Kufa Journal of Engineering
Vol. 17, No. 2, April 2026, P.P. 212-238
Article history: Received 25 February 2025, last revised 23 June 2025, accepted 5 July 2025

OPTIMIZED DEEP LEARNING TECHNIQUES TO PREDICT LEARNERS' ONLINE PERFORMANCE

Safa Ridha Albo Abdullah

Department of Cyber Security, College of Information Technology, University of Babylon, Babil, Iraq. Email: safaruda@uobabylon.edu.iq.

https://doi.org/10.30572/2018/KJE/170213

ABSTRACT

Existing models for predicting students' academic performance show limitations which result in their inability to make accurate predictions. The system achieved operational efficiency together with precise predictions, which created problems for its application in extensive online educational platforms. The system fails to predict accurately because it depends solely on basic dataset attributes, which prevents it from understanding complex user behavior patterns. The research introduces a refined Deep Neural Network (DNN) classifier to predict student performance. The research connects these study gaps with results obtained from UK Open University Learning Analytics Dataset (OULAD) which includes Social Science and STEM courses. The study introduced new features that track student behaviors through different activities which include Engagement, Total Pre-Course Activities, Average, Studying, Discussing, Examining, Working, and Total Post-Course Activities. The research team selected their most predictive features through Particle Swarm Optimization (PSO), which resulted in a feature reduction from 40 to 12. The research team optimized the hyperparameters through Grid Search, which tested learning rates between 0.0001 and 0.1 and assessed three hidden unit configurations: 64, 128, and 256. The proposed model achieved accuracies of 90% (Social Science) and 88% (STEM), which showed better performance than previous studies that used the same datasets with lower computational requirements. The results enable educators to perform personalized educational interventions which will improve student results during online courses.

KEYWORDS

Deep Learning, Predicting Student Performance, Grid Search Optimizer, Heuristic Search, Feature Selection.

1. INTRODUCTION

The academic institutions need to forecast student academic performance through data mining methods which enable them to design specific intervention strategies that improve educational outcomes (Abdullah and Al-Azawei, 2025). The solution to study complex behavioral data from Virtual Learning Environments (VLE) and demographic data requires researchers to face three major obstacles: necessary elements must be chosen from extensive, messy databases, and the analysis process demands substantial processing power and current methods need to achieve better results than their existing performance (Kandula and Kumar, 2023). To provide targeted support and interventions, the effective prediction of students' performance is necessary for educational institutions (Abdullah and Al-Azawei, 2023). However, educational institutions can go above and beyond to address the needs of students who require more support by identifying and performing targeted efforts, ultimately improving their results (Kandula and Kumar, 2023).

The prediction of students' performance is one of the most essential aspects of educational data mining, which can help tutors leader the learning method to improve students' achievement and education quality (Liu et al., 2023). Quality is the primary concern in the realm of education,, and educators have always wanted to progress the quality of instruction and learning (Liu et al., 2023). Effective prediction methodologies and features are required to help teachers and educators create relevant lesson plans that will support students in achieving the expected learning objectives (Nawang, Makhtar and Hamzah, 2021). Using machine learning and deep learning algorithms is one of extremely relevant method (Aljameel et al., 2021). These models allow authorities to implement alternative actions to support at-risk students by anticipating dropout and informing the pertinent authorities in advance, (Andrade-Girón et al., 2023). In contrast to traditional prediction models, Deep neural networks (DNNs) are examples of deep learning-based prediction models that have a large number of features that can be accurately analyzed (Kriegeskorte and Golan, 2019, Wu et al., 2019). As compared to individual learning models, Deep learning models are more effective for many learning tasks (Wotaifi and Dhannoon, 2023b). Furthermore, when a model learns, it excerpts more profoundly important information, thereby significantly increasing the precision of the model's predictions (Poirion et al., 2021).

A Deep Neural Network (DNN) model, optimized with Particle Swarm Optimization (PSO) for feature selection and Grid Search for hyperparameter tuning in this research, to address computational complexity, scalability, and accuracy challenges in predicting student outcomes across Social Science and STEM courses. Starting Virtual Learning Environment (VLE) interactions and demographic data, by generating novel features (e.g., Engagement, Discussing) the model aims to enable real-time interventions and enhance prediction accuracy. There are Challenges such as the high computational complexity of deep learning models, the difficulty of selecting relevant features from large and noisy educational datasets, and the need for improved accuracy over existing methods that remain significant barriers to effective student performance prediction, despite the potential of DNNs. Accordingly, the study contributions may be listed as follows:

Eight new features, were generated to capture complex behavioral patterns and improve the model's accuracy compared to models without these features improve the model's accuracy.

Using Particle Swarm Optimization (PSO) in order to determining the most relevant features that could affect the performance of students on different courses (All-courses, Social Science courses, and the four STEM (Science, Technology, Engineering, and Mathematics) courses.

Prediction of students' academic performance, using deep learning methods instead of traditional machine learning methods.

Compared to previous studies, improved accuracy in predicting students' performance, targeting an improvement. These contributions address challenges in computational complexity, feature selection, and predictive accuracy, enabling effective early interventions.

The rest of this paper is organized as follows. Section 2 presents the theoretical background and related work; Section 3 describes the research methodology for the proposed system; Section 4 reports and discusses the research results, and Section 5 draws a conclusion and suggests possible future research directions.

2. THEORETICAL BACKGROUND AND RELATED WORK

2.1. Deep Learning

Deep learning is a subset of machine learning that can achieve high accuracy in several domains, such as image processing, audio-recognition, and text categorization, among others, by utilizing multi-layer neural networks (Abdulaal et al., 2024). The idea behind deep learning methods, which are based on artificial neural networks (ANNs), is to divide work into layers, each of which is in charge of performing a specific function (Q. Li et al, 2022). While there are other methods used in deep learning, DNN, convolutional neural networks (CNNs), recurrent neural networks (RNNs) and are the most crucial (Al-Hmouz, 2020).

