' and 'Failure based on student's performance. Assessments are made after the trained model is applied to the test set and produces projected categories based on the test set. In figure 8 the result shows the accuracy projection graph for training and validation for our proposed hybrid model. It demonstrates that our model outperforms earlier research models with the same parameters, and the total accuracy of our model is 95%, and the sensitivity, specificity, accuracy, and F-Measure has been calculated as formulas given below and table 2 is showing the Senstivity, Specificity, Accuracy, and F-Measure values for our proposed work.

$$\text{Senstivity} = \frac{TP}{TP+FN} * 100$$

$$\text{Specificity} = \frac{TP}{TN+FP} * 100$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} * 100$$

F-Measure = (2 * Precision * Recall) / (Precision + Recall)

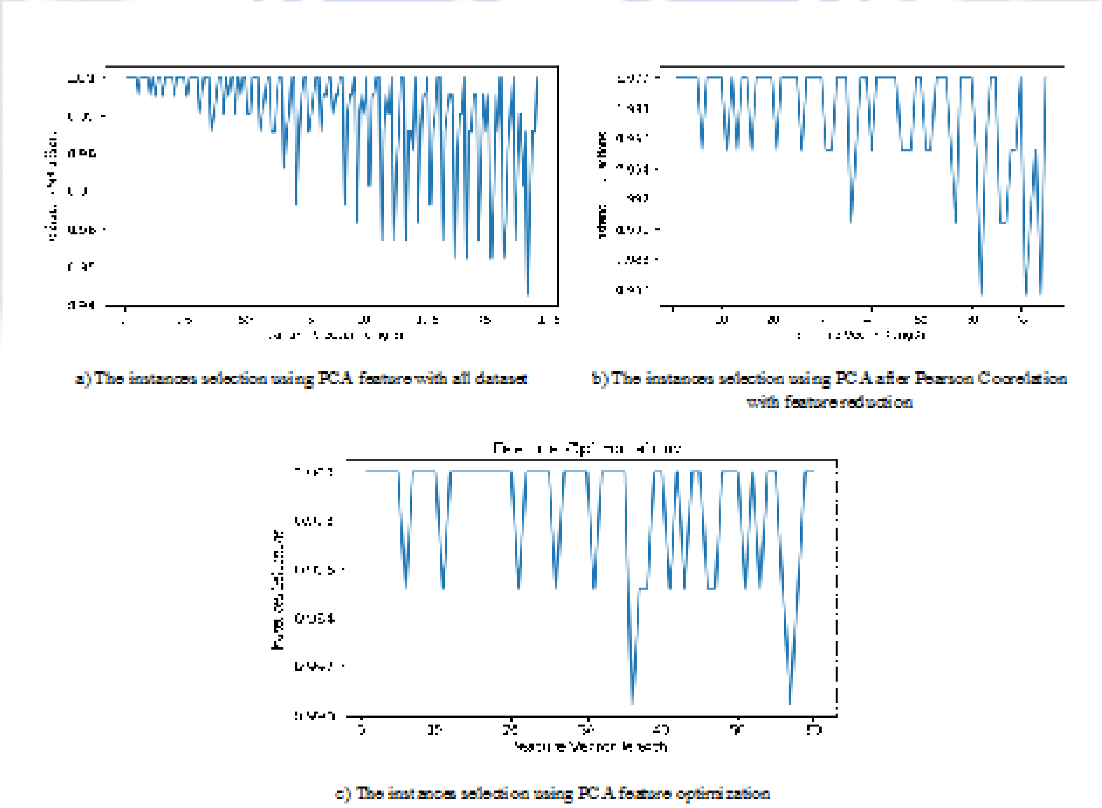


Figure 6. The instance are identified with the attributes (such as age, address, guardian, health, study time, etc.)

