Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Student performance prediction with BPSO feature selection and CNN classifier

Safira Begum [1,*], Sunita S. Padmannavar [2]

[1]Department of Computer Applications, Visvesvaraya Technological University–RRC, Belgaum, India
[2]Department of Computer Applications, KLS Gogte Institute of Technology, Belgaum, India

ARTICLE INFO

Article history:
Received 2 April 2022
Received in revised form
18 June 2022
Accepted 29 July 2022

Keywords:
Binary particle swarm optimization
Convolution neural network
Data mining
Deep learning
Educational data mining

ABSTRACT

Educational Data Mining (EDM) is gaining great importance as a new interdisciplinary research field related to some other areas. It is directly related to data mining (DM), the latter being a fundamental part of knowledge discovery in databases (KDD). This data is growing more and more and contains hidden knowledge that could be very useful for users (both teachers and students). It is convenient to identify such knowledge in the form of models, patterns, or any other representation scheme that allows better exploitation of the system. Data mining is revealed as the tool to achieve such discovery, giving rise to EDM. In this complex context, different techniques and learning algorithms are usually used to obtain the best results. Recently educational systems are adopting artificial intelligent systems, especially in the educational context, specific areas for extracting relevant information, such as EDM, which integrates numerous techniques that support the capture, processing, and analysis of these sets of records. The main technique associated with EDM is Machine Learning, which has been used for decades in data processing in different contexts, but with the advent of Big Data, there was an intensification in the application of this technique to extract relevant information from a huge amount of data. This paper proposes the student performance prediction using CNN (Convolution Neural Network) and BPSO (Binary Particle Swarm Optimization) based feature selection method. In this study, classifiers are made for 2-class and 5-class predictions. The proposed system claims an outperforming accuracy of 96.6% with various previous research works as well as found that the majority of attributes related to school activities as compared to data on demographic and socioeconomic characteristics.

1. Introduction

In recent years, education has changed as a result of technological advances available that are directed to the instrumentation of the educational sector, both in software aimed at teaching and in the digital administration of academic records by institution managers. Data Mining is an area of study whose objective was to extract patterns and relevant knowledge from the data. One of the ways to extract these patterns is to use techniques of Learning from Machines, a sub-area of Artificial Intelligence, whose main objective is to develop computational techniques capable of acquiring knowledge automatically. These patterns can be useful to build predictive models to assist in decision-making processes associated with student dropout and their performance analysis. The mining area of educational data is constantly growing. More focus has been given to the jobs area for predicting student environments of learning, where there is constant information on student activities is being recorded (Salal et al., 2019). It is among those that have received the most attention from the academic community. Some research works utilized information about the student's performance before joining high school (Khan and Ghosh, 2021). However, few works have been carried out in the context of on-site technical courses (Andrade et al., 2021). Machine Learning and Statistics can explore student data and detect patterns, and hidden factors that characterize the behavior and performance of students (Mangina and Psyrra, 2021). Various techniques in data mining have been used to predict

* Corresponding Author.
Email Address: safirabgm@gmail.com (S. Begum)
https://doi.org/10.21833/ijaas.2022.11.010
Corresponding author's ORCID profile:
https://orcid.org/0000-0003-2883-9994

performance, for example, Decision Tree, Naive Bayes, KNN (K-Nearest Neighbour), regression models, and Artificial Neural Networks (Bhatia, 2019). Traditional data mining systems (Namoun and Alshanqiti, 2020) may not be better suited to the problem, due to the fact that there is no linearity between the factors that involve prediction or academic performance. In recent years, the trend of Deep Learning (DL) has been gaining strength, especially in the ability to identify parents on a complex basis (Kriegeskorte and Golan, 2019). DL showed great potential in various data mining tasks related to classification, and regression expanding the applications of the Artificial neural network (ANNs), widely used in the area of EDM and Learning Analytics (LA) (Kriegeskorte and Golan, 2019).

The increase in the volume of data was accompanied by the development of new techniques for EDM, such as the resumption of studies with Deep Learning (Zhao et al., 2019; Yang et al., 2019) point out four gaps in the EDM area:

1. Unavailability of consistent data sets that are large enough to analyze the educational system and its performance
2. Need for integration and versatility in sets of data
3. Fewer works were performed using hybrid techniques
4. Need to compare methods

In this context, this study aims to predict student performance; for this, a common dataset from the UCI machine learning repository is utilized, with this it was possible to compare techniques already consolidated in the scope of the EDM, with the machine learning technique. The approach used was supervised learning for classification (Ajibade et al., 2018), in which students' grades are predicted, but these were divided into four categories, their numerical values not being used. With these predictions, it was also possible to verify if the attributes that make up the database are sufficient to generate effective models in predicting student performance, in addition to evaluating BPSO-based feature selection to ensure redundancy management and proper training in supervised classification. Finally, this study intends to make available to those interested in the area, a document that presents in detail how to carry out the educational data mining process. To this end, this document is structured as follows: Literature review is presented in section 2; section 3 deals with the main aspects of the pre-processing and feature selection techniques, as well as the architecture of the CNN classifier used in this study; section 4 presents the results achieved with this investigation; finally, section 5 describes the authors' conclusions with the development of this study.

