Research Article

Integrating LA and EDM for Improving Students Success in Higher Education Using FCN Algorithm

Monika Hooda, Chhavi Rana, Omdev Dahiya, Jayashree Premkumar Shet, and Bhupesh Kumar Singh

1Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, Haryana, India
2Department of Computer Science and Engineering, University Institute of Engineering and Technology, Maharshi Dayanand University, Rohtak, Haryana, India
3School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab, India
4College of Science and Arts, an Nabhanya, Qassim University, Qassim, Saudi Arabia
5Arba Minch Institute of Technology, Arba Minch University, Arba Minch, Ethiopia

Correspondence should be addressed to Bhupesh Kumar Singh; dr.bhupeshkumarsingh@amu.edu.et

Received 28 January 2022; Revised 13 February 2022; Accepted 17 February 2022; Published 29 March 2022

Academic Editor: Vijay Kumar

Copyright © 2022 Monika Hooda et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

EDM and LA are two fields that study how to use facts to get more academic learning and enhance the students’ entire performance. Both areas are concerned with a broad range of issues such as curriculum strategies, coaching, mental well-being of students, learning motivation, and academic achievement. The COVID-19 pandemic highly disrupted the higher education sector and shifted the old, chalk-talk teaching-learning model to an online learning format. This meant that the structure and nature of teaching, learning, assessment, and feedback methodologies also changes. With the empowerment in technology, timely and effective feedback is provided by the teachers to achieve greater learning. Through these studies, it is noted that negative feedback discourages the effort and achievement of learners, so it should be carefully crafted and delivered. In this work, a new methodology is planned based on an improved FCN (fully connected network). The key impartial of the proposed method is to regulate the assessment of the quality of students in Higher Education HE. The proposed methodology is composed of different phases: the first phase is data acquisition, in which the data are gathered from various sources for training and testing of the proposed method. The second phase is data orientation, in which the information is oriented in a specific file format. After that, data are cleaned, and preprocessing methods are applied. In the fourth phase, a machine learning-based model is developed to predict student’s academic performance. The fully connected neural network is enhanced with LA to better assess student quality in higher education. The proposed work is evaluated with the OULAD database, which was gathered from the students of Open University. The proposed methodology has attained an accuracy of 84%, more significant than the conventional ANN model accuracy rate. The proposed methodology’s Recall, F1-score, and precision rates are 0.88, 0.91, and 0.93, respectively.

1. Introduction

Higher education has grown tremendously in recent years. Many new institutions, colleges, and universities are being built by both the private and public sectors to promote education and the well-being of students [1, 2]. EDM is a new area that examines information in an academic environment using various data mining tools and methodologies. It gives an in-depth understanding of the teaching and learning activities, allowing for more current educational management, student’s performance analysis, and a better understanding of learning processes to develop more productive teaching approaches. In 2008, Montreal, Canada, hosted the first international scientific conference on EDM. The International EDM Society was created in 2011, and the Journal of EDM began posting in 2009 [3].

Developing sophisticated models that represent a broader linguistic structure is a legitimate objective of
statistical language learning, and verifying the quality of these systems becomes increasingly challenging as the complexity increases [4, 5]. The primary goal of using educational DM (data mining) to evaluate academic data is to assess the contents, model techniques, and analyze the learner’s comments. Educational data mining (EDM) monitors the students’ performance and helps the learners enhance their performance. The EDM system stores the general information of the students, which can be used to calculate the educational performance. Whenever the learners lack in any course, it gives an alarm to the students. Also, it gives a performance chart of the students that can reflect in the students’ learning [6, 7]. Data mining (DM) in education, commonly referred to as learning analytics, is a relatively new interdisciplinary study subject (EDM). It focuses on the developing strategies for analyzing the numerous forms of data collected in an education setting [8, 9]. Educational data mining is a strategy or method for extracting patterns using vast amounts of student records. In the educational field, many data mining approaches aid in analyzing student’s academic performances and discovering different sequences. Educational Data Mining and learning analytics are engaged in both types of information, namely, how learners begin and what the students acquire to estimate their outcomes. Educational data mining aids in the administration of student records and datasets. The Educational Data Mining technique transforms relevant data from learning environments into essential data used for the academic research and identifying students’ learning patterns [10].

EDM can be thought of as a type of advanced statistics. EDM technologies are currently linked to curriculum monitoring systems such as Moodle, the blackboard that gives organized massive amounts of information [11]. It incorporates a variety of psychological aspects to examine students’ behavior, which can be beneficial in increasing their learning capacity. A study was done among children to determine the need for a system that has the ability to improve their comprehension and learning capacities [12, 13]. Learning analytics is a valuable tool for uncovering hidden patterns in the raw collected data from academic educational contexts. Learning analytics transforms unstructured data into specific knowledge that may be used to help teachers, learners, and the contexts in which activities take place. Learning analytics can assist in determining the most appropriate educational conceptual framework for a specific project or course. Universities can discover and expand impactful approaches to enhance particular practice and curriculum objectives by linking resources used by the students and staff with specific educational goals. Educators can use learning analytics to determine courses strongly linked to a student’s summative assessment and those related to a dramatic difference in performance and curriculum dropouts. Educators can quickly identify which tests are practical and need to be revised. Universities may accurately assess how and why the platform is being implemented through the access to the information on improving the educational activities and promoting the implementation where it will have the most impact [14]. Educational Data Mining (EDM) is concerned with creating, investigating, and applying different computerized algorithms to discover and analyze fascinating correlations in data sources that would otherwise be difficult to locate [15, 16]. Universities can also evaluate the usage of the third-party services; a small number of academic staff may employ and therefore do not reflect the hefty premium. As mentioned earlier, EDM and LA play a vital role in improving and enhancing student’s performance. It analyzes students’ performance and provides genuine reviews on the interpretation; the feedback term is used. The primary purpose of feedback is to assist the students in addressing any perceived shortcomings that may have been found due to completing an assessment item [17]. Feedback is used to narrow the gap between a student’s present knowledge and the training objective they need. As a result, feedback can be categorized into three main questions: What am I doing, how am I moving, and what should I do next?

