International Journal of Computational and Electronic Aspects in Engineering
Vol. 5, Issue 3, Sept 2024, pp 80-89
https://doi.org/10.26706/ijceae.5.3.20240802

Enhanced System for Prediction of Students' Performance Using Deep Learning

Mina Basheer Gheni

College of Business Economics, Al-Nahrain University, Baghdad, Iraq

*Correspondence: mina@nahrainuniv.edu.iq

Abstract: This work created an improved deep learning model for predicting student academic achievement. Students' data is gathered from online learning platforms, offline learning systems in schools, or through questions and responses. Many writers could use the collected data to better understand student behavior and, as a result, improve learning levels and student performance. To serve as legitimate input for a deep learning model, the gathered data must be processed first. The deep learning model's hyper-parameters were optimized using a genetic method. The OULAD dataset was utilized for validation. The results demonstrated that CNN plus a genetic algorithm is an effective strategy for predicting academic success.

Keywords: Prediction; Deep Learning; Genetic Algorithm; Optimization; OULAD

Article – Peer Reviewed
Received: 25 July 2024
Accepted: 17 Sept 2024
Published: 30 Sept 2024

Cite this article: Mina Basheer Gheni, "Enhanced System for Prediction of Students' Performance Using Deep Learning", *International Journal of Computational and Electronic Aspects in Engineering*, RAME Publishers, vol. 5, issue 3, pp. 80-89, 2024.
https://doi.org/10.26706/ijceae.5.3.20240802

1. Introduction

Education does a very essential role in our lives. The primary objective of education is to assist each individual to have a prosperous position in community. Education is too a way to assist people enhance their experience and knowledge. An untrained person cannot write and read, and therefore does not have access to knowledge. whose he can gain out of books and other means, that is, he is removed from the external world, and in turn, the cultured man lives in a room whose windows are all open to the outer surface world. So, it is a good instruction that determines the future of the person, and community [1].

One of the most important domains of knowledge in data mining and its various uses to unearth hidden knowledge is mining academic student data to help registration systems and admission systems that assist students at every stage of their education. Data mining has a close relationship with machine learning and deep learning. In general, assessing students' academic achievement has been an important topic in a variety of academic disciplines. The an estimation approach can provide important data for teachers to improve student learning, particularly for pupils with low study performance [2, 3].

Many studies, including [4], [5], have employed machine learning using various methodologies to analyze and enhance education performance in schools and universities. Many studies have found that familial, personal, social, psychological, socioeconomic, and environmental factors all have a direct impact on student academic achievement.

The purpose of this publication is to predict academic accomplishment using a deep learning and machine learning models. It also evaluates these models and investigates ways to improve them in order to create optimal prediction models. The validation is performed using the OULAD dataset. The implemented algorithms are

- Logistic Regression (LR) and
- Convolution Neural Network (CNN).

2. Relevant Works

Numerous developed states have turned their education system into a fully automated system to help students and teachers also participate in an interactive and successful educational process. To improve the educational process of students and their training courses, authors in [6] relied on the decision tree algorithm to rank a group of features that significantly affect the educational performance of students. The NN (neural network) is one more widely used practices and methodologies for analyzing educational data in depth. Authors in [7] used a neural network to envisage the academic progress of undergraduate degree students.

[8] used the Naïve Bayes algorithm to calculate students' cumulative grade point average (CGPA) at acceptance based on their academic history. They also applied data mining techniques to forecast students' academic behavior. [9] conducted a comprehensive survey, analyzing more than 60 articles on student performance. Their research focused on three major areas: assessing learning outcomes, examining models for predicting student performance, and identifying the essential elements influencing student achievement. They found that machine learning approaches, including prediction and regression models, were widely used to forecast student performance. They concluded their investigation by identifying important research challenges and making recommendations to encourage further work in this field.