2. Literature review

Different types of EDM (Educational Data Mining) have been proposed in a bibliographic study with the objective of higher accuracy via optimal feature selection and deep learning (Hussain et al., 2019). Chaudhury et al. (2016) claimed that EDM is defined as the area of research that focuses on the development of techniques to explore sets of data collected in educational settings. According to the authors, the nature of these data is different from that observed in the data traditionally used in mining tasks, demanding adaptations and new approaches. At the same time, this diversity in the data represents a potential for the implementation of fundamental resources to help improve education. Therefore, techniques and tools are needed to assist in the task of verifying, interpreting, and relating these data, in order to generate useful and relevant knowledge, which, according to Shetu et al. (2021) was already a goal of DM techniques, employed to identify patterns of behavior and find insights that lead to improvements in products and services.

Various research works (Hamoud, 2016; Pojon, 2017) were fundamental for an improvement and understanding of the evolution of EDM in the course of its consolidation as a research area and are summarized in the sequence. Ünal (2020) provided an overview of the data mining techniques that were used to predict student performance, in publications dated between 2002 and 2015. The study is also based on focused on how prediction algorithms could be used to identify the most among the diversity of student data. The authors of Athani et al. (2017) also came to the conclusion that Machine Learning was the most used technique and regarding the effectiveness of the algorithms the Neural Networks had the greatest precision (98%) for predicting student performance, followed by Decision Trees (91%), then Support Vector Machines and KNN with the same efficiency (83%), finally, the least accurate was Naive Bayes (76%).

Ma and Zhou (2018) proposed a new model to analyze the learning profile and engagement of students in an optimized support vector machine and to validate its accuracy the authors compared it with widely used algorithms–Linear Regression and Support Vector Machines. In addition to this example, the research by Singh et al. (2020) aimed to predict student performance, using a Learning Neural Network Profundo, which was compared to Logistic Regression algorithms (Srivastava et al., 2020).

Deep Learning (DL), or Deep Learning, can be defined as a subarea of Machine Learning, characterized by the use of several layers of information for feature extraction, transformation, and parent analysis. It is also characterized the use of learning algorithms that seek to identify relationships between the given through statistical models (Hussain et al., 2021).

3. Methodology

The use of hybrid methods can be considered the one that most coincides with the reality of experiments carried out in the EDM area today, as many researchers have used EDM techniques. The flow diagram of the proposed approach is shown in Fig. 1.

3.1. Feature extraction

3.1.1. Participants

This work is framed within the field of analytics applied to education. The goal is to propose a model through machine learning algorithms using MATLAB, in order to establish the possibility of predicting the academic performance of students in secondary education in two Portuguese schools. In the first step, data were collected from the public data repository UCI Machine Learning (UCI, 2014). Student attributes include grades, and demographic characteristics (social and school) and were gathered through school reports and questionnaires. Two sets of data were provided of 1044 students related to performance in two different subjects: Mathematics and Language.

Feature extraction (Zaffar et al., 2020) is an important step in classification because the effectiveness of a learning model depends on input variables (substantial features) that describe student characteristics and can be used to predict student performance.

This data refers to the performance of secondary school students in two Portuguese schools. Attribute data (including student grades, demographic, social, and academic characteristics) were collected using school reports and questionnaires. Two sets of performance data are provided in two different subjects: Mathematics (Math) and Portuguese (Po). In Athani et al. (2017), both datasets were modeled using binary/five-level classification and regression assignment. Important note: The target G3 attribute has a strong correlation with the G2 and G1 attributes. This is because G3 is the final value (issued in the 3rd period) and G1 and G2 are the values of the 1st and 2nd periods. It is more difficult to predict G3 without G2 and G1, but these predictions are much more useful.

3.2. Attribute selection by BPSO

Binary particle swarm optimization (Kumar and Bharti, 2019) is used to select a subset of M features from a set of N features of a database, (where M<N), in order to reduce redundancy in the database so that optimal results can be achieved. BPSO finds optimal results from the search space of the candidate solution using the equation below:

$$v' = w * v + c_1 r_1 (xP_{best} - x(t)) + c_1 r_1 (xG_{best} - x(t)) \quad (1)$$
$$x(t+1) = x(t) + v' \quad (2)$$

where,
$x(t+1)$: New position of the particle in search space
$x(t)$: Current position of the particle in search space
w: Intertia (assumed to be 0.8)
v: Velocity of particle
c1, c2: Cognitive and social attraction coefficients (assumed to be 1.414)
r_1, r_2: Random numbers $\in (0,1)$
xP_{best}: Local best position
xG_{best}: Global best position