EDM-based recommendation systems can also tap into scientific relations collected in course-learning databases. The recommender platform’s objective is to facilitate learners across a whole program using a competency-based evaluation approach [18]. Learners must attain escalating levels of achievement with each program competency through productive projects. Learners may find suggestions that are helpful in advancing toward the next level of expertise. The primary goal of the EDM-based recommendation systems is to discover individuals who are comparable concerning the learning curves. In EDM and LA, different techniques such as ML (machine learning), DM, and statistics are all used for the assessment and feedback. The data for learning analysis are collected from higher education institutions. The existing majority of data processing approaches are developed to remove useless information and knowledge; there is a need for a process that
efficiently works on educational learning [19]. The learning environment is crucial for analyzing the outcomes and determining whether or not an approach will work in the modern concentration on upgrading and customizing, empowering communities in educational places.

Assessment and feedback play an important role in the learning process and student’s quality outcomes. The assessment process acts as a detecting system for students to predict students’ grades or progress in a course. On the other hand, feedback helps the students to adjust ongoing teaching and learning to improve their achievement of intended instructional outcomes. The development of online learning in higher education requires schools and teachers to shift their thinking and practices in terms of learning effectiveness. In this work, an ML-based method is developed for the assessment and feedback to or enhance students’ outcomes or performance.

The data-driven feedback to students explains the “why” of the predictions. In this research work, we employ machine learning in combination with learning analytics to provide data-driven feedback to the students that supports teachers and students in their success in the course in comparison with the previous studies. A combination of these input variables has not yet been used collectively in predicting the student’s performance and providing decision-based feedback that brings novelty to this piece of work.

The existing methods are suffered from different issues such as data imbalance issues, misclassification issues, and insufficient features. There is a need for a methodology to reduce such problems, which facilitates such matters. A machine learning-based improved fully convolutional network is proposed to assess the student’s performance. An improved fully connected network is a group of interlinked input or outcome labels, and load exists on each link. The proposed model that is working with an OULAD database is separated into a 70:30 ratio for training and testing of the model. The improved FCN model predicts student’s performance and provides feedback for further improvements.

The work sections are organized as follows: A detailed overview of EDM and LA is presented in Section 1. The various types of EDM and learning analytics techniques are surveyed in Section 2. The problem statement is stated in Section 3. In Section 4, the research methodology with architecture is depicted. The simulation setup with results and comparative analysis is elaborated in Section 5.

2. Related Work

This section described the existing EDM and LA methods that are surveyed, and comparison tables are depicted for better analysis. Göppert et al. (2021) [20] designed a model for the calculation of academic performance based on an ANN method. A model validation method based on ANN was presented toward online scheduling challenges involving continuously interlinked automation systems. Because the adaptive framework of the system necessitates the simple visual selection, precisely defined modeling was not a viable option due to the long compute periods required. The estimation of such a performance measure to quickly evaluate a dynamic system condition was selected as the application deployment assistance challenge for such a project. The improvements from the AI code generator AlphaZero were used for this challenge. In particular, an ANN was developed, which could recommend positive work routing options and quantify the effect of operations. The analysis reveals that the learning algorithm could accurately anticipate desirable outcomes, including over 95% accuracy, and calculate the makespan with less than 3% of error. Chango et al. (2021) [21] proposed the information fusion approach based on the blended learning. The information fusion system was used to determine the university students’ overall educational excellence integrating different, multidimensional data across blended educational contexts. Information about first-year university graduates was collected and preprocessed in various ways, including classroom sessions, practical classes, interactive Moodle workshops, and a midterm test. The main goal was to determine where the information fusion method yields the most significant outcomes with our dataset. The findings indicate that aggregates and picking the best characteristics method with fractional order data generate the best predictions. According to the finest estimation techniques, the degree of attentiveness in classroom sessions, Moodle examination results, and Moodle discussion participation were also the most acceptable characteristics for predicting students’ ultimate achievement in proposed programs. Mubarak et al. [22] proposed the CONV-LSTM hypermodel, which combines CNN with the LSTM approach to autonomously select knowledge utilising MOOC raw data collection and evaluating whether each participant will fail or pass the tests. The authors discussed the class imbalance problem that indicates that algorithms would be influenced to produce better results for the overwhelming class examples while producing bad outcomes for the minorities. The system forecast was erroneous, resulting in a significant negative result occurrence. An expense method was applied inside the loss function to encourage higher prediction accuracy, including the numerous categorization consequences for false-negative alarms and false-positive aspects. As contrasted to the baseline approaches, the suggested method was more efficient. Fotso et al. [23] designed a deep learning approach to estimate instructional practices (learner interactions) within the learning experience, allowing students as well as course faculty members to gain a piece of information about how people acquire knowledge. The authors had utilized relevant information from the UNESCO-IICBA (International Institute for Capacity Building in Africa) MOOC framework, which was constructed for the professional development in Africa, to conduct the assessment. They looked at various contents, including regional, social developmental, and learning-specific traits. The authors were proposed that the system can predict a learner’s permit behavior in response to several other MOOC courses by looking at how he watches videos or takes quizzes. Al Nagi et al. [24] used multiple classification algorithms on an educational database for online courses to select the optimum model to classify academic achievement based on the key variables that may lead to the different results. ANN, DT, KNN, and SVM were
some of the classifiers employed. Experiments were performed with actual statistics, as well as the algorithms were assessed using four performance indicators: Precision, accuracy, f-measure, and Recall. The artificial neural network and the decision tree produced the best results among several predictors. Raga et al. [25] used the deep neural network design and Internet communication features as the training sets. The authors were investigated with constructing a forecasting model for academic success in the initial stages of the teaching and learning process. Moodle's activity logs were used to derive the online behavior parameters. A maximum of 885 entries had been used from the students who were enrolled in three separate programs over 16 multiple categories. Firstly, several measurements were taken to find the model parameters for just the highest convolutional neural network (CNN) architecture, as it was used as a foundation classification model. This result backs with earlier research findings. The accuracy performance for forecasting examination findings for a specific course was 91.07%, with a ROC AUC value of 0.88. In contrast, the overall efficiency with forecasting midterm consequences was 80.36%, with a ROC AUC value of 0.70. In Table 1, various existing educational data mining and data analytics methods. The comparison is based on the implemented techniques, comparative methods, and problems.