2.1.1. Deep Neural Networks (DNN)

The fundamental design of a DNN is motivated by the anatomy of the brain of the human. The construction of a DNN has several processing units, and these are interconnected. In brief, the perceptron produces output after receiving input (Rani et al., 2023). Equation 1 illustrates the basic concept of a DNN: input (i) is combined with a bias (b) and then weighted by (w) before being summed.

$$O = (\sum(w * i) + b) \tag{1}$$

where, O = Output, f = Activation function, w = Weight, i = Input, and b = Bias.

The range of weight is -1 to 1. The point at which anything meaningful can happen will take longer to reach if all the weights are set extremely low. Conversely, employing high starting weights will raise the possibility of being stuck in a local optimum too soon (Kareem and Jasim, 2022b). The activation function in DNN activates and deactivates nodes, and then carries out transformation in a nonlinear manner (Abdulaal et al., 2025). Frequently utilized activation functions include sigmoid, softmax, and rectified linear (ReLU) (Rani et al., 2023).

2.1.2. Grid Search Optimizer

One popular technique for determining a classification model's proper hyperparameters is the grid search optimizer method. If there are enough grid nodes, there is the possibility it can obtain the optimal solution (Erdogan Erten, Bozkurt Keser and Yavuz, 2021). To summarize, the search optimizer approach determines which hyperparameter combination will produce the greatest outcomes for the model's performance (Ismael et al., 2025). In the grid search optimizer approach, the dataset is randomly separated into test, validation, and training sets, utilizing a train test split (Erdogan Erten, Bozkurt Keser and Yavuz, 2021) . The accuracy of predictions can be greatly impacted by determining the optimal solution.

2.2. Previous Work

The Open University Learning Analytics Dataset (OULAD) serves as a widely used educational data mining resource which provides free access to its dataset that contains demographic details and academic performance data and student conduct records of 32593 students who studied 22 Social Science and STEM courses during the 20132014 academic year (Kuzilek, Hlosta and Zdrahal, 2017). The system provides complete Virtual Learning Environment (VLE) user activity tracking through its seven interactive database tables. OULAD has been used in numerous research studies to investigate the elements that influence students' academic success in Massive Open Online Courses (MOOCs). Recent research studies from 2024 to 2025 have utilized OULAD and other datasets to forecast student achievement through sophisticated deep learning methodologies.

(Alnasyan, Basheri and Alassafi, 2025) The OULA dataset was utilized to forecast student achievement in Virtual Learning Environments through the implementation of seven deep learning models which included ResNet, NODE, AutoInt, TabNet, TabTransformer (TT), SAINT, and GatedTabTransformer (GTT). In multi-class classification when combined with the Tomek Links method, the SAINT model achieved the highest accuracy of 94.33% in binary classification and 73.22%. Statistical analysis confirmed SAINT and AutoInt as the top-performing models, demonstrating their effectiveness in handling complex and imbalanced educational data. The study investigates how resampling methods such as SMOTE and ROS and ADASYN and RUS and Tomek Links solve class imbalance problems. In a Similar, (Alnasyan et al., 2025) established the KANFormer model as a deep learning architecture which integrates multi-head self-attention mechanisms with Kolmogorov–Arnold networks for student performance prediction.

The model evaluation combined the Open University Learning Analytics Dataset with ASSISTments2012 and xAPI and Mathematics and Portuguese Language Courses dataset. KANFormer achieved better performance results than traditional models and advanced state-of-the-art models by attaining 95.14% accuracy in binary classification and 82.73% accuracy for multi-class classification. Academic assessments served as the primary factors that predicted student success according to SHAP analysis results. The Synthetic Minority Oversampling Technique (SMOTE) was used to solve class imbalance problems which led to better performance metrics and maintained equal class distributions. (Ouahi, Khoulji and Kerkeb, 2024) The study utilized the Open University Learning Analytics Dataset (OULAD), which contained data from 32,593 undergraduate students from the academic year 2014–2015. Researchers used Long Short-Term Memory (LSTM) neural network technology to forecast student performance based on their first contact with Virtual Learning Environment (VLE) systems. The model reached 60% accuracy from learner interaction data after analyzing the first three course weeks. The research helped to discover main factors which determine student achievement in online learning environments thus allowing educational institutions to make better decisions. Additional study (Ben Said, Abdel-Salam and Hazaa, 2024) developed a dual-pathway deep learning framework to estimate student performance in online courses by analyzing clickstream data from the Open University Learning Analytics Dataset (OULAD) and converting time series data into images through the Gramian Angular Field method. The model used demographic data together with assessment data to improve its prediction accuracy. The system attained 60% accuracy using 5th week data yet its performance improved to over 90% when all course information became accessible. The research demonstrated that deep learning methods deliver better results than standard machine learning techniques even though researchers encountered obstacles from imbalanced data distributions.

Finally, (Moubayed et al., 2023) The researchers applied deep learning methods through Convolutional Neural Networks and Recurrent Neural Networks with Long Short-Term Memory to forecast student results during online course assessments. The study used three separate datasets which came from different locations around the world. The deep learning models showed better performance than traditional machine learning models on two datasets while they reached similar results with the third dataset, which showed their capability to predict student performance accurately. In Dataset 1 the RNN-LSTM model achieved an accuracy of 82% and in Dataset 3 the CNN model achieved an accuracy of 91%.

The evaluation of these methods through critical analysis shows their most important restriction. Real-time educational environments cannot use SAINT and KANFormer models because their accurate results require excessive computational power to function properly. While CNN/RNN-LSTM and two-pathway models lack interpretability or perform badly with partial data, early prediction models like LSTM suffer from low accuracy due to restricted features. These gaps demonstrate the necessity of a model that strikes a balance between interpretability, efficiency, and accuracy. By employing a smaller feature set (12 features compared to 40 in previous models) and managing unbalanced data pass and fail without SMOTE, the suggested DNN with PSO resolves these problems and achieves competitive performance (Section 4.3).