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

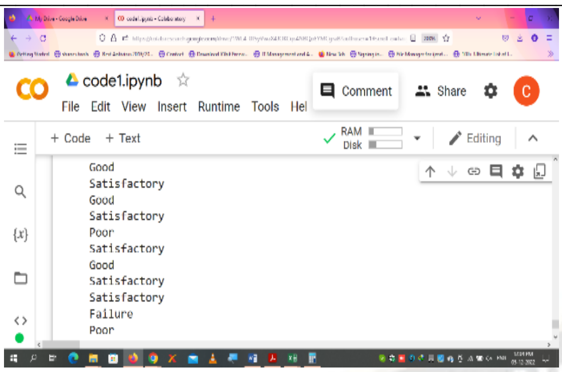


Figure 7. Depicts the classification outcome

Where, "TP represents True Positive", "TN represents True Negative", "FP represents False Positive" and "FN represents False Negative".

TABLE II. EXPERIMENTAL EVALUATION OF BFO-ALO MODEL

Parameters	Values
Accuracy	94%
Specifity	98%
Sensitivity	85%
F-Measure	83%

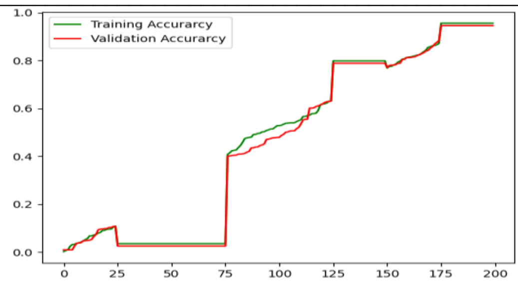
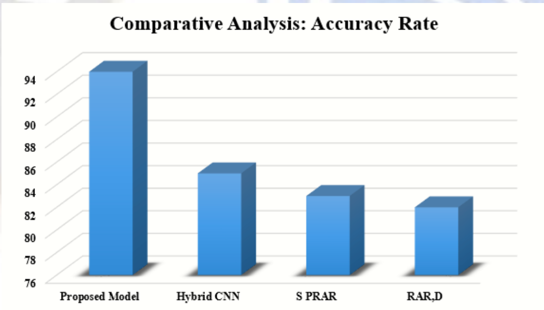


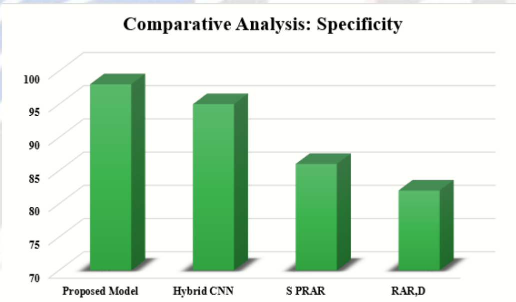
Figure 8. Training and validation accuracy graph

V. COMPARATIVE PERFORMANCE EVALUATION OF PROPOSED BFO-ALO MODEL

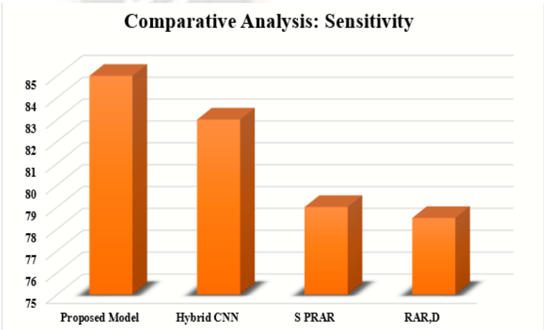
In this part, a comparative evaluation of the proposed BFO-ALO model was conducted using four performance metrics: Accuracy, Sensitivity, Specificity, and F- measure. In figure9, all comparative analysis graphs are obtained using the same parameters as the parameters of the proposed model. Table 3 and Figures 9a, 9b, 9c, and 8d show the comparison analysis graph as performance differences between the proposed model and prior research models for precision, specificity, sensitivity, and F-measure.