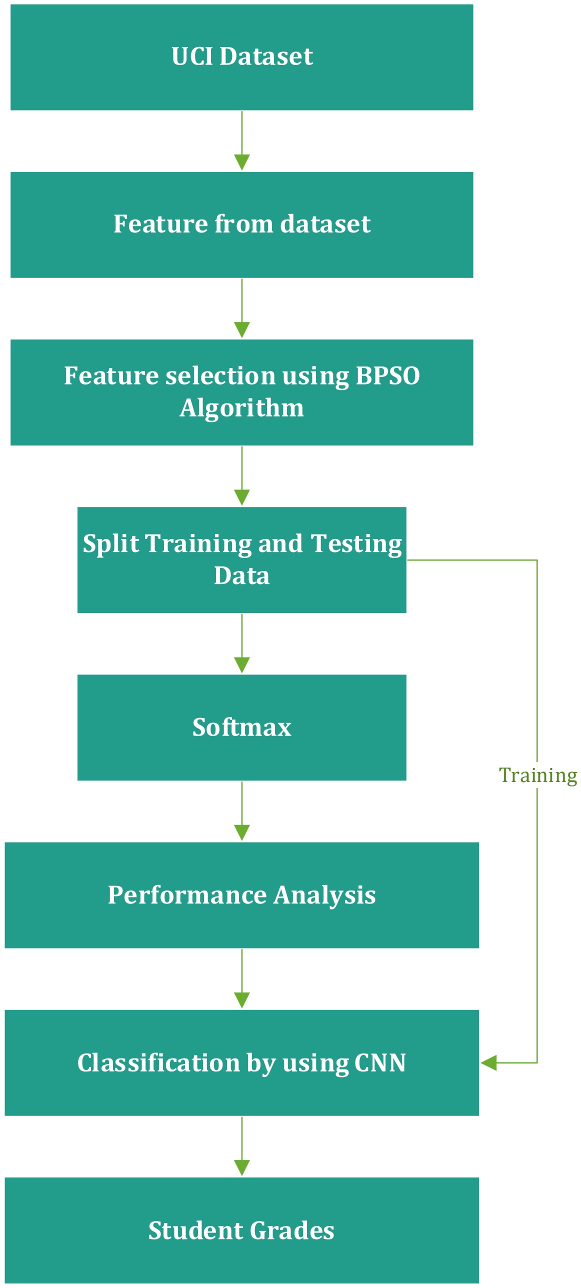

Fig. 1: Flow diagram of the proposed research

3.2.1. Search space and particle coding

Very first we need to create a search space where all possible candidate solutions can be defined. Let's a feature has N attributes i.e., $A = \{A_1 A_2 A_3 \ldots \ldots A_{N-1} A_N\}$ then we need to code each candidate solution to a binary string of length N. we represent each attribute by 1 or 0 to show its presence in the selected subset where the MSB of the binary string represents A_1 and LSB represents A_N. To perform a valid training of classifier we need at least one attribute of data so the lower boundary of search space becomes 1 coded in the binary string of length N. If all the attributes are selected (worst case) for training, in that case, the upper boundary of the search space becomes 2^N coded in the binary string of length N. For example, if the database has a feature of 7 attributes $A = \{A_1 A_2 A_3 \dots \dots A_7\}$ then N becomes 7 and the total possible subset are 2^7= 128. If a candidate solution is [0 0 1 0 0 1 1] then selected subset will be $S = \{A_3 A_6 A_7\}$.

3.2.2. Initialization

Binary particle swarm optimization initializes n uniformly distributed random binary numbers between 1 and 2^N. All candidate solutions are valid and satisfy all constraints of the fitness function.

3.2.3. Fitness function

Fitness Function receives database S of selected attributes and divides data into two sub-databases, one used for training of classifier and the second used for testing. The Accuracy of the testing database is the output of the fitness function which has to be maximized.

$$fitness = \frac{\alpha}{n} \tag{3}$$

where,
α: Correctly classified test cases
n: Total test cases

3.2.4. Update swarms

Swarms are updated using Eqs. 1 and 2. In every iteration each candidate is compared with its value in the previous iteration if the new value gives better accuracy, it is updated otherwise the old value is retained. The best among all is known as global minima. The BPSO approach is presented in Fig. 2.

3.3. Classification by convolutional neural network

A convolutional neural network works by detecting simple information in its first layers (Phan et al., 2018), to later combine this information in the deep layers of the network. The origin of convolutional neural networks lies in their similarity to the primary visual cortex of biological brains. In a biological brain, there are neurons specialized in the detection of simple patterns grouped in receptive fields. Activation by a stimulus to this visual cortex spreads through these receptive fields, specialized in simple tasks such as line and edge detection, and later to complex task fields such as rotation of the visual stimulus, in which neurons necessary are activated.