Brahim [26] used machine learning techniques to predict students’ performance in the DEEDS dataset. The author obtained performance results in terms of accuracy, precision, Recall, F1-score, and ROC, with the help of classifiers Random Forest, SVM, naïve Bayes, logistic Regression, and MLP classifiers. The results showed that RF performed best with an accuracy score of 97% and an F1-score of 97%. Afzaal et al. [27] used a machine learning-based approach to provide data-driven feedback and action recommendations for learners in an LMS environment to support students’ self-regulation and to improve their performance. The results showed that the performance of the Random Forest RF is best in terms of all evaluation measures, that is, 0.75% of accuracy, precision, recall, and F-measure.

The existing methods of EDM and LA with proposed parameters and comparison parameters are depicted in Table 2. The conclusion of current forms of EDM and LA with future enhancement is also defined in Table 2.

3. Main Challenges and Problems

The landscape of knowledge in education has modified over the last two decades. OL (Online learning) has become ubiquitous. Constant analysis of information through learner-centered assessment is the vital situation which is necessary for the success of today’s online courses. Offline courses need to be assessed in such a manner that could help the learner in identifying potentially weak areas and ways of improving. LA efforts on transforming education by changing the nature of teaching, learning, and assessment. Lately, investigators have begun promoting and advocating the use of LA which is “interpretation of a wide range of data produced by and gathered on behalf of students to assess academic progress, predict future performance, and spot potential issues” [28, 29].

The principles for the quality of students were not studied so far, and the perceptions of all stakeholders, namely, industries, faculty, student, and the requirements for quality of faculty, were also completely ignored. As the higher education system is undergoing a colossal change, with privatization and globalization of education, this study will aid the development of the system by bringing in a socially relevant tool [30] and suggestions to the policymakers which will enhance the quality of higher education institutions in India.

Assessment in higher education has traditionally focused on retaining knowledge and its application in limited contexts as measured by paper and pencil tests and academic assignments such as writing term papers. The increasing amount of data generated in digital learning contexts provides opportunities to benefit from learning analytics and challenges related to interoperability, privacy, and pedagogical and organizational models [31, 32]. Consequently, new methodologies and technological tools are necessary to analyze and make sense of these data and provide personalized scaffolding and services to stakeholders including students, faculty/teachers and administrators, and parents. Pedagogical and organizational models must also be incorporated to utilize personalized framing and services to ensure productive learning and teaching [33]. In addition, access to data from different sources raises several concerns related to data sharing and interoperability and the protection of privacy for individuals and business interests for institutions.

One of the challenges that HEIs face in today’s online learning context is the lack of real-time and immediate feedback from the instructors especially in an emergency situation like the COVID-19 outbreak, as good feedback increases students’ satisfaction and learning outcomes. Also, OL limited student-to-student interaction in online learning, which indirectly affects student outcomes [34].

4. Proposed Methodology

Data mining-based proposed work includes decision-making and complex processing. The following is a proposed framework that outlines that the phases and rules may simplify the complete research work procedure. The LA and EDM methods not only help in laying out the plans to manage with DM schemes but also ease the documented procedure and help in fast industry-level standards (educational database), thus doing better quality work. The proposed methodology is one of the world’s widely used DN procedure structures. Typically, it is accessible to use, has well-defined steps, and is highly generic in the future work, making it the most well-suited method for any DM-based project. Then, a similar procedure was followed for students’ learning result prediction. Several analyses that connect student’s attrition prediction have also utilized a similar DM research methodology. The proposed work has defined several phases, as shown in Figure 1, such as (i) data acquisition, (ii) data orientation, (iii) data cleaning, (iv)
modeling using improved FCN model, and (v) evaluate the performance metrics.