3. RESEARCH METHODOLOGY

Over four time periods (quarter 1 [Q1], quarter 2 [Q2], quarter 3 [Q3], and quarter 4 [Q4]), an improved model based on deep learning was proposed for the two course datasets: OULA Science courses and OULA Social Science courses dataset (Waheed et al., 2023). To provide a constant indicator of the eventual results for students, the model was applied for a long enough period assuming that they would maintain their current academic status. The classification procedure deployed assessment scores, as well as behavioral and demographic variables. These crucial characteristics were identified using a PSO feature selection technique.

3.1. Dataset

As mentioned previously, the Open University (OU) dataset was used in this study (Aljohani, Fayoumi and Hassan, 2019). This dataset contains students’ demographic, course enrolment, assessment, clickstream, and final performance score data for training machine learning models. The Open Data Institute (ODI) (Kuzilek, Hlosta and Zdrahal, 2017) has formally verified this dataset, which is freely accessible via OULA. Within the OULA dataset, clickstream, course, student, and registration data are distributed through seven tables, with the students serving as the main focus of information. Over a period of nine months, data on 32,593 students were collected. Final performance of the students' was divided into four classes/grades: failed students (22%), passed students (38%), students with the award of Distinction (9%), and students who had withdrawn (31%). In the current study, two classes—"pass" and "fail"—were created from the reported student performance to provide a binary categorization. The class labels for withdrawal and failure were combined into one label, "fail," while the pass and distinction labels were merged into one label, "pass," to produce this categorization. The students were able to choose from seven modules, each of which was offered at least twice a year. The seven tables containing the raw data were linked by identification columns and provided data on assessments, types of assessment, submission dates for assessments, courses, types of VLE materials, clickstreams showing students' VLE engagement, course registration details, and students' demographic information (see Table 1).

Table 1. The OULA dataset used.

Selected Dataset	Course Name	No.of Instances
OULA-Science courses dataset	C, D, E, and F	21,402
OULA Social Science courses dataset	A, B, and G	11,191

3.2. Data Preprocessing

The dataset was obtained from a large number of data files in an unprocessed, organized format (M.Steinbach, P.Tan and V.Kumar, 2006). However, before the data were utilized in the machine learning algorithms, they were pre-processed (Mansour et al., 2024). Preprocessing consists of a variety of methods and approaches (Qasrawi et al., 2021), including data-cleaning, addressing missing values, categorical data-coding, and detecting inconsistent data (M.Steinbach, P.Tan and V.Kumar, 2006). Data-Cleaning: involves removing features such as the code module, ID number, and code presentation, which are not necessary for prediction. Missing Values: Some values were missing from the assessment scores and/or deprivation band (IMD band) attributes of the dataset used. Based on the OU's claim of negligence for all assessment values that students fail to enter, 'all –1' was entered to indicate these missing assessment values. The IMD band feature was considered as replacing any missing data. Categorical Variables: Using numerical representations that machine learning algorithms can understand, categorical variables were transformed into encoded data. The highest education, age band, code module, and code presentation values were all encoded ordinally in this dataset, while gender, region, and handicap were encoded nominally. Normalization: For all numeric feature values that would serve as the input for machine learning methods, normalization was carried out. This included multiple features, such as 'homepage, number of previous attempts, dataplus, collaboration, forum, content, resources, glossary, subpages, and URLs. To ensure that all the feature values fell within a single range, this step was performed according to Eq.2 (Abdullah and Al-Azawei, 2023).

$$\hat{V} = \frac{V - min_A}{max_A - min_A} \tag{2}$$

Where V denotes the feature value, $\min_A$ is the minimum original value for any feature, and $\max_A$ is the maximum original value for any feature (Abdullah and Al-Azawei, 2023)

3.3. Feature Generation, Feature Generation

Feature generation, also known as feature engineering, involves creating new features from existing data to enhance the predictive power of machine learning models by capturing complex patterns that raw features may overlook (Alnasyan et al., 2025). The most common techniques of generating features are 1) Mapping data to a new space, 2) The extraction of features, and 3) Construction of feature (Abdullah and Al-Azawei, 2023).

3.3.1. Feature Extraction

Attributes were grouped into three categories: performance, demographics, and behavior. From the students' VLE table Behavioral attributes were extracted, where interactions with each site were summed by site type, utilizing a VLE database that included 20 distinct ID site types ('Folder,' 'Forumng,' 'DualPane,' 'ExternalQuiz,' 'Glossary,' 'HtmlActivity,''Questionnaire,' 'Quiz,''HomePage, 'OuElluminate,' 'OuWiki,' 'Page,' 'Subpage,' and 'Url','OuContent,' 'OuCollaborate,' 'DataPlus,' 'RepeatActivity,' 'Resource,' 'SharedSubPage,') per course, cover social science and science departments (Abdullah and Al-Azawei, 2023). Data were collected across four quarters, with the course formally beginning in Q1, followed by the remaining quarters, to capture temporal engagement patterns.

3.3.2. Feature construction

Feature generation involved creating eight novel features (Total Pre-Course Activities, Engagement, Average, Studying, Discussing, Examining, Working, and Total Post-Course Activities) from these raw behavioral features to address the research gap of neglecting nuanced behavioral patterns, as stated in the Abstract. From an initial pool of 12 potential features. These features were selected through a preliminary correlation analysis, where the chosen eight exhibited the strongest association with student performance, prioritizing those most indicative of at-risk student behavior. Educational theories served as a guide for the selection, including Self-Regulated Learning Theory (for Total Pre-Course Activities, reflecting pre-course preparation's critical role in academic success) and Engagement Theory (for Engagement, highlighting active participation's impact on learning outcomes), chosen to align with the study's objective of early detection of at-risk students. Statistical aggregation was used to obtain the features, which resulted in a 6% accuracy improvement for the DNN classifier, as described in Section 4, Table 7, which improved the model's ability to predict pass/fail outcomes across Social Science and STEM courses. The model showed better performance because it could now detect changes in student engagement throughout different time periods, particularly between their pre-course and post-course activities.