Convolutional neural networks have played a very vital part in the history and evolution of artificial neural networks. They are a clear example of using the biological and physiological study of the brain (CNNs present similarities with human vision) for the development of artificial algorithms within the area of machine learning and more specifically deep learning. They were also one of the first models of neural networks to obtain good results and performance, already being used at the end of the last century for the development of commercial applications.

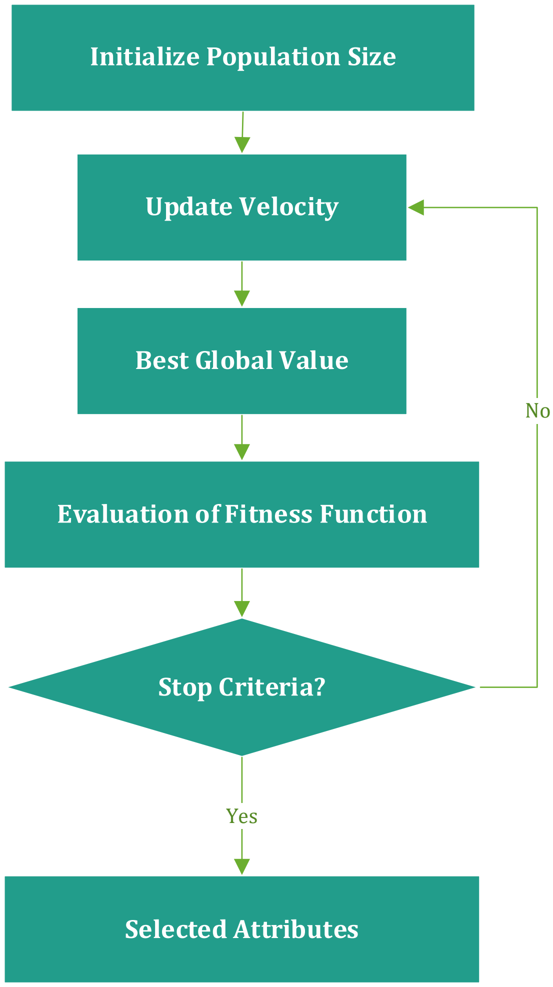

Fig. 2: Selection of attributes by a BPSO algorithm

The structure of a typical convolutional neural network consists of the following types of layers and algorithms (Syu et al., 2018).

Convolutional layer: Convolution is an operation in which an input image is received and a filter or "kernel" is subsequently applied to obtain a characteristic map. The convolution, in turn, provides the reduction of the input image parameters due to the smaller size of the applied filter compared to the original input. Fig. 3 provides a convolution operation. The size of the feature map is managed by three parameters, which are expressed as follows:

1. Depth refers to the number of filters applied on the same convolutional layer.
2. Step or "Stride": It is the number of pixels with which we slide the filter or kernel over the input image matrix. For example, if we move 2 pixels at a time, the stride has a value of two.

3. Zero-padding: This means filling the matrix of the input image with zeros around the border of the image so that there are no calculation errors when applying the filter to the pixels that form the border.

1	1	1	0	0
0	1	1	1	0
0	0	1	1	1
0	0	1	1	0
0	1	1	0	0

Image

4	3	4
2	4	3
2	3	4

Convolved feature

Fig. 3: Convolution operation (Singh et al., 2020)

After each convolution operation, an additional operation called ReLU (Rectified Linear Unit) is usually applied. This is a non-linear operation and is given by:

$$Output = \max(0, Input) \tag{4}$$

ReLU is applied to each pixel and replaces all pixels with negative values with values equal to 0 (Srivastava et al., 2020). The purpose of ReLU is to introduce non-linearity in the convolutional neural network because the convolution operations are linear. Research on activation functions indicates that ReLU is much faster for deep network training (Hussain et al., 2021). The algorithm for a convolutional layer to generate an output volume from an input volume is as follows:

Given an input volume of size $W_1 \times H_1 \times D_1$ where W is the width, H is the height and D is the depth; Using the hyperparameters K (number of filters), F (size of the receptive field or filter), S (value of stride) and P (value of zero padding), an output volume of size $W_2 \times H_2 \times D_2$ is produced where:

$$W_2 = \frac{(W_1 - F + 2P)}{S+1} \tag{5}$$

$$H_2 = \frac{(H_1 - F + 2P)}{S+1} \tag{6}$$

$$D_2 = K \tag{7}$$

Pooling layer: It is placed after the convolutional layer. Its use is to reduce the height and width dimensions of your input for the next convolutional layer. The operation used in this project is max-pooling. It is shown in Fig. 4.