4.1. Data Acquisition. This research methodology will use the Open University (OU) which is an anonymized student’s data commonly known as Open University Learning Analytics Dataset (OULAD). It contains data about students, their course-related information, their demographic information, and their interaction with the VLE. The datasets contain records of 32,593 students, enrolled in 22 courses from February to October time duration. This dataset will organize into seven relational tables (vle, studentvle, studentinfo, studentregistration, courses, assessments, and studentassessment) that can be downloaded as seven different ∗ csv or excel files. The proposed model will read all the seven databases and combine the student information, assessments, and courses database. Each table has

| Author name | Techniques | Proposed method | Comparative methods | Problems |
|-------------|------------|----------------|---------------------|----------|
| Göppert, A. et al. [20] | Artificial neural network | Dynamic interconnected based assembly systems for student performance prediction | ANN, Greedy heuristic approaches JRIP REPTREE PART | Need to use a greedy approach for better results |
| Chango et al. [21] | Data diffusion-based approach | Multi-mode data fusion based student academic prediction model | SVM (support vector machine) Logistic regression (LR) | Need to extract semantic level features |
| Mubarak et al. [22] | Combination of CNN and LSTM | A deep learning-based analytic model | SVM (support vector machine) Logistic regression (LR) | Misclassification issues in complex data |
| Fotso et al. [23] | Deep neural network | Deep learning-based model for learner performance prediction | SVM, ANN | For more effective results, need to work on more features |
| Al Nagi et al. [24] | DT | Machine learning-based model for student performance prediction | SVM, ANN | Poor feature extraction results |
| Raga et al. [25] | A deep neural network-based system | Blended learning-based DNN approach for student performance prediction | SVM, ANN | Limited information |
| Brahim (2022) [26] | Logistic regression | Machine learning-based model for student performance prediction | SVM, ANN, Naïve Bayes | Poor feature extraction results; more advanced machine learning algorithms are needed for feature extraction |
| Afzaal et al. [27] | KNN (k nearest neighbors) SVM Random-forest MLP | Machine learning-based approach, a dashboard that provides data-driven feedback for assessment of students outcomes | KNN (k nearest neighbors) SVM | The dashboard that provides feedback and recommendations does not provide evidence about improved student knowledge about course contents; the sample size is small |
defined all the grouped data, and a set of PK (primary key) and FK (foreign key) relates all the seven tables.

4.2. Data Orientation. The student’s communicating information with the vle is saved in the student_vle database. The proposed work will write a script that sums the click made by various student events and keep it as a * csv file [sumclick.csv] then, and the proposed work will be reading this excel * csv file and combining it with the last database.

4.3. Data Cleaning and Preprocessing. The proposed work will use the data cleaning phase. This phase manages with preparing the information to utilize for modeling. It will check the missing values; the data orientation as per the need is the most vital phase that might bring incorrect outcomes if missed. This proposed methodology will use a student registration database as below:

There was wrong in the database; a few unregistered students from the courses were marked as failed, but the actual label should be reserved, so this proposed methodology will manage that case (Figure 2(a)).

Resolve NANs: the present NAN information has been managed in this phase. It has been shown on the database official page that the student who has a date as NAN will have 300 as date, and then this work has managed NANs in the data column as same.

Encode class values. This work has encoded all the feature sets to numeric so that the proposed work has fed them as input to the out model.

Normalize. This proposed model has normalized the information using a min and max scaler so that the proposed model converges fast and performs better.

Split. This proposed model has split and interchanged the database into train, test, and validate set for training, validation, and testing proposed model, as shown in Figures 2(a), 2(b), and 2(c).

4.4. Modeling. To implement a structure based on ML methods for student’s learning, result estimate utilizing several VLE interactivity, academic learning metrics, and

| Author name          | Proposed parameters | Conclusion                                                                 | Future enhancement                                                                 |
|----------------------|---------------------|---------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| Göppert et al. [20]  | Accuracy            | The findings show that the learning algorithm can accurately anticipate pleasant outcomes including over 95% correctness as well as predict the makespan with less than 3% of distortion. | A reinforcement learning-based approach will be used for effective outcomes and more data will be collected. |
|                      | Mean absolute error |                                                                           |                                                                                  |
|                      | Validation error    |                                                                           |                                                                                  |
| Chango et al. [21]   | Accuracy AUC        | The efficient outcomes in AUC values and accuracy came from using ensembles and finding the best features technique from variational data sets. | Intelligent data aggregation based system will be developed by using abstract features. |
|                      |                     |                                                                           |                                                                                  |
| Mubarak et al. [22]  | Accuracy            | The proposed methodology improves the class imbalance issues.             | Student textual data will be mined in the future for better outcomes of the model. |
| Fotso et al. [23]    | Accuracy            | The proposed model can efficiently predict the failure and dropping out students. | Inaccurate results will be improved using the optimization technique. |
|                      | F- measure          |                                                                           |                                                                                  |
| Al Nagi et al. [24]  | Accuracy Acc Rec Pre| The proposed model can handle imbalanced data efficiently.                | For efficient outcomes of the proposed model accuracy classifiers will be enhanced in future. |
|                      |                     |                                                                           |                                                                                  |
| Raga et al. [25]     | Accuracy AUC        | The proposed work is a type of project to create a system that may be used in specific collaborative learning contexts to provide an autonomous assessment as well as educator guidance. | For the early warning of poor performance of students, the planned model will be enhanced with a hybrid approach. |

Table 2: Existing methods of educational data miningEDM and data analytics: parameters, conclusion, and future enhancement.