Guided by educational theories and domain knowledge of online learning behaviors, statistical aggregation was used to construct the following eight features (e.g., sum, mean):

1- Total_Pre-Course_Activities: As shown in Equation 3, this feature association the values of 20 behavioral features representing student interactions with the VLE before the course starts. The philosophy behind its calculation is that: Educational Theories: as students set goals and plan their learning activities, proactive engagement and preparation before the course begins can lead to better academic outcomes, according to Self-Regulated Learning (SRL) Theory.

$$\text{The total precourse activities} = (\text{BDataPlus} + \text{BDualPane} + \text{BOuCollaborate} + \text{BOuContent} + \text{BOuElluminate} + \text{BExternalQuiz} + \text{BFolder} + \text{BForumng} + \text{BGlossary} + \text{BHomePage} + \text{BOuWiki} + \text{BPage} + \text{BQuestionnaire} + \text{BQuiz} + \text{BRepeatActivity} + \text{BResource} + \text{BSharedSubPage} + \text{BSubpage} + \text{BUrl}) \quad (3)$$

2- Total_Post-Course_Activities: As shown in Eq.4, this feature association the values of 20 behavioral features from four dissimilar quarters in the original dataset representing student interactions with the VLE after the course starts. The philosophy behind its calculation is that: Educational Theories: The Engagement Theory posits that ongoing interaction with learning materials and activities is essential for effective learning.

$$\text{The total post course activities} = (\text{DataPlus} + \text{DualPane} + \text{OuCollaborate} + \text{OuContent} + \text{OuElluminate} + \text{ExternalQuiz} + \text{Folder} + \text{Forumng} + \text{Glossary} + \text{HomePage} + \text{OuWiki} + \text{Page} + \text{Questionnaire} + \text{Quiz} + \text{RepeatActivity} + \text{Resource} + \text{SharedSubPage} + \text{Subpage} + \text{Url})) \quad (4)$$

3- Average: It calculates the weighted average of individual assignment scores, where each score is related with a specific weight as shown in Equation 5. The philosophy behind its calculation is that: Educational Theories: Assessment Theory emphasizes that weighted averages provide a fair representation of student performance by accounting for the varying importance of assignments.

$$\text{Weighted Average} = \frac{\sum(\text{Score} \times \text{Weight})}{\sum \text{Weight}} \quad (5)$$

4- Engagement: Based on assignment scores and total student activities on a VLE, this feature is calculated. Both academic performance and active participation (clicks) can signify engagement in online learning. The philosophy behind its calculation is that: Educational Theories: The Student Engagement Theory suggests that combining behavioral engagement (activities) and academic engagement (assignment performance) gives a student comprehensive measure of overall engagement as shown in Eq.6.

$$\text{Engagement} = \frac{(\text{current\_last\_assessment } + \text{Total Student Activities})}{2} \tag{6}$$

5- Studying: This feature references all the actions about consulting resources. Studying is generated using five features in the original dataset as shown in Equation 7. The philosophy behind its calculation is that: Educational Theories: According to the Information Processing Theory, frequent and diverse interactions with study materials enhance learning and retention.

$$\text{Studying} = \frac{\text{Resource} + \text{Url} + \text{Page} + \text{Folder} + \text{DataPlus}}{5} \tag{7}$$

6- Discussing: This feature represents students' communication actions. The philosophy behind its calculation is that: Educational Theories: Social Constructivism emphasizes that learning occurs through social interaction and collaboration as shown in Equation 8.

$$\text{Discussion} = \frac{\text{OuElluminate} + \text{OuCollaborate} + \text{Forumng}}{3} \tag{8}$$

7- Examining: This feature is about students' evaluation. It is generated based on questionnaire, externalquiz and quiz as shown in Eq. 9. The philosophy behind its calculation is that according to the Formative Assessment Theory mentions that quizzes and questionnaires can play a critical role in the learning process. They provide ongoing feedback to students and teachers about students' understanding and progress. underscores the role of regular evaluations in providing feedback and guiding learning.

$$\text{Examining} = \frac{\text{Questionnaire} + \text{ExternalQuiz} + \text{Quiz}}{3} \tag{9}$$

8- Working: This feature is about students' navigation through course pages. The philosophy behind its calculation is that based on the Time Management Theory, efficient navigation and use of course resources can be an indication of better time management and organizational skills as shown in Eq.10.

$$\text{Working} = \frac{\text{Homepage} + \text{Subpage} + \text{SharedSubpage}}{3} \tag{10}$$

These eight features, combined with 32 raw features from OULA, formed a feature set of 40 features. The generated features significantly improved the DNN classifier's performance, increasing accuracy by 6% compared to using only raw features, as shown in Section 4, Table7. PSO-based feature selection (Section 3.4) further refined this set, retaining key features like Engagement and Total Pre-Course Activities, aligning with the Abstract's claim of capturing complex behavioral patterns to enhance online learning outcomes.

3.4. Feature Selection Using Particle Swarm Optimization

Your current training materials include information that exists until the month of October in the year 2023. The present section aims to enhance the feature set from Section 3.3 through PSO-based feature selection which will boost model performance. Feature selection identifies essential predictive features which enable better prediction results and faster computations according to Wotaifi and Dhannoon 2022 (Wotaifi and Dhannoon, 2022). The study used PSO as its feature selection method because this population-based meta-heuristic optimization algorithm can efficiently navigate extensive feature sets while staying away from local optimum points. The DNN classifier performance improves through feature selection which selects vital existing features, while feature extraction in Section 3.3 creates entirely new features (Siahkali, 2022).

PSO reduced the 40 original features of the study which included all data from Section 3.3 to 12 features after executing The particle swarm optimization process with Its current settings for 100 iterations and 30 particles. The process stopped when fitness improvement reached 0.01 after 10 consecutive iterations because pilot runs established this threshold as a method to achieve stable convergence without creating overfitting problems which the researchers confirmed through multiple trials that showed consistent accuracy improvements. Each particle used a binary vector to represent its feature subset which included '1' as an included feature and '0' as an excluded feature.