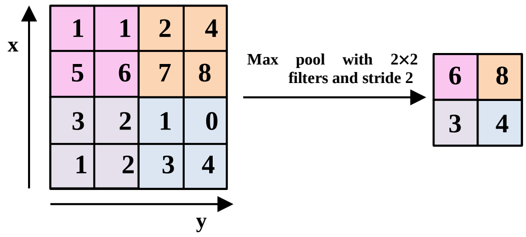

Fig. 4: Max-pooling operation (Singh et al., 2020)

This operation divides the input into rectangles or squares as the case may be and, according to each region generated, creates an output in which the maximum value of each region is included, as can be seen in Fig. 4.

The algorithm for a pooling layer to generate an output volume from an input volume is as follows:

Given an input volume of size $W_1 \times H_1 \times D_1$ where W is the width, H is the height and D is the depth; Using the hyperparameters F (receptive field or filter) and S (stride value), an output volume of size $W_2 \times H_2 \times D_2$ is produced where:

$$W_2 = \frac{(W_1 - F)}{S+1} \tag{8}$$

$$H_2 = \frac{(H_1 - F)}{S+1} \tag{9}$$

$$D_2 = D_1 \tag{10}$$

Fully connected layer: At the end of all the layers of the network, this layer is placed, which will have as many neurons as the number of classes to be predicted.

Finally, Fig. 5 shows a minimal structure of a convolutional neural network with the layers described above.

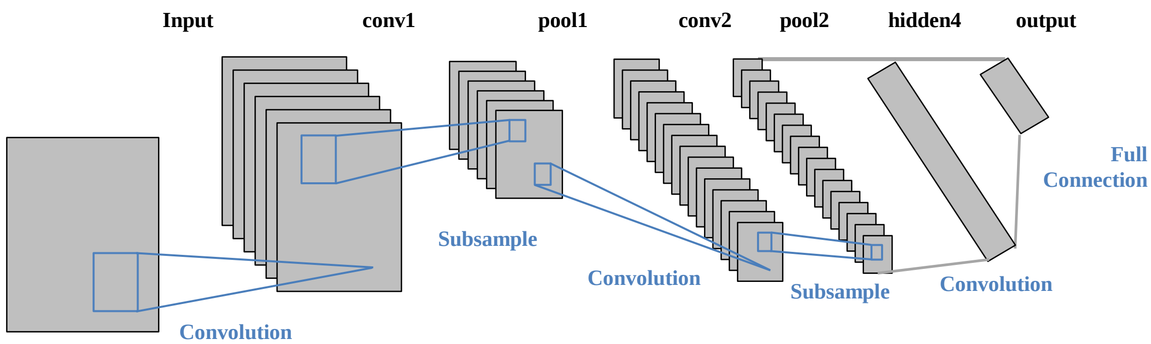

Fig. 5: Typical architecture of a deep convolutional neural network (Syu et al., 2018)

In this last layer, the softmax function is used, which "compresses" the elements of a vector into values in the range [0, 1]. This function allows representing the outputs of the convolutional neural network as a probability distribution (He et al., 2020). The softmax function is defined as:

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \quad (11)$$

For all $j = 1, 2, \ldots, K$. The softmax function normalizes a vector z of real values to a vector $\sigma(z)$ where all its components add up to 1, turning it into a vector of probabilities.

4. Simulation results

UCI dataset: In the first step, data were collected from the public data repository UCI Machine Learning (UCI, 2014). Second Step: Resource extraction and data cleaning (Pre-Processing and Transformation).

Regarding the second stage, the data were pre-processed to suit the application of the EDM techniques-ML algorithms-used to predict student performance:

1. Joining the data of the students of the two subjects in a single base, for all the students–as the data available in UCI Machine Learning was divided into two databases data (one for math subject data and another for Portuguese data), a script was created to join the data in a single base.
2. Transformation of the G3 attribute from numeric to classification levels-in this procedure each range of notes received a value in character format, assigning a category/class for records: Grades between 20 and 16="A," grades between 15 and 11="B," grades between 10 and 4="C," scores between 4 and 0="D." This approach was inspired by the study developed by Yang et al. (2019), the authors who also aimed to predict the student performance, more specifically the prediction of student grades. as detailed in the related work section the authors divided the student results into classes: From 85 to 100-A; from 70 to 85-B; from 60 to 70-C; and >60–D, therefore becoming a classification problem. In this sense, this approach was considered quite effective and adherent to the study developed here, so we chose to do it in a similar way, performing the categorization of grades into four classes A, B, C, and D.

This step in this study intends to evaluate the effectiveness of predicting the performance of the models generated by the algorithms submitted to the databases. To verify the results of a model of classification, two items are needed: The evaluation methods and the interpretation metrics, the two must be applied together so that it is possible to observe whether a model is effective or not.

The methods indicate how this model will be evaluated, and the metrics translate the results of the application of these methods to numbers that can be interpreted.

For this study, the evaluation method used was Training and Testing, in which the data is randomly divided into two portions, one for training and one for testing, so generally, 85% of the instances are for training and 15% to test. The algorithm, when applied to the training base, collects information about the attributes of the instances and generates a classification or regression model based on those attributes and information, after that this model is applied to the test base (which contains records different from the training base) and then the evaluation metrics are calculated on that application.