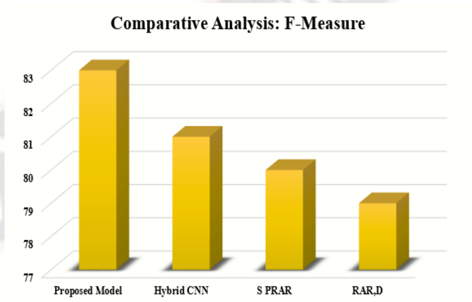
a) Comparative Analysis: Accuracy Rate.



b) Comparative Analysis: Specificity.



c) Comparative Analysis: Sensitivity.



d) Comparative Analysis: F-Measure.

Figure 9. Comparative analysis based on various parameters.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

TABLE III. THE COMPARISON ANALYSIS TABLE AS PERFORMANCE DIFFERENCES BETWEEN THE PROPOSED MODEL AND PRIOR RESEARCH MODELS.

Parameter	RAR, D	S PRAR	Hybrid CNN	Proposed
Accuracy	82%	83%	85%	94%
Specifity	85%	86%	95%	98%
Sensitivity	78.5%	79%	83%	85%
F-Measure	79%	80%	81%	83%

.

Table 3 shows the comparative performance analysis of the proposed BFO-ALO model and the existing RAR,D, S PRAR and Hybrid CNN. It is obvious from the table 4 that the performance of proposed BFO-ALO model outperformed RAR,D, S PRAR and Hybrid CNN with accuracy rate is 94 percent, specificity 98 percent, sensitivity is 85 percent, and F-score or F-measure 83 percent.

VI. CONCLUSION AND FUTURE SCOPE

The purpose of this study was to evaluate how well two deep learning models, BFO and ALO, could be used as intelligent tools for the investigation of academic data sets. The research that was carried out for this study put a spotlight on the ability of deep learning models to recognize trends while analysing the academic performance of students. But the study's authors found that the models' ability to classify cases wasn't as good as it could have been because of likely anomalies in the academic data set and the small number of variables used to describe the cases. But the outcome demonstrates that the accuracy projection graph for training and validation for our proposed hybrid model outperforms earlier research models with the same parameters, and the overall accuracy of our model is 95%.

The experimental assessment will be expanded in the future to include more academic data sets, and further work would be undertaken to understand the intriguing RARs that were mined. Instead of only using two classes, the RAR mining experiment would be broadened so that it may mine crucial rules for the whole range of grades (4–10). (pass and fail). To further enhance mining efficiency, we will investigate methods for identifying outliers and abnormalities within the data sets. It will allow the noise to have less of an influence on the learning process, which will in turn improve the performance. Deep learning categorization models that can accurately predict students' academic achievement are something we are considering pursuing as the logical next step in our study.

ACKNOWLEDGMENT

The author of the manuscript did not receive any funding from the external sources.

REFERENCES

[1] Shahiri, Amirah Mohamed, and Wahidah Husain. "A review on predicting student's performance using data mining techniques." *Procedia Computer Science* 72 (2015): 414-422.

[2] Lee, SeungYeon, HuiSooChae, and Gary Natriello. "Identifying User Engagement Patterns in an Online Video Discussion Platform." *International Educational Data Mining Society* (2018).

[3] Feng, Wenzheng, Jie Tang, and Tracy Xiao Liu. "Understanding dropouts in MOOCs." In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 517-524. 2019.

[4] Huang, Shaobo, and Ning Fang. "Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models." *Computers & Education* 61 (2013): 133-145.

[5] Clarke, Emma, and John Visser. "Pragmatic research methodology in education: possibilities and pitfalls." *International Journal of Research & Method in Education* 42, no. 5 (2019): 455-469.

[6] Galassi, John P. Strengths-based school counseling: Promoting student development and achievement. Routledge, 2017.

[7] Sin, Katrina, and LoganathanMuthu. "APPLICATION OF BIG DATA IN EDUCATION DATA MINING AND LEARNING ANALYTICS--A LITERATURE REVIEW." *ICTACT journal on soft computing* 5, no. 4 (2015).

[8] Ameen, A. O., Alarape, M. A., &Adewole, K. S. (2019). STUDENTS'ACADEMIC PERFORMANCE AND DROPOUT PREDICTION. *Malaysian Journal of Computing*, *4*: 278-303.