[10] and [11] used various classification algorithms, including random forest, decision tree, Naïve Bayes, and rule-based. [10] discovered that the rule-based technique had the maximum accuracy of 71.3% when analyzing data from 497 students that included features such as GPA, family income, gender, race, subject grades, and university admission mode. [11] discovered that the decision tree technique produced the best results, with an accuracy of 66.9%, when using data from 300 students that included demographics, family factors, and pre-enrollment attributes (prior GPA).

[12] used classification and regression algorithms to predict student academic progress. Their categorization model predicted whether students would pass or fail, but the regression model predicted student grades. The dataset includes 5779 records with various parameters, such as age, gender, scholarships, and student status. The classification methods employed were KNN, classification and regression trees, AdaBoost, Random Forest (RF), Naïve Bayes (NB), and SVM. For regression, they used RF, AdaBoost, SVM, classification and regression trees, and ordinary least squares.

[13] created a new prediction system to measure student success in scientific academia by combining clustering and classification approaches. This technique was validated using real-time data collected from several Indian universities. The validation showed that the use of classification and clustering methods produced good accuracy.

Finally, it is worth mentioning some modern applications of artificial inelegant. Jyotsna et al demonstrated the use of deep learning in plant disease prediction [14]. Additionally, a chat bot is supposed by Mothankar et al to start a conversation between humans and machines [15]. Also, Prajapati et al has devised AI tools to design smart cart for physically challenged person [16]. Artificial Neural Network is widely applied in many applications like solar energy cells [17]. At last, simulation and numerical investigation are key aspects of modern AI tools like work of Muhammed Shefeek et al [18].

Not all researchers have highlighted the important features that influence educational student's behavior (performance). In our work, we are going to study important features and DL model and compare results with other studies. We also will use an optimization algorithm to optimize developed algorithms.

3. The Developed Method

3.1 Open University Learning Analytics Dataset

The Open University Learning Analytics Dataset (OULAD) includes both quantitative and qualitative data about students, their activities, and their disciplines. This dataset, compiled over nine months (February to October), contains information about the Virtual Learning Environment (VLE). The dataset consists of numerous CSV tables. The tables are described as follows [19].

- studentInfo.csv: This file provides the students' demographic information as well as their results.
- Assessments.csv: This file contains information regarding assignments and presentations. Each presentation contains a distinct set of exams and a final exam.
- studentAssessment.csv: This file contains student grades for subjects, resources, and assignments. It contains no results when a student fails to complete an assessment.
- studentVle.csv: This file contains information on all student activities and interactions with subjects on the VLE.

3.2 Dataset Processing

Dataset might contain columns that need to be processed. The operations on dataset columns to be ready for the ML model are named Dataset cleaning [20]:

- Convert String columns to numeric columns. The conversion might be:
    - Row string value such as “-1” is converted to number -1
    - Convert categorical values such as gender features “Female” “Male” to 0, 1 respectively.
- Replace nan values with replaced with the value (-1).
- Remove null values by replace them with defined value.
- Join all tables into one table.

Each processed dataset is assigned a main key; the final step is to combine all processed datasets into a single dataset.

3.3 Genetic Algorithm for Optimization

The genetic method (GA) is a learning method that uses crossover to combine the biases and weights of two effective neural networks. This crossover procedure seeks to create an optimized neural network with better biases and weights. GA frequently produces optimal solutions, providing useful insights into the problem at hand [21].

Initially, the agent's weights must be optimized. A set of random biases and weights is constructed, resulting in the initial agent's neural network. The agent goes through a series of tests, each of which results in a score. This technique is done several times to generate a population of agents. The top ten percent of the population is chosen for crossover. Mutations can occur during crossing. The genetic algorithm (GA) requires a cost function to minimize. When this cost function approaches a minimum, the tested weights are deemed optimal for the agent. This simple GA approach gradually improves accuracy and performance. In our situation, the agent is a neural network whose weights require modification to improve model performance.

Suppose the parameter is PM; we begin with the initial value and must alter it after updating the GA loss function. The loop terminates after a certain number of repetitions. GA will optimize the hyper parameters of a deep learning model (convolution neural network).