A model was generated for the BPSO-CNN algorithm used in this study, from its application in the maths database, it should be noted that for the application of the algorithms all attributes and BPSO selected features were used in the database for 2 classes described in Table 1, and all attribute and optimized attributes used in the final grade classified in 5 categories: A, B, C, D, and F. The proposed methods give higher accuracy of 93.3 % for 2 classes whereas 5 classes prediction gives 86.21 %.

A model was generated for the BPSO-CNN algorithm used in this study, from its application in the Portuguese database, it should be noted that for the application of the algorithms all attributes and BPSO-selected features were used in the database for 2 classes described in Table 1, and all attribute and optimized attributes used in the final grade classified in 5 categories: A, B, C, D, and F. Various results of the proposed method are provided in Tables 2-6. The proposed methods give higher accuracy of 96.67% for the 2 classes whereas 5 classes prediction gives 86.11 %.

Table 7 and Table 8 present a comparative analysis of the proposed work with previous work for UCI Maths and UCI Portuguese datasets respectively.

Table 1: BPSO parameter

Inertia factor	Min=0.95, Max=0.99
No of population	50
Acceleration constant	C1=c2=0.25
Maximum iteration	50

Table 2: Selected features

No. of features without optimization	No. of features with BPSO
32	27

Table 3: 2 classes Maths

Parameters	Features	BPSO features
Accuracy	0.9114	0.9333
Error	0.0886	0.0667
Sensitivity	0.9114	0.9333
Specificity	0.9114	0.9778
Precision	0.9114	0.9427
False Positive Rate	0.0886	0.0222
F1_score	0.9114	0.9328
Matthews Correlation Coefficient	0.8228	0.9152
Kappa	0.8228	0.8222

Table 4: 5 classes Maths

Parameters	Features	BPSO features
Accuracy	0.8440	0.8621
Error	0.1560	0.1379
Sensitivity	0.9025	0.8611
Specificity	0.6101	0.9521
Precision	0.9025	0.8958
False Positive Rate	0.3899	0.0479
F1_score	0.9025	0.8603
Matthews Correlation Coefficient	0.5126	0.829
Kappa	0.5126	0.6322

Table 5: 2 classes Portuguese

Parameters	Features	BPSO features
Accuracy	0.9385	0.9667
Error	0.0615	0.0333
Sensitivity	0.9385	0.9667
Specificity	0.9385	0.9889
Precision	0.9385	0.9682
False Positive Rate	0.0615	0.0111
F1_score	0.9385	0.9669
Matthews Correlation Coefficient	0.8769	0.9562
Kappa	0.8769	0.9111

Table 6: 5 classes Portuguese

Parameters	Features	BPSO features
Accuracy	0.8559	0.8611
Error	0.1441	0.1389
Sensitivity	0.9100	0.8611
Specificity	0.6398	0.9537
Precision	0.9100	0.8875
False Positive Rate	0.3602	0.0463
F1_score	0.9100	0.8601
Matthews Correlation Coefficient	0.5498	0.8262
Kappa	0.5498	0.6296

To compare the results obtained with the methods proposed in this work with the results obtained in other works, tests were carried out with the methods of classification used in some works. This comparative study was carried out as described. The proposed BPSO-CNN method for Mathematics and Portuguese datasets outperforms traditional CNN with an improvement of 1.6 % in accuracy. Whereas the Logistic Regression based method gives 62.05% and 67.69% in mathematics and Portuguese dataset respectively, which is very low compared to the proposed method.

5. Conclusion

The main objective of this study was to predict the performance of students using a public dataset and compare the prediction effectiveness of the model generated by CNN. A hybrid structure of CNN-BPSO was the most accurate. With that, it was possible to verify that a database composed of records of notes, frequency, and characteristics demographics, (social and school) is sufficient to carry out the generation of effective models in the prediction of student performance.

Table 7: Performance analysis of existing and proposed BPSO-CNN feature selection with different classifiers for UCI Maths data

Methods	Accuracy	Selected features
Logistic Regression (SVM) (Mason et al., 2018)	62.05%	"Sex, Fedu, sex, Pstatus, sex, Mjob, sex, study-time, age, reason, Medu, sex, guardian, sex, Fjob, famsize, address, travel-time, sex, sex"
KNN (Deepika and Sathyanarayana, 2019)	62.3%	"Internet, higher, Fjob, Pstatus, nursery, activities, sex, Mjob, famsize, address, schools-up, Medu, Fedu, age, travel-time, paid, reason, failures, study-time"
Proposed CNN	91.14%	"freetime, famrel, school, romantic, guardian, higher, studytime, famsup, internet, age, nursery, Medu, Fedu, paid, activities, Mjob, Fjob, address, Pstatus, schoolsup, famsize, reason, failures, sex, traveltime"
Proposed BPSO-CNN	93.33%	"freetime, famrel, school, higher, studytime, famsup, internet, age, nursery, Medu, Fedu, paid, activities, Mjob, Fjob, Pstatus, schoolsup, famsize, reason, failures, sex, traveltime"