[9] Agrawal, H., &Mavani, H. (2015). Student performance prediction using machine learning. *International Journal of Engineering Research and Technology*, *4:* 111-113.

[10] Silva, J., Romero, L., Solano, D., Fernandez, C., Lezama, O. B. P., & Rojas, K. (2021). Model for predicting academic performance through artificial intelligence. In Computational Methods and Data Engineering, 519-525).

[11] PADMA, E., TAPALA YASIN, and MADDIPATI SRIVATSAV. "Students Performance Prediction Using Machine Learning Algorithms."

[12] Li, Gang. "Construction of Sports Training Performance Prediction Model Based on a Generative Adversarial Deep Neural Network Algorithm." *Computational Intelligence and Neuroscience* 2022 (2022).

[13] Yağcı, Mustafa. "Educational data mining: prediction of students' academic performance using machine learning algorithms." *Smart Learning Environments* 9, no. 1 (2022): 1-19.

[14] Shreem, Salam Salameh, HamzaTurabieh, Sana Al Azwari, and FaizBaothman. "Enhanced binary genetic algorithm as a feature selection to predict student performance." *Soft Computing* 26, no. 4 (2022): 1811-1823.

[15] Najm, Ihab Ahmed, Jasim Mohammed Dahr, AlaaKhalafHamoud, Ali Salah Alasady, WidAkeelAwadh, Mohammed BM Kamel, and AqeelMajeedHumadi. "OLAP Mining with Educational Data Mart to Predict Students' Performance." *Informatica* 46, no. 5 (2022).

[16] Alachiotis, Nikolaos S., Sotiris Kotsiantis, EvangelosSakkopoulos, and Vassilios S. Verykios. "Supervised machine learning models for student performance prediction." *Intelligent Decision Technologies* Preprint (2022): 1-14.

[17] Trakunphutthirak, Ruangsak, and Vincent CS Lee. "Application of educational data mining approach for student academic performance prediction using progressive temporal data." *Journal of Educational Computing Research* 60, no. 3 (2022): 742-776.

[18] Chango, Wilson, RebecaCerezo, and Cristóbal Romero. "Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses." *Computers & Electrical Engineering* 89 (2021): 106908.

[19] Silva, Jesus, Ligia Romero, Darwin Solano, Claudia Fernandez, Omar Bonerge Pineda Lezama, and Karina Rojas. "Model for predicting academic performance through artificial intelligence." In *Computational Methods and Data Engineering*, pp. 519-525. Springer, Singapore, 2021.

[20] Surenthiran, S., R. Rajalakshmi, and S. S. Sujatha. "Student performance prediction using atom search optimization based deep belief neural network." *Optical Memory and Neural Networks* 30, no. 2 (2021): 157-171.

[21] Hussain, Sadiq, Silvia Gaftandzhieva, MdManiruzzaman, RositsaDoneva, and ZahraaFadhilMuhsin. "Regression analysis of student academic performance using deep learning." *Education and Information Technologies* 26, no. 1 (2021): 783-798.

[22] Huang, Chenxi, Junsheng Zhou, Jinling Chen, Jane Yang, Kathy Clawson, and YonghongPeng. "A feature weighted support vector machine and artificial neural network algorithm for academic course performance prediction." *Neural Computing and Applications* (2021): 1-13.

[23] Tsiakmaki, Maria, GeorgiosKostopoulos, Sotiris Kotsiantis, and OmirosRagos. "Fuzzy-based active learning for predicting student academic performance using autoML: a step-wise approach." *Journal of Computing in Higher Education* 33, no. 3 (2021): 635-667.

[24] Bujang, SitiDianah Abdul, Ali Selamat, Roliana Ibrahim, OndrejKrejcar, Enrique Herrera-Viedma, Hamido Fujita, and Nor AzuraMdGhani. "Multiclass prediction model for student grade prediction using machine learning." *IEEE Access* 9 (2021): 95608-95621.