3.4 Convolution neural networks (CNN)

A convolution neural network (CNN) is a form of feed-forward neural network that thrives in situations where target information can be represented by a hierarchy of local features. It has several convolutional and fully connected layers, making it suitable for image-based applications [22].

CNNs are well-suited for image processing due to their substantial spatial dependence in local areas and great translation invariance. Similarly, time series data may contain adjacent linked points that remain constant over time transitions. The successful use of deep CNNs for one-dimensional classification and multi-dimensional time sequences is well known. CNNs, like image classification, may extract deep features from a signal's internal structure, making them effective feature engineering tools in predictive processing tasks (comprehensive learning).

CNN needs data with 2 dimensions, so we need to convert the final data table into 2D. For this study, the deep modular CNN consists of the following components:

- Input
- Convolutional layer
- Aggregation layer
- Fully connected (FC) layer
- Output

Each component has its own properties and parameters. The convolutional layer includes:

- No. filter: The number of convolution filters.
- Filter size: The size of the filters (kernel size).
- Strides: The strides of the convolution operation.
- Padding: The process of adding layers of zeros to the input images to prevent issues related to size after convolution.
- Activation: The activation function applied to this layer (default is linear).
- Bias: If True, a bias is used.
- Weights_init: The initialization of weights.
- Trainable: If True, the weights will be trainable.
- Regularizer: Regularization techniques, such as Drop Block and Shake, are used to reduce over fitting and improve generalization performance.

Additional parameters exist for pooling, regression, and fully connected layers. The pooling layer includes:

- Kernel size
- Strides of the pooling operation
- Padding similar to the convolutional layer

The fully connected layer includes:

- Neurons
- Activation_function
- The regression layer includes:
- Loss_function: Crucial in machine learning models as it defines the metric for performance measurement.
- Learning rate: This parameter helps the model converge faster. An incorrect learning rate, whether too low or too high, can significantly affect performance.
- Optimizer: This function calculates and reduces errors to minimize the specified loss function.

Additionally, there is notion known as dropout, which involves removing hidden and visible units from a neural network during the training phase. Dropout refers to dismissing specific neurons during a forward or backward pass based on a randomly chosen keep probability. This method helps to prevent over fitting by excluding these units during specified training iterations.

3.5 Convert Dataset from 1D to 2D

After encoding, we had a dataset of students. We used a 2D CNN in our study, but the initial results were 1D. As a result, we translated our data into a 2D format appropriate for the 2D CNN architecture. We came up with ten numerical features after converting the categorical data. To transform the data to 2D, we repeated each row five times, creating a $5 \times 10 \times 1$ matrix for each sample. The graphic below depicts the results for a single data sample.

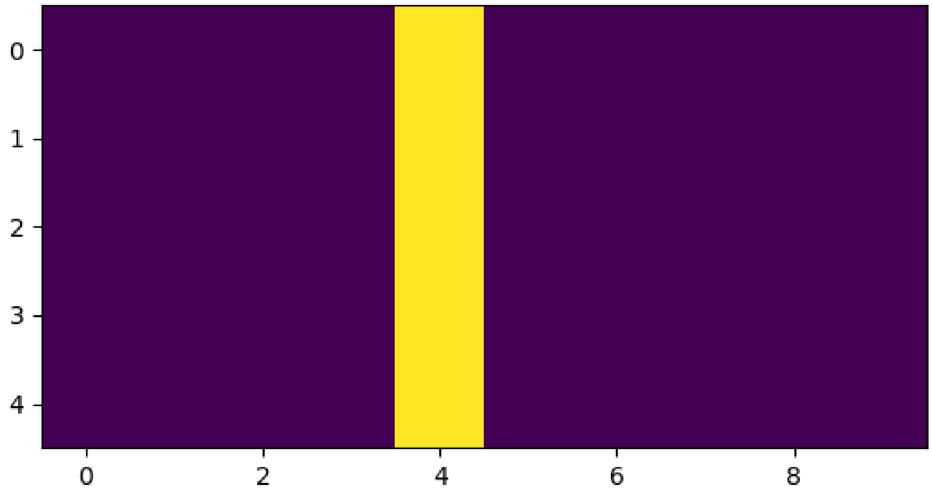

Figure 1. The Constructed 2D Image.