Table 8: Performance analysis of existing and proposed BPSO-CNN with different classifiers for UCI Portuguese data

Methods	Accuracy	Selected Features
Logistic Regression (SVM) (Mason et al., 2018)	67.69%	"sex, traveltime, address, age, studytime, famsize, Mjob, guardian"
RFBT-RF (SVM) (Deepika and Sathyanarayana, 2019)	66.92%	"freetime, famrel, school, romantic, guardian, higher, studytime, famsup, internet, age, nursery, Medu, Fedu, paid, activities, Mjob, Fjob, address, Pstatus, schoolsup, famsize, reason, failures, sex, traveltime"
CNN (Turabieh, 2019)	95%	"freetime, famrel, school, romantic, guardian, higher, studytimeage, nursery, Medu, Fedu, paid, Fjob, address, Pstatus, schoolsup, famsize, reason, failures, sex, traveltime internet higher"
Bagging Classifier (Malini, 2021)	86%	------
Proposed CNN	93.85 %	"freetime, famrel, school, romantic, guardian, higher, studytime, famsup, internet, age, nursery, Medu, Fedu, paid, activities, Mjob, Fjob, address, Pstatus, schoolsup, famsize, reason, failures, sex, traveltime"
Proposed BPSO-CNN	96.6%	"freetime, famrel, school, higher, studytime, famsup, internet, age, nursery, Medu, Fedu, paid, activities, Mjob, Fjob, Pstatus, schoolsup, famsize, reason, failures, sex, traveltime"

Furthermore, it was possible to identify that the attributes referring to the student grades and absences are more predictive of performance than student records. Regarding the performance prediction, the proposed EDM technique was adequate, in which the results achieved are the following: CNN with an accuracy of 91.14% in 2 classes of Maths data, BPSO optimized features with 93.33%; Where 5 classes Maths with 84.40%; CNN-BPSO with 86.21%. Similarly, CNN with an accuracy of 93.85% in 2 classes of Portuguese data, BPSO optimized features with 96.6%; where 5 classes Portuguese with 85.59%; CNN-BPSO with 86.11%. Regarding the set of attributes with the greatest influence on the prediction of student performance, it was identified as a relevant finding.

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Ajibade SSM, Ahmad NB, and Shamsuddin SM (2018). A data mining approach to predict academic performance of students using ensemble techniques. In the International Conference on Intelligent Systems Design and Applications, Springer, Vellore, India: 749-760. https://doi.org/10.1007/978-3-030-16657-1_70

Andrade TLD, Rigo SJ, and Barbosa JLV (2021). Active methodology, educational data mining and learning analytics: A systematic mapping study. Informatics in Education, 20(2): 171-204.

Athani SS, Kodli SA, Banavasi MN, and Hiremath PS (2017). Student academic performance and social behavior predictor using data mining techniques. In the International Conference on Computing, Communication and Automation, IEEE, Greater Noida, India: 170-174. https://doi.org/10.1109/CCAA.2017.8229794

Bhatia P (2019). Data mining and data warehousing: Principles and practical techniques. Cambridge University Press. Cambridge, UK. https://doi.org/10.1017/9781108635592

Chaudhury P, Mishra S, Tripathy HK, and Kishore B (2016). Enhancing the capabilities of student result prediction system. In the Second International Conference on Information and Communication Technology for Competitive Strategies, Udaipur, India: 1-6. https://doi.org/10.1145/2905055.2905150

Deepika K and Sathyanarayana N (2019). Relief-F and budget tree random forest based feature selection for student academic performance prediction. International Journal of Intelligent Engineering and Systems, 12(1): 30-39. https://doi.org/10.22266/ijies2019.0228.04

Hamoud A (2016). Selection of best decision tree algorithm for prediction and classification of students' action. American International Journal of Research in Science, Technology, Engineering and Mathematics, 16(1): 26-32.