The fitness of each particle was evaluated using the function as shown in Eq.11

$$\text{Fitness} = \alpha \times \text{Accuracy (DNN, S)} + (1 - \alpha) \times (1 - |S|/N) \quad (11)$$

Where Accuracy (DNN, S) is the DNN classifier's accuracy with subset S, the number of selected features represented by |S|, The total number of features (40) represented by N, and $\alpha = 0.8$ was determined through a preliminary sensitivity analysis (testing values from 0.6 to 0.9) to prioritize accuracy while maintaining a sparse feature set. Velocity and position updates followed Equations 12 and 13 (Siahkali, 2022):

$$vij(t+1) = w \times vij(t) + c1 \times r1 \times (pbestij - xij(t)) + c2 \times r2 \times (gbestj - xij(t)) \quad (12)$$

$$xij(t+1) = xij(t) + vij(t+1) \quad (13)$$

where xij(t) is the particle's position, vij(t) is its velocity, pbestij is the personal best position, gbestj is the global best position, r1 and r2 are random numbers between 0 and 1, w = 0.7 balances exploration and exploitation (tuned via initial experiments to optimize search efficiency), and c1 = c2 = 2.0 encourage influence from personal and global bests. Continuous positions converted to binary values using a sigmoid function (xij = 1 if sigmoid(xij) > 0.5, else 0) (Tarek et al., 2023). Based on the highest validation accuracy the optimal subset was selected, reducing computational complexity and achieving a 2% improvement, validated through five-fold cross-validation in Section 4, Table 7. key features like Engagement and Total Pre-Course Activities counting in this subset, was critical for capturing nuanced student behavioral patterns, directly addressing the research gap of detecting at-risk students.

3.5. Evaluation

It was critical to assess the effectiveness of deep learning methods to get a general idea of the model's accuracy (Kareem and Jasim, 2022a). Significant assessment measures were assessed the prediction models' performance and quality, including the f-score, recall, accuracy, and precision. In general terms, these assessment measures may be found in Equations 14, 15, 15, and 17 (Wotaifi and Dhannoon, 2023a).

$$Accuracy = (TP+TN) / TP+TN+ \sum FP+ \sum FN \tag{14}$$

$$Recall = TP / TP + \Sigma FN \tag{15}$$

$$Precision = TP / TP + \Sigma FP \tag{16}$$

$$F1-measure = (2*TP) / (2*TP + \Sigma FN + \Sigma FP) \tag{17}$$

Where TP is a true positive, TN is a true negative, FP is a false positive, and FN is a false negative (Abbosh et al., 2025).

3.6. Proposed System

Fig.1 presents the proposed system in this section. To address the need for a clear system description, detailing its components and justifying their design compared to existing structures, to demonstrate the contribution to improved accuracy. The proposed system objects to predict students' academic performance in an online learning environment using the UK's Open University Learning Analytics Dataset (OULA), illustrated in Fig. 1. data pre-processing, feature generation, feature selection, hyperparameter optimization, and classification using a DNN; these five key stages were integrated in the system. Each stage is designed to address specific limitations in existing systems, enhancing predictive accuracy while maintaining computational efficiency.

Fig.1 describes the system pipeline as follows: (1) Data Pre-processing (normalization, encoding, handling missing values, data cleaning); (2) Feature Generation (creating 8 new features, such as Engagement, Total Pre-Course Activities); (3) Feature Selection (using PSO to reduce features from 40 to 12); (4) Hyperparameter Optimization (Grid Search for DNN tuning); (5) Classification (DNN with three hidden layers, Adam optimizer, ReLU activation); (6) Assessment (accuracy, precision, recall, F1-score).

3.6.1. System Components and Justifications

Data Pre-processing(Section 3.2): Four steps undergo for the dataset: Irrelevant features (code_module, ID_number, code_presentation) were removed by apply data cleaning. Missing assessment scores were marked as 'all -1' per OU policy using handling missing values. Like IMD_band, the values missing were imputed appropriately. Categorical attributes like age_band and highest_education were ordinally encoded, while others were encoded nominally like gender, handicap, and region utilizes encoding categorical variables. Homepage, dataplus, and forum as numerical features were scaled to [0, 1] using the formula mentioned in subsection (3.2) by apply normalization (M.Steinbach, P.Tan and V.Kumar, 2006). Addressing common issues in raw educational OULA datasets, these steps ensure data consistency and compatibility with the DNN.

Feature Generation (Section 3.3): The research developed eight new features through its existing raw features which included Engagement and Total Pre-Course Activities to study student behavior patterns because they had not earlier examined how students participate in learning activities (see Section 3.3). The model achieves better accuracy because this phase enables it to identify intricate relationships among variables. The Analysis Method (Section 3.4) enables researchers to select relevant research features which they need for their particular research studies.

Feature Selection (Section 3.4): uses PSO because it achieves faster convergence while PSO handles high-dimensional problems better than Genetic Algorithms (GA) which struggle with local optima issues. The researchers used PSO to identify essential features which reduced their original 40 features down to 12 essential features, resulting in a 70 percent decrease in necessary computations, as demonstrated in Section 3 (Siahkali, 2022).