Each sample or student now is represented by an image such previous. The design of CNN is shown below:

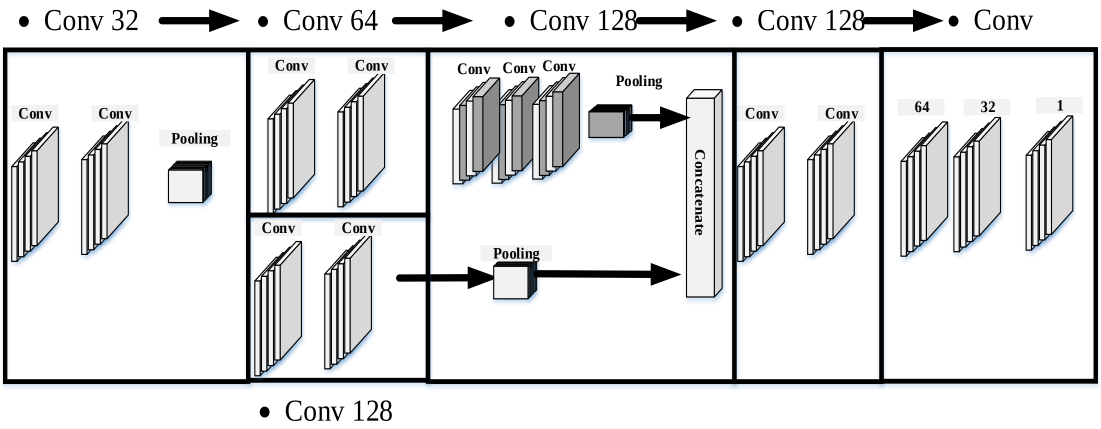

Figure 2. Proposed CNN Structure

Each image is first processed by two wrap layers with 32 filters and then a pooling (assembly) layer. The second stage contains parallel layers: a) two convolution layers with 64 filters, b) two wrap layers with 128 filters.

The third stage contains parallel layers: a) three wrap layers with 128 filters and then an assembly layer, b) pooling layer. The Fourth stage is concatenating layers. Then two convolution layers with 128 layers and three convolution layers with 64, 32, filter.

The concatenation is to mix features from different convolution layers. The dropout layer is added at the end of CNN structure.

4. Implementation

The loop of optimization with 2D CNN is shown below.

The GA optimization loop will retain the loss value for each iteration as well as each loss value with a changing parameter, such as the convolution layer's kernel size.

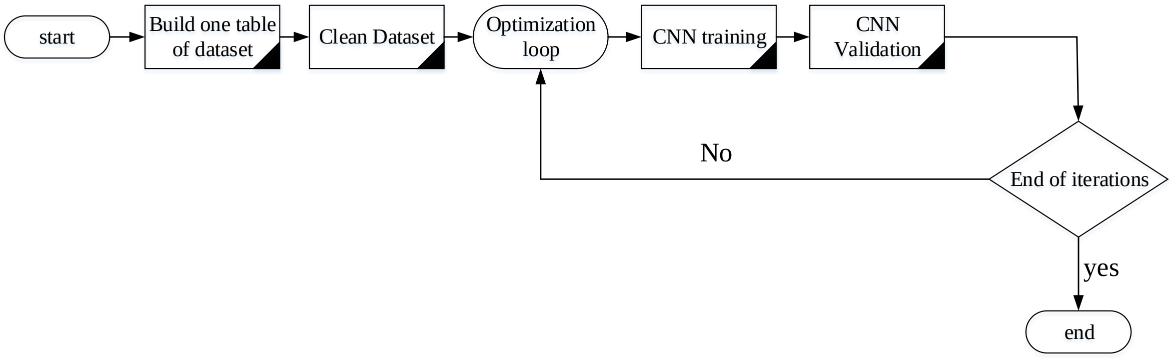

Figure 3. Flowchart of proposed algorithm

4.1 Cleaning Dataset

The first step is to clean the datasets and generate a final file that summarizes all subsets into a single final dataset.