He T, Liu Y, Yu Y, Zhao Q, and Hu Z (2020). Application of deep convolutional neural network on feature extraction and detection of wood defects. Measurement, 152: 107357. https://doi.org/10.1016/j.measurement.2019.107357

Hussain S, Gaftandzhieva S, Maniruzzaman M, Doneva R, and Muhsin ZF (2021). Regression analysis of student academic performance using deep learning. Education and Information Technologies, 26(1): 783-798. https://doi.org/10.1007/s10639-020-10241-0

Hussain S, Muhsion ZF, Salal YK, Theodorou P, Kurtoglu F, and Hazarika GC (2019). Prediction model on student performance based on internal assessment using deep learning. International Journal of Emerging Technologies in Learning, 14(8): 4-22. https://doi.org/10.3991/ijet.v14i08.10001

Khan A and Ghosh SK (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and Information Technologies, 26(1): 205-240. https://doi.org/10.1007/s10639-020-10230-3

Kriegeskorte N and Golan T (2019). Neural network models and deep learning. Current Biology, 29(7): R231-R236. https://doi.org/10.1016/j.cub.2019.02.034 **PMid:30939301**

Kumar L and Bharti KK (2019). An improved BPSO algorithm for feature selection. In: Khare A, Tiwary U, Sethi I, and Singh N (Eds.), Recent trends in communication, computing, and electronics: 505-513. Volume 524, Springer, Singapore, Singapore. https://doi.org/10.1007/978-981-13-2685-1_48

Ma X and Zhou Z (2018). Student pass rates prediction using optimized support vector machine and decision tree. In the IEEE 8th Annual Computing and Communication Workshop and Conference, IEEE, Las Vegas, USA: 209-215. https://doi.org/10.1109/CCWC.2018.8301756

Malini J (2021). Analysis of factors affecting student performance evaluation using education datamining technique. Turkish Journal of Computer and Mathematics Education, 12(7): 2413-2424.

Mangina E and Psyrra G (2021). Review of learning analytics and educational data mining applications. In the 13th Annual International Conference on Education and New Learning Technologies, Palma, Spain: 949-954. https://doi.org/10.21125/edulearn.2021.0250

Mason C, Twomey J, Wright D, and Whitman L (2018). Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression. Research in Higher Education, 59(3): 382-400. https://doi.org/10.1007/s11162-017-9473-z

Namoun A and Alshanqiti A (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. Applied Sciences, 11(1): 237. https://doi.org/10.3390/app11010237

Phan H, Andreotti F, Cooray N, Chén OY, and De Vos M (2018). Joint classification and prediction CNN framework for automatic sleep stage classification. IEEE Transactions on Biomedical Engineering, 66(5): 1285-1296. https://doi.org/10.1109/TBME.2018.2872652 **PMid:30346277 PMCid:PMC6487915**

Pojon M (2017). Using machine learning to predict student performance. M.Sc. Thesis, University of Tampere, Tampere, Finland.

Salal YK, Abdullaev SM, and Kumar M (2019). Educational data mining: Student performance prediction in academic. International Journal of Engineering and Advanced Technology, 8(4C): 54-59.

Shetu SF, Saifuzzaman M, Moon NN, Sultana S, and Yousuf R (2021). Student's performance prediction using data mining technique depending on overall academic status and environmental attributes. In the International Conference on Innovative Computing and Communications, Springer, Delhi, India: 757-769. https://doi.org/10.1007/978-981-15-5148-2_66

Singh M, Verma C, Kumar R, and Juneja P (2020). Towards enthusiasm prediction of Portuguese school's students towards higher education in realtime. In the International Conference on Computation, Automation and Knowledge Management, IEEE, Dubai, UAE: 421-425. https://doi.org/10.1109/ICCAKM46823.2020.9051459

Srivastava AK, Chaudhary A, Gautam A, Singh DP, and Khan R (2020). Prediction of students performance using KNN and decision tree-A machine learning approach. Strad Research, 7(9): 119-125. https://doi.org/10.37896/sr7.9/018

Syu NS, Chen YS, and Chuang YY (2018). Learning deep convolutional networks for demosaicing. ArXiv Preprint ArXiv:1802.03769. Available online at: https://arxiv.org/abs/1802.03769

Turabieh H (2019). Hybrid machine learning classifiers to predict student performance. In the 2nd International Conference on New Trends in Computing Sciences, IEEE, Amman, Jordan: 1-6. https://doi.org/10.1109/ICTCS.2019.8923093

UCI (2014). Student performance data set. UCI Machine Learning Repository. Available online at: https://archive.ics.uci.edu/ml/datasets/student+performance

Ünal F (2020). Data mining for student performance prediction in education. In: Birant D (Ed.), Data mining-methods, applications and systems: 153-163. IntechOpen, London, UK. https://doi.org/10.5772/intechopen.91449

Yang J, Zhang XL, and Su P (2019). Deep-learning-based agile teaching framework of software development courses in computer science education. Procedia Computer Science, 154: 137-145. https://doi.org/10.1016/j.procs.2019.06.021

Zaffar M, Hashmani MA, Savita KS, Rizvi SSH, and Rehman M (2020). Role of FCBF feature selection in educational data mining. Mehran University Research Journal of Engineering

and Technology, 39(4): 772-778.
https://doi.org/10.22581/muet1982.2004.09

Zhao R, Yan R, Chen Z, Mao K, Wang P, and Gao RX (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115: 213-237.
https://doi.org/10.1016/j.ymssp.2018.05.050