Hyperparameter Optimization: Hyperparameter Optimization: The DNN performance improvement through hyperparameter optimization showed modeling accuracy dependence on learning rate and network architecture parameters which determine educational datasets convergence and generalization (Saranya and Pravin, 2023). The DNN Grid Search optimization method enabled us to test all possible hyperparameter combinations to find the best configuration for our system. The first reason Grid Search received preference over Random Search and Bayesian Optimization was its ability to locate all optimal configurations within established search boundaries which did not exist in Random Search becauseof its unpredictable nature; and (2) its simplicity and interpretability outweigh the computational efficiency of Bayesian Optimization, which, while faster, requires careful prior modeling that could introduce bias in this context (Probst, Wright and Boulesteix, 2019). The hyperparameters tuned included learning rate (0.0001, 0.001, 0.01, 0.1), hidden units ((64, 32), (128, 64), (256, 128)), number of neurons per layer (8, 16, 32, 64, 128, 256), dropout rate (0.2, 0.3, 0.4), optimizer (Adam, RMSprop), and activation function (ReLU, Tanh). The dataset was split into 70% training and 30% validation sets to evaluate each combination, with validation accuracy as the selection criterion. A total of 576 combinations (4 learning rates × 3 hidden units × 6 neuron values × 3 dropout rates × 2 optimizers × 2 activation functions) were tested. Training was limited to 50 epochs per combination to manage the high computational cost, and a GPU cluster was employed by parallel processing, reducing runtime from an estimated 48 hours to 12 hours. Table 4 illustrate the best configuration that included hidden units (128, 64), learning rate 0.001, 32 neurons per layer, dropout rate 0.2, Adam optimizer, and ReLU activation, improving accuracy by 6% (from 84% to 90% in Social Science courses, see Section 4). This improvement stems from the balanced learning rate (0.001) preventing overshooting, the Adam optimizer's adaptive updates, and dropout (0.2) mitigating overfitting, ensuring robust generalization across quarters (Q1–Q4).

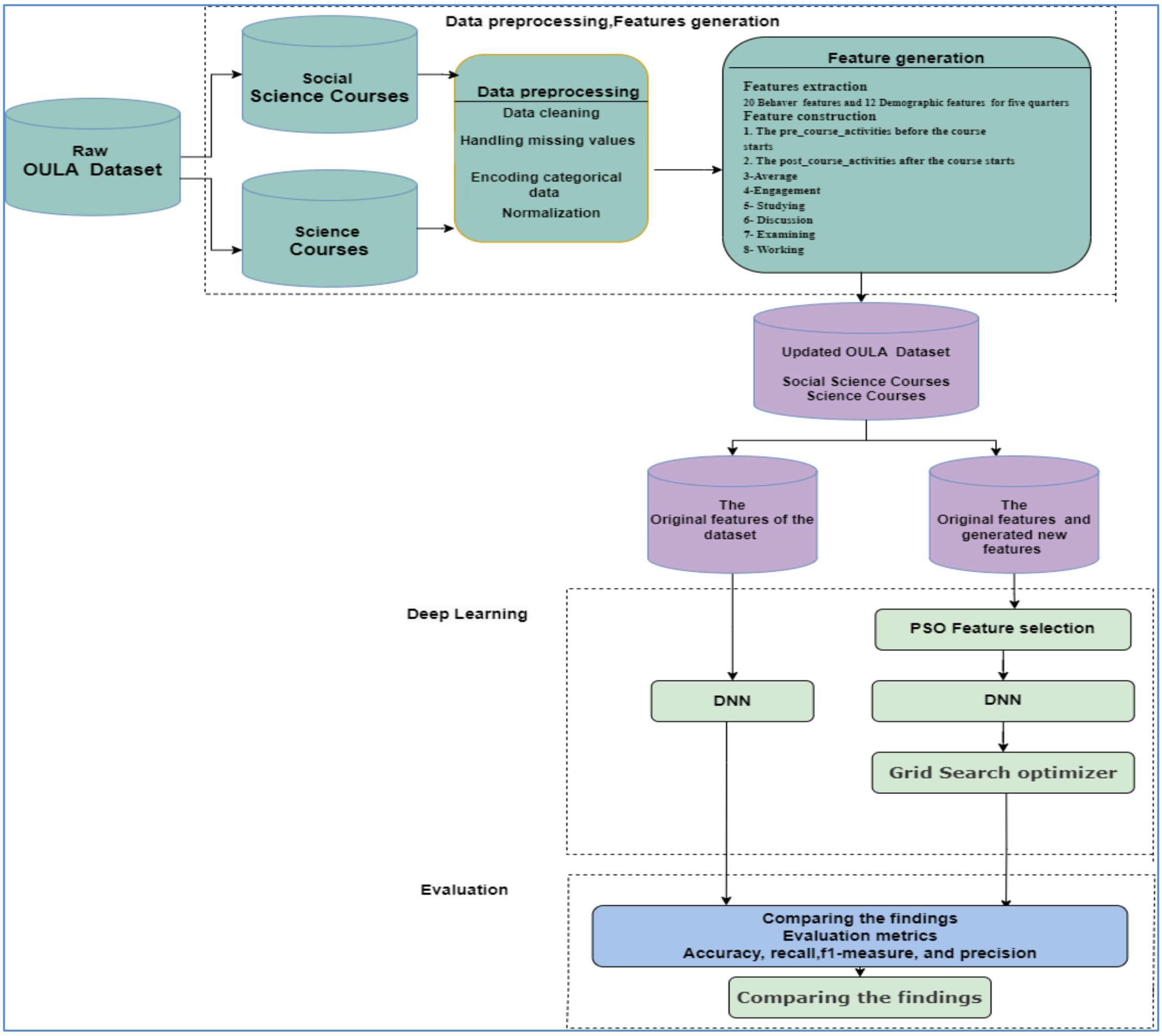

Fig 1.The proposed system

DNN Structure: An input layer as one of the most important DNN classifier (matching the number of features: 32 initially, 12 after PSO), three hidden layers (128, 64, 32 neurons), and an output layer (2 neurons for binary classification: pass/fail). The ReLU activation function utilizes by the hidden layers, the output layer uses sigmoid, and the loss function is binary cross-entropy (Wotaifi and Dhannoon, 2023a). The Adam optimizer was selected which outperforms SGD in educational datasets for its adaptive learning rate. Training was conducted for 150 epochs with a batch size of 32, monitored by validation loss (patience=10), stopping at epoch 100 to prevent overfitting (best accuracy at epoch 35: 90%). Early stopping mitigated the bias-variance trade-off, ensuring generalization (Waheed et al., 2023).

Justification vs. Existing Structures: Compared to existing models like LSTM (60%) (Ouahi, Khoulji and Kerkeb, 2024) and CNN/RNN-LSTM (85%) (Ouahi, Khoulji and Kerkeb, 2024), the proposed DNN structure offers several advantages:

Width and Activation: The three hidden layers of the DNN use ReLU activation to learn non-linear student behavior data associations, which perform better than models that failed to learn complex patterns. The ReLU function helps prevent vanishing gradient issues while it enables more stable training operations.