Abbreviations: SVM: support vector machine; CNN: convolutional neural network; LSTM: long-short term memory; DNN: deep neural network; ANN: artificial neural network; DT: decision tree; KNN: k-nearest neighbor; EDM: educational data mining; LA: learning analytics; OULAD: open university learning analytics dataset; AUC: area under the curve; Rec: recall; Pre: precision; acc: accuracy; MAE: mean absolute error rate.

Figure 1: General flow chart.
wide surveyed has been analyzed in Section 2. This literature survey defined DL- and ML-based methods in student’s attrition prediction. Generally, various FCNs (fully connected network), ANN, and optimization (SGD) algorithms have been implemented to attain the research objective.

4.4.1. Improved Fully Connected Network. An improved fully connected network is a group of interlinked input/output labels, and load exists on each link. In this work, 21 input variable features and three output variable features for each class have been implemented. At the training stage, the system acquires knowledge through load arrangement to estimate the accurate labels of the input module. It is exceptionally capable of deriving explanations from complex or nonspecific data. It can also be used for design formation and recognizing patterns that are complicated to be observed by either human beings or any other computer method. It is suitable for rated consistently inputs and outcomes. This system recognizes input samples and is ideal for predicting the outputs. It is designed as hierarchical spatial features from lower- to higher-level patterns. It comprises pooling, convolution, and fully connected layers. When the layers are weighed, then the structural design is created. The main functions of fully convolutional neural networks samples are categorized into different areas.

4.4.2. Artificial Neural Network. Similar to the human nervous system, Artificial Neural Networks [35] include neurons, some of which are coupled in multiple connection elements. Nodes are the name for such neurons. In the study of intelligent machines, an ANN aims to duplicate the structure of interconnected neurons that made the human brain such that devices can perceive things that make judgments in a human-like fashion. Machines are programmed to act as interlinked cells in the brain to create an artificial neural network.

4.4.3. SGD (Stochastic Gradient Descent) Optimization. It is an iterative process used for objective function optimization. As it substitutes the original gradients (derived from the complete data set) with an approximation, these can be considered a probabilistic approach of stochastic gradient enhancement. The above improves the computing cost, particularly in extensive optimization techniques, leading to faster repetitions in exchange for such a reduced convergence speed. The term “stochastic” refers to systems or procedures with a randomized probability associated with that. As a result, a few selections are randomly chosen rather than selecting the entire database for every repetition in stochastic gradient descent (SGD). The word “batch” is often used in gradient descent to indicate the whole instances from the database employed to compute the value for every repetition. Stochastic gradient descent is a well-known method for training neural networks. Stochastic gradient descent is a sophisticated alternative to the stochastic gradient that addresses the gradient descent computation fundamental flaw.

4.5. Computation. This phase includes computing the model for reliable outcomes aligning with the proposed objectives. The parameters were selected to calculate the structure accuracy rate, precision, recall, etc. The maximum value of prediction accuracy rate suggests better proficiency of the model to classify the SL (student learning) result. TPR (true positive rate), called recall, is the research model’s capability to classify precise courses accurately. F1-score is the complete proposed model’s test of prediction accuracy rate, searching for the perfect balance between precision and recall rate. These performance metrics are compared with the existing model ANN.

Pseudocode in the proposed model:

- Import all the req_libraries
- Read all 7 datasets
- Combine the above databases to form the final database

![Figure 2: (a) Class distribution in training set, (b) class distribution in validation set, and (c) class distribution in testing set.](image)
Orient click data
Merge click data to final database
Clean database
Remove inconsistency
Resolve NANs
Encode categorical Variables
Normalized data
Split the database
Define model architecture
Set optimizer and set tuned hyper-parameters
Initialize model wts (Xavier Uniform Initialization)
Train
Save model wts
Test
Provide feedback based on prediction model

5. Simulation Setup

This section represents the simulation setup, and a series of phases have been surveyed. Before that, the research model was prepared for computation. Data collecting and preprocessing have been approved, as discussed in Section 4. The proposed work is evaluated with the OULAD database [36], which was gathered from the students of Open University. With about 170,000 participants (students) registered in different schemes, Open University UK is the oldest institution of higher learning in the United Kingdom. A virtual learning environment is used to distribute course-related study resources. The OULAD database includes information from 32,592 participants and 22-course sessions, participant (student) demographics, evaluation sessions, results, and clickstream data from the participants (students) of VLE interactions. The research methodology flow chart of the final process is defined in Figure 3.

An improved FCN model has been implemented to define the information that utilized classifiers that met the research idea. The proposed model has been executed on PYTHON. The proposed model has been distributed into a 70 : 30 training and testing ratio. A set of different databases with changing the size of records have been produced to test the proposed presentation of individual methods with modifying data sizes. The input feature vector set contained variables and predicted output variables for the particular class. This proposed model has improved the accuracy rate and optimized the error rates compared with the existing ANN model.