id_assessment	id_student	date_submitted	is_banked	score
1752	11391	18	0	78
1752	28400	22	0	70
1752	31604	17	0	72
1752	32885	26	0	69
1752	38053	19	0	79
1752	45462	20	0	70
1752	45642	18	0	72
1752	52130	19	0	72
1752	53025	9	0	71
1752	57506	18	0	68
1752	58873	19	0	73
1752	59185	18	0	67
1752	62155	17	0	73
1752	63400	19	0	83
1752	65002	17	0	66
1752	70464	19	0	59
1752	71361	19	0	82

Figure 4. shows each student's assessment score for a certain date.

The major table is student information, which comprises each student's final results in the modules.

code_module	code_presentation	id_student	gender	region	highest_education	imd_band	age_band	num_of_prev_attempts	studied_credits	disability	final_result
AAA	2013J	11391	M	East Anglian Region	HE Qualification	90-100%	55<=	0	240	N	Pass
AAA	2013J	28400	F	Scotland	HE Qualification	20-30%	35-55	0	60	N	Pass
AAA	2013J	30268	F	North Western Region	A Level or Equivalent	30-40%	35-55	0	60	Y	Withdrawn
AAA	2013J	31604	F	South East Region	A Level or Equivalent	50-60%	35-55	0	60	N	Pass
AAA	2013J	32885	F	West Midlands Region	Lower Than A Level	50-60%	0-35	0	60	N	Pass
AAA	2013J	38053	M	Wales	A Level or Equivalent	80-90%	35-55	0	60	N	Pass
AAA	2013J	45462	M	Scotland	HE Qualification	30-40%	0-35	0	60	N	Pass
AAA	2013J	45642	F	North Western Region	A Level or Equivalent	90-100%	0-35	0	120	N	Pass
AAA	2013J	52130	F	East Anglian Region	A Level or Equivalent	70-80%	0-35	0	90	N	Pass
AAA	2013J	53025	M	North Region	Post Graduate Qualification		55<=	0	60	N	Pass
AAA	2013J	57506	M	South Region	Lower Than A Level	70-80%	35-55	0	60	N	Pass
AAA	2013J	58873	F	East Anglian Region	A Level or Equivalent	20-30%	0-35	0	60	N	Pass
AAA	2013J	59185	M	East Anglian Region	Lower Than A Level	60-70%	35-55	0	60	N	Pass
AAA	2013J	62155	F	North Western Region	HE Qualification	50-60%	0-35	0	60	N	Pass
AAA	2013J	63400	M	Scotland	Lower Than A Level	40-50%	35-55	0	60	N	Pass
AAA	2013J	65002	F	East Anglian Region	A Level or Equivalent	70-80%	0-35	0	60	N	Withdrawn
AAA	2013J	70464	F	West Midlands Region	A Level or Equivalent	60-70%	35-55	0	60	N	Pass

Figure 5. Student Info table

The final table after cleaning looks as follows:

A	B	C	D	E	F	G	H	I	J
date_submitted	disability	final_result	highest_education	id_student	num_of_prev_attempts	score	studied_credits	sum_of_sum_click	label
27	1	3	3	588775	0	62	120	4104	1
18	1	3	5	591774	0	82	60	0	0
19	1	2	4	603861	0	76	60	0	0
18	1	1	5	606143	0	59	60	1944	0
19	1	3	4	704156	0	73	120	0	0
19	1	4	3	705379	0	90	120	0	0
19	1	2	2	721259	0	66	120	0	0
26	1	3	2	749412	0	45	60	733	0
21	1	4	3	760729	0	77	60	0	0
50	1	3	3	905042	0	58	60	0	0
19	1	3	3	949618	0	85	60	0	0
30	1	3	4	958987	0	61	60	0	0
20	1	3	3	968578	0	67	60	4777	1
19	1	2	3	969076	0	78	180	0	0
19	1	3	4	971027	0	84	60	0	0
19	1	3	4	978739	0	85	60	0	0
17	1	3	2	1035023	0	82	60	0	0
16	1	3	4	1105478	0	83	60	0	0
17	1	2	3	1352868	0	68	60	0	0
19	1	1	3	1401935	0	80	60	0	0
18	1	3	4	1402638	0	83	60	0	0
18	1	1	3	1414443	0	65	120	0	0