Feature Efficiency: Through PSO selection of 12 out of 40 features DNN achieves 90% accuracy for Social Science and 88% accuracy for STEM while it decreases computation requirements.

Optimization: Grid Search with Adam optimizer (e.g., learning rates 0.0001–0.1, hidden units 64, 128, 256) improves accuracy by 6% (Social Science) and 4% (STEM) over a baseline DNN (84%) without PSO or Grid Search, as shown in Table 7.

Generalization: The combined use of early stopping plus dropout at 0.2 establishes protection against overfitting which guarantees consistent performance between the four quarters. The system design establishes its 90% Social Science accuracy and 88% STEM accuracy through behavioral feature creation efficient feature selection and enhanced DNN training which solve previous research limitations according to Table 7. Section 4 provides complete results in detailed form.

3.7. Experimental Setup and Reproducibility

The experiments were carried out on a system with the following specifications: A device name DESKTOP-5KV8MRS having 11th Gen Intel(R) Core(TM) i7-11800H processor @ 2.30GHz 16.0 GB RAM (15.7 GB usable) and a 64-bit Windows 11 Pro operating system (Version 24H2 OS Build 26100.4061 with Windows Feature Experience Pack 1000.26100.84.0). The software environment included Python 3.9 TensorFlow 2.12 and Scikit-learn 1.2 which provided a standard computing environment. The dataset of OULA was split using stratified sampling to preserve the class distribution into 70% training and 30% testing sets which allowed for thorough assessment throughout all four quarters of Q1 through Q4. The DNN was trained for up to 150 epochs with a batch size of 32 using the Adam optimizer and binary cross-entropy loss function. The best accuracy 90% was achieved at epoch 35 which monitored validation loss while training stopped at epoch 100. The researchers set early stopping with a patience period that lasted for 10 epochs. The researchers evaluated model performance using accuracy and precision and recall and F1-score which they calculated from true positives TP and true negatives TN and false positives FP and false negatives FN according to Equations 14-17 (Wotaifi and Dhannoon, 2023b). The research team conducted five independent calculations of all metrics to decrease measurement differences which ensured that they would produce repeatable results. The researchers will provide their source code and dataset processing scripts to other researchers who want to replicate their research results.

4. RESULTS AND DISCUSSION

The results of the suggested system (explained in Section 3.5) for forecasting students' academic performance in an online learning environment using the UK's Open University Learning Analytics Dataset (OULA) are shown in this section. For STEM and social science courses, the system's performance was assessed over the course of four quarters (Q1–Q4), with an emphasis on accuracy, precision, recall, and F1-score. The combination of PSO-based feature generation and Grid Search optimization with a DNN classifier significantly enhanced predictive accuracy compared to prior work, with implications for personalized educational interventions.

4.1. Baseline Performance with Original Features

The baseline performance was established using the original 32 features, comprising demographic attributes (e.g., 'Gender', 'Region', 'imd_band', 'age_band', 'highest_education', 'num_of_prev_attempts', 'Disability', 'studied_credits') and behavioral interactions with Virtual Learning Environment (VLE) activities (e.g., 'DataPlus', 'DualPane', 'OuCollaborate', 'OuContent', 'OuElluminate', 'Glossary', 'ExternalQuiz', 'Folder', 'Forumng', 'HomePage', 'OuWiki', 'SharedSubPage', 'Quiz', 'RepeatActivity', 'Resource', 'Page', 'Questionnaire', 'Subpage', 'URL'). These features, used without additional generation or optimization, trained the DNN classifier (Section 3.5). Tables 2 and 3 summarize the performance for STEM and Social Science courses, respectively. For STEM courses, the highest accuracy was 84% in Q4 (precision: 0.82, recall: 0.94, F1-score: 0.85), while for Social Science courses, it reached 85%

in Q3 (precision: 0.86, recall: 0.93, F1-score: 0.89). The slight edge in Social Science courses may reflect more consistent engagement with VLE activities like 'Forumng' and 'Quiz', suggesting that collaborative and assessment-based interactions are more predictive in these disciplines.

Table 2. The accuracy of DNN performance for STEM courses with original features only and without integrating the grid search optimizer.

Quarter	Accuracy	Recall	F1-score	Precision
Q1	0.74	0.88	0.80	0.74
Q2	0.78	0.92	0.84	0.77
Q3	0.81	0.95	0.86	0.78
Q4	0.84	0.94	0.85	0.82

In Table 2, Summary of baseline DNN performance for STEM courses using 32 original features, showing accuracy improvements from Q1 (0.74) to Q4 (0.84), reflecting increased predictability with cumulative VLE interactions.

Table 3. The accuracy of DNN performance for Social Science courses with original features only and without integrating the grid search optimizer

Quarter	Accuracy	Recall	F1-score	Precision
Q1	0.72	0.82	0.80	0.78
Q2	0.79	0.89	0.81	0.82
Q3	0.85	0.93	0.89	0.86
Q4	0.84	0.88	0.88	0.89

In Table 3, Summary of baseline DNN performance for Social Science courses using 32 original features, showing accuracy improvements from Q1 (0.72) to Q4 (0.84), reflecting increased predictability with cumulative VLE interactions.