5.1. Result Analysis. In this research work, a prediction model is designed, and based on this, immediate feedback is provided to the students to enhance the quality of outcomes in the courses they enrolled in. This research paper has organized a student’s complete learning result prediction structure. The proposed study aims to implement an ESWS (Early-stage warning system) that expects the comprehensive examination’s outcome of a scholar. The designed model takes students’ demographic information, academic data, and students’ interaction with vle as input and students’ overall results in their final examinations as an output. Based on this output, a feedback module is designed which helps the learners to improve their examinations results and the instructor of the course to modify course contents and their instructional strategies for the success of the course. An improved FCN model has been developed in a recent study to search out the best execution for student data. Figure 4 defines the comparative analysis of all the applied deep learning methods in the form of accuracy, precision, F1-score, and recall.

This proposed model has used Xavier Uniform Distribution to initialize various network layers’ weights. It helped us to enhance the prediction performance of the proposed model. The training model has classified the student’s performance as “PASS,” “WITHDRAW,” and “FAILURE.” After the prediction provides feedback to the students according to their outcomes, it will train on 70% of the database and has used 10% for the validation set. The proposed model used a weighted sampler to load data in a train, validate, and test load. Then, each of them has a similar ratio of samples for each class. The proposed model has used the SGD (Stochastic Gradient Descent) optimizer. This optimizer has reduced the dimensionality of the feature matrix. This proposed model has used loss function (cross-entropy loss) and hyper-parameters such as epochs = 900; batch_size = 16; learning_rate = 0.07. The proposed model has printed the training loss, validated loss, train accuracy rate, and validated accuracy rate for each epoch. Figure 4 defines train/validate accuracy rate and train/validate loss, respectively. Then, the proposed model has stored the best weights.

The proposed model has performed a test on 20% of the dataset, as shown in Table 3. It offers the performance metrics precision, recall, and F1-score.

Our research work achieves the highest value of precision, recall, and F1-score, i.e., 0.93, 0.88, and 0.91, respectively, when the parameter is 1 and “support” is 26,235, which is quite commendable. But all the three decrease with a rise in the number of parameters. The accuracy score of our designed FCN model is 84% score, which shows that the improved FCN model works fairly well.

The proposed model prediction process has been completed and given the feedback of students’ performance.

Immediate feedback provided will help the learners to engage and modify their behaviors in the learning environment. Also, it helps students in improving retention and their outcomes in the course. Making students engage with the feedback process will enhance learning and advance assessment performance. Also, it helps instructors to modify and improve teaching strategies and leads to the success of the course to a great extent. Students must pay more attention to the student courses they are enrolled in to perform better. A cluster of students is likely to fail as per the proposed prediction model. It has achieved feedback as “STUDENT_No is likely to get failure, requirements to pay attention.” So, there will be students who are predicted to get
Passed. They will get the feedback as “STUDENT_NO, is likely to attain PASS, keep up the good work.” Finally, there will be students who are LIKELY TO WITHDRAW. For them, the student course creator will attain FEEDBACK as “STUDENT_NO is LIKELY TO WITHDRAW.” So, course creators can involve and pay more focus to those students to understand their issues.

5.2. Evaluation. The proposed model performance metrics are described as below:

5.2.1. Precision. Precision is the number of correct class predictions, which genuinely belongs to the true positive is measured by precision. It is the ratio of accurately expected positive findings to expected positive occurrences. Statistically, this can be detailed as follows:

$$\text{Precision} = \frac{TP}{TP + FP}$$  \hspace{1cm} (1)

In equation (1), $TP = \text{true positive}$ and $FN = \text{false positive}$.

5.2.2. Recall. Recall is the number of correct category expectations made from all positive instances within the
database is measured by the recall. The percentage of relevant samples, which are successfully retrieved, is known as recall. Statistically, this can be detailed as follows:

\[
\text{Recall} = \frac{TP}{TP + FN}
\]  

In equation (2), TP = true positive and FN = false negative.

5.2.3. F1-Score. F1-score is the weighted mean of recall, and precision is the F1-score. As a result, the score considers false negatives and false positives. Statistically, this can be detailed as follows:

\[
F - Score = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}
\]  

5.2.4. Accuracy. Accuracy is the value of the proportionality of precise classified value, where sum_of TP and TN to the sum of TP, TN, FP, and FN. Statistically, this can be stated as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]  

In equation (4), TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

Figure 5 defines the comparison analysis with the improved FCN model and the existing ANN model. Current surveys in this area of theoretical result prediction, such as the classification model, define the parameter of accuracy rate. The accuracy rate is one of the most significant performance calculations. So, calculating the accuracy rate for this simulation, the improved FCN model has gained a high performance of 84%, followed by ANN with 78%. The accuracy rate values for all classification methods are observed in Table 4. It is perfect from the table that the ANN model has been the minimum performer in the form of accuracy rate for the data sizes.

Figure 6 defines how accurately the classifier verifies the class. It is the ratio of TP’s verified and the total of a TP and FP. This is the simulation arrangement with deep learning classification models. The precision principles for each class are evaluated distinctly. Figure 6 defines the comparative analysis with proposed and existing models. This proposed model has improved the precision rate compared with the current ANN model. Table 5 shows that the improved FCN and ANN models have been reliable and accurate in predicting precise classes.