Figure 6. final table after cleaning

4.2 Classifiers Results

The hyper-parameters which GA has optimized are:

- Stride was defined for all layers the same and was optimized.
- The kernel size of each convolution layers was the same and was optimized.
- The dropout layer ratio was optimized.
- The Learning rate in the optimizer stage.

Table 1. Stride optimization results

Stride	Accuracy	Loss
0	0.85	0.23
1	0.82	0.33
2	0.72	0.45

Table 2. kernel size optimization results

Kernel size	Accuracy	Loss
3*3	0.83	0.24
2*2	0.82	0.31
4*4	0.76	0.44
5*5	0.75	0.45

Table 3. learning rate optimization results

Learning rate	Accuracy	Loss
0.1	0.8	0.28
0.01	0.82	0.26
0.001	0.84	0.22
0.0001	0.89	0.18
0.00001	0.83	0.23
0.000001	0.78	0.35

Table 4. Dropout rate optimization results

Dropout rate	Accuracy	Loss
0.1	0.61	0.52
0.2	0.71	0.39
0.3	0.84	0.21
0.4	0.81	0.28

After all, we mix best values of all parameters as follow: stride=0, learning rate=0.0001, dropout rate=0.3, kernel size=3*3. The final accuracy was during training is shown below:

```
Step 15, Minibatch Loss= 0.2859, Training Accuracy= 0.849
Step 15, Minibatch Loss= 0.2590, Training Accuracy= 0.880
Step 15, Minibatch Loss= 0.2548, Training Accuracy= 0.898
Step 15, Minibatch Loss= 0.3160, Training Accuracy= 0.864
Step 15, Minibatch Loss= 0.2478, Training Accuracy= 0.893
Step 15, Minibatch Loss= 0.2677, Training Accuracy= 0.896
Step 15, Minibatch Loss= 0.2132, Training Accuracy= 0.928
Step 15, Minibatch Loss= 0.3506, Training Accuracy= 0.842
Step 15, Minibatch Loss= 0.2570, Training Accuracy= 0.896
Step 15, Minibatch Loss= 0.2514, Training Accuracy= 0.870
Step 15, Minibatch Loss= 0.3070, Training Accuracy= 0.879
Step 15, Minibatch Loss= 0.2565, Training Accuracy= 0.890
Step 15, Minibatch Loss= 0.2954, Training Accuracy= 0.890
Step 15, Minibatch Loss= 0.2618, Training Accuracy= 0.888
Step 15, Minibatch Loss= 0.2410, Training Accuracy= 0.912
Step 15, Minibatch Loss= 0.2957, Training Accuracy= 0.883
Step 15, Minibatch Loss= 0.2569, Training Accuracy= 0.889
Step 15, Minibatch Loss= 0.2750, Training Accuracy= 0.882
Step 15, Minibatch Loss= 0.3434, Training Accuracy= 0.868
Step 15, Minibatch Loss= 0.2441, Training Accuracy= 0.891
Step 15, Minibatch Loss= 0.2269, Training Accuracy= 0.898
Step 15, Minibatch Loss= 0.3651, Training Accuracy= 0.823
***********************
```

Figure 7. Accuracy and loss during training phase

Comparing the results with the state of the arts (all research used the OULAD dataset)

Table 5. Results Comparison

Study	Techniques	Accuracy
[23]	"Decision tree, random forest, extreme gradient boosting, and multilayer perceptron"	78.2 %
[24]	CNN and Long Short-Term Memory	61 %
[25]	Decision Tree	83.14 %
[26]	Naïve Bayes	63.8 %
[26]	Hybrid 2D CNN	88 %
Ours	2D CNN with Genetic Algorithm	92.8% (train accuracy) 89.3% (test accuracy)

The results above showed that our proposed system (Genetic Algorithm with CNN) outperformed other researches that has worked on the same dataset.