[25] Ojajuni, Opeyemi, FolusoAyeni, OlagunjuAkodu, Femi Ekanoye, Samson Adewole, Timothy Ayo, Sanjay Misra, and Victor Mbarika. "Predicting student academic performance using machine learning." In *International Conference on Computational Science and Its Applications*, pp. 481-491. Springer, Cham, 2021.

[26] Czibula, Gabriela, IstvanGergelyCzibula, Diana-Lucia Miholca, and Liana Maria Crivei. "A novel concurrent relational association rule mining approach." *Expert Systems with Applications* 125 (2019): 142-156.

[27] Crivei, Liana Maria, Gabriela Czibula, George Ciubotariu, and Mariana Dindelegan. "Unsupervised learning based mining of academic data sets for students' performance analysis." In *2020 IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pp. 000011-000016. IEEE, 2020.

[28] Miholca, Diana-Lucia, Gabriela Czibula, and IstvanGergelyCzibula. "A novel approach for software defect prediction through hybridizing gradual relational association rules with artificial neural networks." *Information Sciences* 441 (2018): 152-170.

[29] Nagaraj, Kalyan, G. S. Sharvani, and Amulyashree Sridhar. "Emerging trend of big data analytics in bioinformatics: a literature review." *International Journal of Bioinformatics Research and Applications* 14, no. 1-2 (2018): 144-205.

[30] CRIVEI, LIANA. "INCREMENTAL RELATIONAL ASSOCIATION RULE MINING OF EDUCATIONAL DATA SETS." *StudiaUniversitatis Babes-Bolyai, Informatica* 63, no. 2 (2018).

[31] Learning, D. (2020). Deep learning. *High-Dimensional Fuzzy Clustering*.\

[32] Rajesh, Ganesan, X. MercilinRaajini, R. AshokaRajan, M. Gokuldhev, and C. Swetha. "A multi-objective routing optimization using swarm intelligence in IoT networks." In *Intelligent Computing and Innovation on Data Science*, pp. 603-613. Springer, Singapore, 2020.

[33] Dong, Dawei, Zhiwei Ye, Yu Cao, ShiweiXie, Fengwen Wang, and Wei Ming. "An improved association rule mining algorithm based on ant lion optimizer algorithm and FP-growth." In 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), vol. 1, pp. 458-463. IEEE, 2019.

[34] Czibula, Gabriela, Andrei Mihai, and Liana Maria Crivei. "S PRAR: A novel relational association rule mining classification model applied for academic performance prediction." Procedia Computer Science 159 (2019): 20-29.\

[35] Poudyal, Sujan, Mahnas J. Mohammadi-Aragh, and John E. Ball. "Prediction of Student Academic Performance Using a Hybrid 2D CNN Model." Electronics 11.7 (2022): 1005.

[36] Akour, M.; Alsghaier, H.; Alqasem, O. The effectiveness of using deep learning algorithms in predicting students achievements. Indones. J. Electr. Eng.Comput. Sci. 2020, 14, 388–394.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 10 Issue: 2s
DOI: https://doi.org/10.17762/ijritcc.v10i2s.5913
***Article Received**: 09 October 2022 **Revised**: 16 November 2022 **Accepted**: 22 December 2022*

[37] Cortez, Paulo, and Alice Maria Gonçalves Silva. "Using data mining to predict secondary school student performance." (2008).

[38] Hu, Hongping, Yangyang Li, Yanping Bai, Juping Zhang, and Maoxing Liu. "The improved antlion optimizer and artificial neural network for Chinese influenza prediction." Complexity 2019 (2019)..

[39] Majhi, Ritanjali, Ganapati Panda, Babita Majhi, and Gadadhar Sahoo. "Efficient prediction of stock market indices using adaptive bacterial foraging optimization (ABFO) and BFO based techniques." Expert systems with applications 36, no. 6 (2009): 10097-10104.

[40] Passino, Kevin M. "Biomimicry of bacterial foraging for distributed optimization and control." *IEEE control systems magazine* 22, no. 3 (2002): 52-67.

[41] S. Mirjalili, "The ant lion optimizer," Advances in Engineering Software, vol. 83, pp. 80–98, 2015.