Figure 7 shows the recall comparison analysis with the proposed improved FCN model and the existing ANN algorithm. It only chooses accuracy as the parameter does not give an accurate indication of the proposed algorithm’s performance. Then, the recall parameter that considers the verification of the classification models’ capacity to accurately classify the TC (true class) has been utilized. The parameter of recall value for all the methods is defined in Table 6.

Figure 8 defines the F1-score to provide the complete presentation of the method in the form of Precision and Recall. If there is a perfect balance between the recall and precision values, the F1-score is 0.91, which is the accurate value for the deep learning method. The proposed method outcomes for the F1-score are defined in Table 7. The improved FCN model has been the most precise in performance with the maximum values.
6. Conclusion and Future Work

This work planned an enhanced EDM system with an improved FCN algorithm for the assessment of quality of students in Higher Education (HE) system by integrating learning analytics. Both EDM and LA are essential for students’ performance assessment. EDM is a strategy or method for extracting patterns using vast amount of student records, and learning analytics is a valuable tool for uncovering hidden patterns in raw collected data. A machine learning-based FCN approach is used to find the assessment and feedback or enhance student outcomes. The outcome defines that the I-FCN (improved fully connected network) has been the best fit for this recent proposed work with the maximum precision rate among all the other developed supervised learning method. The accuracy rate and F1-score attained by the research model are rather estimable. Several epochs execute on distinct data sizes ensued in the best probable precision rate. The various existing educational data mining and LA methods, such as ANN, SVM, DNN, and KNN, are compared and analyzed. The existing methods are suffered from different issues such as data imbalance issues, misclassification issues, and insufficient features. To reduce such problems, the FCN framework is improved and attained better results. The proposed methodology is executed in the PYTHON model environment. This proposed model has improved the accuracy rate and optimized the error rate and compared to the existing ANN model. In the future, data size will be increased to train and test the model. An optimization technique will be used to enhance the accuracy of the method. A hybrid feature extraction process will be implemented for more efficient outcomes.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S. Meri Al Absi, A. Hasan Jabbar, S. Oudah Mezan, B. Ahmed Al-Rawi, and S. Thajee Al_Attabi, “An experimental test of the performance enhancement of a Savonius turbine by modifying the inner surface of a blade,” Materials Today Proceedings, vol. 42, pp. 2233–2240, 2021.
[2] M. K. Loganathan, C. M. Tan, B. Mishra, T. A. Msagati, and L. W. Snyman, “Review and selection of advanced battery technologies for post-2020 era electric vehicles,” in Proceedings of the 2019 IEEE Transportation Electrification Conference (ITEC-India), pp. 1–5, IEEE, Bengaluru, India, 2019, December.
[3] X. Du, J. Yang, I. L. Hung, and B. Shelton, “Educational data mining: a systematic review of research and emerging trends,” Information Discovery and Delivery, vol. 48, 2020.
[4] M. Sehgal and D. Bhargava, “Knowledge mining: an approach using comparison of data cleansing tools,” Journal of Information and Optimization Sciences, vol. 39, no. 1, pp. 337–343, 2018.
[5] M. N. Kumar, V. Jagota, and M. Shabaz, "Retrospection of the optimization model for designing the power train of a formula student race car," Scientific Programming, vol. 2021, Article ID 9465702, 9 pages, 2021.
mathematical problems in engineering

[6] T. Thilagaraj and N. Sengottayanan, “A review of educational data mining in higher education system,” in Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering, pp. 349–358, Hyderabad, India, June, 2017.

[7] Y. S. Mitrofanova, A. A. Sherstobitova, and O. A. Filippova, “Modeling smart learning processes based on educational data mining tools,” in Smart Education and E-Learning 2019, pp. 561–571, Springer, Singapore, 2019.

[8] R. K. Garg, J. Bhola, and S. K. Soni, “Healthcare monitoring of mountaineers by low power wireless sensor networks,” Informatics in Medicine Unlocked, vol. 27, Article ID 100775, 2021.

[9] Z. Yan, Y. Yu, and M. Shabaz, “Optimization research on deep learning and temporal segmentation algorithm of video shot in basketball games,” Computational Intelligence and Neuroscience, vol. 2021, Article ID 4674140, 10 pages, 2021.

[10] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, “Mining in educational data: review and future directions,” in Advances in Intelligent Systems and Computing, pp. 92–102, Springer, Singapore, 2020.

[11] G. Mahajan and B. Saini, “Educational Data Mining: a state-of-the-art survey on tools and techniques used in EDM,” The International Journal of Computer Applications & Information Technology, vol. 12, no. 1, pp. 310–316, 2020.

[12] M. K. Loganathan, S. S. Neog, and S. Rai, “Process safety and performance improvement in oil refineries through active redundancy and risk assessment method-A case study,” in Proceedings of the 2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), pp. 98–102, IEEE, Bangkok, Thailand, 2018, December.

[13] S. Saralch, V. Jagota, D. Pathak, and V. Singh, “Response surface methodology based analysis of the impact of nanoclay addition on the wear resistance of polypropylene,” The European Physical Journal - Applied Physics, vol. 86, no. 1, Article ID 10401, 2019.