5. Conclusion

This paper addressed an essential difficulty in predicting pupils' academic achievement. The key contribution is the use of genetic algorithms to discover optimal values for 2D CNN hyper-parameters. The findings demonstrated that combining CNN with an optimization strategy can result in high accuracy in the field of study.

References

[1] A. Satyanarayana and M. Nuckowski, “Data Mining using Ensemble Classifiers for Improved Prediction of Student Academic Performance,” *CUNY Academic Works Publications*, vol. 79, p. 8, 2016, [Online]. Available: https://academicworks.cuny.edu/ny_pubs/79/

[2] S. Huang and N. Fang, “Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models,” *Computers and Education*, vol. 61, no. 1, pp. 133–145, 2013, doi: 10.1016/j.compedu.2012.08.015.

[3] K. Cohen, L., Manion, L., & Morrison, *Research Methods in Education*, 6th ed. London: Routledge, 2007. doi: https://doi.org/10.4324/9780203029053.

[4] W. Ware and J. Galassi, “Using Correlational and Prediction Data to Enhance Student Achievement in K-12 Schools: A Practical Application for School Counselors,” *Professional School Counseling*, vol. 9, no. 5, pp. 344–356, 2006, doi: 10.5330/prsc.9.5.73184524064708t7.

[5] S. K. Yadav and S. Pal, “Data mining: A prediction for performance improvement of engineering students using classification,” *arXiv preprint arXiv:1203.3832*, 2012.

[6] M. Quadri and D. Kalyankar, “Drop out feature of student data for academic performance using decision tree techniques,” *Global Journal of Computer*, vol. 10, no. 2, pp. 2–5, 2010, [Online]. Available: http://computerresearch.org/stpr/index.php/gjcst/article/viewArticle/128

[7] P. M. Arsad, N. Buniyamin, and J. L. A. Manan, “A neural network students’ performance prediction model (NNSPPM),” in *2013 IEEE International Conference on Smart Instrumentation, Measurement and Applications, ICSIMA 2013*, IEEE, 2013, pp. 1–5. doi: 10.1109/ICSIMA.2013.6717966.

[8] N. T. N. Hien and P. Haddawy, “A decision support system for evaluating international student applications,” in *Proceedings - Frontiers in Education Conference, FIE*, IEEE, 2007, pp. F2A-1. doi: 10.1109/FIE.2007.4417958.

[9] A. Namoun and A. Alshanqiti, “Predicting student performance using data mining and learning analytics techniques: A systematic literature review,” *Applied Sciences (Switzerland)*, vol. 11, no. 1, pp. 1–28, 2021, doi: 10.3390/app11010237.

[10] F. Ahmad, N. H. Ismail, and A. A. Aziz, “The prediction of students’ academic performance using classification data mining techniques,” *Applied Mathematical Sciences*, vol. 9, no. 129, pp. 6415–6426, 2015, doi: 10.12988/ams.2015.53289.

[11] F. J. Kaunang and R. Rotikan, “Students’ academic performance prediction using data mining,” in *Proceedings of the 3rd International Conference on Informatics and Computing, ICIC 2018*, IEEE, 2018, pp. 1–5. doi: 10.1109/IAC.2018.8780547.

[12] P. Strecht, L. Cruz, C. Soares, J. Mendes-Moreira, and R. Abreu, “A Comparative Study of Classification and Regression Algorithms for Modelling Students’ Academic Performance,” in *Proceedings of the 8th International Conference on Educational Data Mining*, Madrid, Spain: ERIC, 2015, pp. 392–395. [Online]. Available: https://www.educationaldatamining.org/EDM2015/proceedings/short392-395.pdf

[13] B. K. Francis and S. S. Babu, “Predicting Academic Performance of Students Using a Hybrid Data Mining Approach,” *Journal of Medical Systems*, vol. 43, no. 6, p. 162, 2019, doi: 10.1007/s10916-019-1295-4.

[14] J. Jyotsna, P. Ramteke, and P. Baxla, “Plant Disease Prediction Using Deep Learning,” *International Journal of Computational*

and Electronic Aspects in Engineering, vol. 3, no. 2, Aug. 2022, doi: 10.26706/ijceae.3.2.arset1002.

[15] T. P. Mothankar, P. S. Maski, S. Uikey, and P. S. Asatkar, "Artificial Intelligence Based College Enquiry Chatbot," *International Journal of Computational and Electronic Aspects in Engineering*, vol. 2, no. 2, Jun. 2021, doi: 10.26706/ijceae.2.2.20210411.

[16] A. K. Prajapati, A. Yadav, A. K. Yadav, A. Singh, and V. P. Singh, "Smart Cart for Physically Challenged Person," *Journal of Production and Industrial Engineering*, vol. 4, no. 1, Feb. 2023, doi: 10.26706/jpie.4.1.icramen202310.

[17] A. A. Hadi, "The Impact of Artificial Neural Network (ANN) on the Solar Energy Cells: A Review," *International Journal of Computational and Electronic Aspects in Engineering*, vol. 5, no. 1, pp. 30–41, 2024, [Online]. Available: https://www.rame.org.in/pdf/ijceae/paper5/v5i1paper4.html

[18] K. MuhammedShefeek and P. K. Venkitraj, "Numerical Analysis on Natural Convection from Dual Heating Element in an Enclosure," *International Journal of Analytical, Experimental and Finite Element Analysis*, vol. 2, no. 1, pp. 1–4, 2014, [Online]. Available: https://www.rame.org.in/pdf/papers/issue2/v1paper9.html

[19] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Data Descriptor: Open University Learning Analytics dataset," *Scientific Data*, vol. 4, no. 1, pp. 1–8, 2017, doi: 10.1038/sdata.2017.171.

[20] X. Chu, I. F. Ilyas, S. Krishnan, and J. Wang, "Data cleaning: Overview and emerging challenges," in *Proceedings of the ACM SIGMOD International Conference on Management of Data*, 2016, pp. 2201–2206. doi: 10.1145/2882903.2912574.

[21] O. Kramer, *Genetic Algorithm Essentials*, vol. 679. in Studies in Computational Intelligence, vol. 679. Cham: Springer International Publishing, 2017. doi: 10.1007/978-3-319-52156-5.

[22] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, 2018, doi: 10.1007/s13244-018-0639-9.

[23] A. Rivas, A. González-Briones, G. Hernández, J. Prieto, and P. Chamoso, "Artificial neural network analysis of the academic performance of students in virtual learning environments," *Neurocomputing*, vol. 423, pp. 713–720, 2021, doi: 10.1016/j.neucom.2020.02.125.

[24] X. Song, J. Li, S. Sun, H. Yin, P. Dawson, and R. R. M. Doss, "SEPN: A Sequential Engagement Based Academic Performance Prediction Model," *IEEE Intelligent Systems*, vol. 36, no. 1, pp. 46–53, 2021, doi: 10.1109/MIS.2020.3006961.

[25] S. Rizvi, B. Rienties, and S. A. Khoja, "The role of demographics in online learning; A decision tree based approach," *Computers and Education*, vol. 137, pp. 32–47, 2019, doi: 10.1016/j.compedu.2019.04.001.

[26] E. N. Azizah, U. Pujianto, E. Nugraha, and Darusalam, "Comparative performance between C4.5 and Naive Bayes classifiers in predicting student academic performance in a Virtual Learning Environment," in *2018 4th International Conference on Education and Technology, ICET 2018*, IEEE, 2018, pp. 18–22. doi: 10.1109/ICEAT.2018.8693928.