[14] M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg, “Advances in learning analytics and educational data mining,” in Proceedings of the ESANN2015, pp. 297–306, Bruges, Belgium, 2015, January.

[15] J. Bhola, S. Soni, and G. K. Cheema, “Recent trends for security applications in wireless sensor networks-A technical review,” in Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), pp. 707–712, IEEE, New Delhi, India, 2019, March.

[16] S. Vyas and D. Bhargava, “Big data analytics and cognitive computing in smart health systems,” in Smart Health Systems, pp. 87–100, Springer, Singapore, 2021.

[17] N. Scaller, A. Peasgood, and J. Mullan, “Learning analytics in higher education,” Frontiers in Psychology, vol. 8, p. 176, 2016.

[18] A. Pardo, O. Poquet, O. Poquet, R. Martinez-Maldonado, and S. Dawson, “ Provision of data-driven student feedback in LA & EDM,” Handbook of learning analytics, pp. 163–174, 2017.

[19] A. Hicham, A. Jeghal, A. Sabri, and H. Tairi, “A survey on educational data mining [2014-2019],” in Proceedings of the 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), pp. 1–6, IEEE, Fez, Morocco, 2020, June.

[20] A. Göppert, L. Mohring, and R. H. Schmitt, “Predicting performance indicators with ANNs for AI-based online scheduling in dynamically interconnected assembly systems,” Production Engineering, vol. 15, pp. 1–15, 2021.

[21] W. Chang, R. Cerezo, and C. Romero, “Multi-source and multimodal data fusion for predicting academic performance in blended learning university courses,” Computers & Electrical Engineering, vol. 89, Article ID 106908, 2021.

[22] A. A. Mubarak, H. Cao, and I. M. Hezam, “Deep analytic model for student dropout prediction in massive open online courses,” Computers & Electrical Engineering, vol. 93, Article ID 107271, 2021.

[23] J. E. M. Fotso, B. Batchkari, R. Nkambou, and G. Okereke, “Algorithms for the development of deep learning models for classification and prediction of behaviour in MOOCS,” in Proceedings of the 2020 IEEE Learning with MOOCS (LWMOOCs), pp. 180–184, IEEE, Antigua Guatemala, Guatemala, 2020, October.

[24] E. Al Naci and N. Al-Madi, “Predicting students performance in online courses using classification techniques,” in Proceedings of the 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA), pp. 51–58, IEEE, Valencia, Spain, 2020, October.

[25] R. C. Raga and J. D. Raga, “Early prediction of student performance in blended learning courses using deep neural networks,” in Proceedings of the 2019 International Symposium on Educational Technology (ISET), pp. 39–43, IEEE, Hradec Kralove, Czech Republic, 2019, July.

[26] G. B. Brahim, “Predicting student performance from online engagement activities using novel statistical features,” Arabian Journal for Science and Engineering, vol. 486, pp. 1–19, 2022.

[27] M. Afzaal, J. Nouri, A. Zia et al., “Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation,” Frontiers in Artificial Intelligence, vol. 4, Article ID 723447, 2021.

[28] A. B. E. D. Ahmed and I. S. Elaraby, “Data mining: a prediction for student’s performance using classification method,” World Journal of Computer Application and Technology, vol. 2, no. 2, pp. 43–47, 2014.

[29] G. J. A. Baars, T. Stijnen, and T. A. W. Splinter, “A model to predict student failure in the first year of the undergraduate medical curriculum,” Health Professions Education, vol. 3, no. 1, pp. 5–14, 2017.

[30] R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, “Predicting student performance from LMS data: a comparison of 17 blended courses using Moodle LMS,” IEEE Transactions on Learning Technologies, vol. 10, no. 1, pp. 17–29, 2016.

[31] V. Jagota, M. Luthra, J. Bhola, A. Sharma, and M. Shabaz, “A secure energy-aware game theory (SEGaT) mechanism for coordination in WSANs,” International Journal of Swarm Intelligence Research, vol. 13, no. 2, pp. 1–16, 2022.

[32] T. K. Lohani, M. T. Ayana, A. K. Mohammed, M. Shabaz, G. Dhiman, and V. Jagota, “A comprehensive approach of hydrological issues related to ground water using GIS in the Hindu holy city of Gaya, India,” World Journal of Engineering, p. 6, 2021.

[33] T. Dharmanawan, H. Ginardi, and A. Munif, “Dropout detection using non-academic data,” in Proceedings of the 2018 4th International Conference on Science and Technology (ICST), pp. 1–4, IEEE, Yogyakarta, Indonesia, 2018, August.

[34] M. W. L. Yeung and A. H. Y. Yau, “A thematic analysis of the academic experience of higher education students’ perceptions of online learning in Hong Kong under COVID-19: challenges, strategies and support,” Education and Information Technologies, vol. 27, no. 1, pp. 181–208, 2022.

[35] A. Jain, Multi-class Classification to Track Students, academic outcome (Doctoral dissertation, Dublin, National College of Ireland), Dublin, Ireland, 2019.

[36] J. Kuzilek, M. Hlosta, and Z. Zdrahal, “Open university learning analytics dataset,” Scientific Data, vol. 4, no. 1, Article ID 170171, 2017.