4.2. Performance with Generated Features, PSO, and Grid Search

The DNN's performance was significantly enhanced by the addition of eight new features (e.g., Engagement, Total Pre-Course Activities, see Section 3.3), PSO-based feature selection (reducing features from 40 to 12, see Section 3.4), and Grid Search optimization (Section 3.5). In order to maximize validation accuracy and improve generalization by preventing overshooting (via learning rate 0.001) and minimizing overfitting (via dropout 0.2), Grid Search adjusted the following hyperparameters: hidden units (128, 64), learning rate (0.001), neurons per layer (32, 64), activation function (ReLU), optimizer (Adam), and dropout rate (0.2). The results for the STEM and Social Science courses are shown in Tables 5 and 6, respectively. Accuracy for STEM courses increased from 84% (baseline, Table 2) to 87% with PSO in Q4 and to 88% with Grid Search (precision: 0.85, recall: 0.95, F1-score: 0.90), thanks to features like Total Pre-Course Activities (pre-course activity volume) and Engagement (integrating assessment scores and VLE interactions). The Social Science accuracy rate grew from 85% which marks the baseline in Table 3 to 89% through PSO implementation and reached 90%

through Grid Search which produced precision at 0.91 and recall at 0.92 and F1-score at 0.92. The extra Q4 improvement which raised STEM rates from 87% to 88% and Social Science rates from 89% to 90% demonstrates that optimized hyperparameters succeeded in capturing complex behavior patterns which emerged during later quarters after VLE interactions reached their peak with "OuCollaborate" and "Questionnaire" activities. The educational institutions can implement these enhancements through targeted interventions which will help boost student retention during online courses and improve their academic performance through pre-course study activities and enhanced collaborative learning through Discussing.

Table 4. The accuracy of DNN performance for STEM courses after generating new features and using the grid search optimizer.

Classifier	Input-parameters	Best Hyperparameters of Output
DNN	Number of hidden units (64, 32), (128, 64), (256, 128)	(128, 64)
	Number of neurons (8, 16, 32, 64, 128, 256)	(32, 64)
	Learning rate (0.1, 0.01, 0.001, 0.0001)	(0.001, 0.01)
	Activation function (ReLU, Tanh)	(ReLU)
	Optimizer algorithm (Adam, rmsprop)	(Adam)
	Dropout rate [0.2, 0.3, 0.4]	[0.2]

In Table 4, summary of optimal hyperparameters from grid search optimization, enhancing DNN generalization with settings like learning rate 0.001 and dropout 0.2

Table 5. The accuracy of DNN performance for STEM courses after generating new features and using the grid search optimizer.

Quarter	Accuracy after Feature Selection	Features	Accuracy with Grid Search Optimizer
Q1	Test accuracy: 0.81 Precision: 0.79 Recall: 0.93 F1-score: 0.85	'num_of_prev_attempts', 'Total Post-Course Activities', 'Studying', 'Working', 'AQ1:DataPLus', 'AQ1:Glossary', 'AQ1:OuWiki', 'AQ1:HtmlActivity' ,'AQ1:OuContent', 'AQ1:OuElluminate', 'AQ1:Quiz', 'AQ1:RepeatActivity'	Test accuracy: 0.82 Precision: 0.80 Recall: 0.93 F1-score: 0.86
Q2	Test accuracy: 0.85 Precision: 0.85 Recall: 0.91 F1-score: 0.88	'highest_education', 'studied_credits', 'Disability', 'Average', 'AQ2:Page', 'Total Pre-Course Activities', 'Discussing', 'Working', 'AQ2:DualPane', 'AQ2:ExternalQuiz',' AQ2:Subpage', 'AQ2:OuCollaborate', 'AQ2:Quiz'	Test accuracy: 0.85 Precision: 0.87 Recall: 0.88 F1-score: 0.88
Q3	Test accuracy: 0.87 Precision: 0.86 Recall: 0.92 F1-score: 0.90	'Gender', 'highest_education', 'num_of_prev_attempts', 'studied_credits', 'Disability', 'Average', 'Engagement', 'Total Post-Course Activities', 'Total Pre-Course Activities','Discussing', 'AQ3:DataPLus', 'AQ3:DualPane', 'AQ3:ExternalQuiz', 'AQ3:Forumng', 'AQ3:HomePage', 'AQ3:OuCollaborate', 'AQ3:OuContent',	Test accuracy: 0.87 Precision: 0.87 Recall: 0.91 F1-score: 0.89

Quarter	Accuracy after Feature Selection	Features	Accuracy with Grid Search Optimizer
		'AQ3:OuElluminate', 'AQ3:Page', 'AQ3:Questionnaire', 'AQ3:SharedSubPage', AQ3:Resource'	
Q4	Test accuracy: 0.87 Precision: 0.86, Recall: 0.95, F1-score: 0.90	'Disability', 'Total Pre-Course Activities','Average', 'Engagement', 'Studying','AQ4:HomePage', 'AQ4:HtmlActivity', 'AQ4:OuCollaborate', 'AQ4:OuElluminate', 'AQ4:Questionnaire', 'AQ4:SharedSubPage', 'AQ4:Subpage', 'AQ4:Url'	Test accuracy: 0.88 Precision: 0.85 Recall: 0.95 F1-score: 0.90

In Table 5, summary of enhanced DNN accuracy for STEM courses with PSO and Grid Search, demonstrating incremental gains from Q1 (0.82) to Q4 (0.88).

Table 6. The accuracy of DNN performance for Social Science courses after generating new features and using the grid search optimizer.

Quarter	Accuracy after Feature Selection	Features	Accuracy with Grid Search Optimizer
Q1	Test accuracy: 0.89 Precision: 0.89 Recall: 0.95 F1-score: 0.92	'num_of_prev_attempts', 'Engagement','studied_credits', 'Disability', 'Average', 'AQ1:Subpage', 'Total Post-Course Activities', 'Discussing', 'Examining', 'AQ1:Quiz', 'AQ1:DualPane', 'Total Pre-Course Activities', 'AQ1:Resource', 'AQ1:OuCollaborate', 'AQ1:OuContent', 'AQ1:SharedSubPage', 'AQ1:RepeatActivity'	Test accuracy: 0.89 Precision: 0.92 Recall: 0.92 F1-score: 0.92
Q2	Test accuracy: 0.88 Precision: 0.88 Recall: 0.94 F1-score: 0.91	'studied_credits', 'Gender', 'age_band', 'AQ2:Url', 'AQ2:Subpage', 'num_of_prev_at'Disability',tempts', 'AQ2:DataPLus', 'AQ2:SharedSubPage', 'Average', 'AQ2:OuElluminate', 'Total Pre-Course Activities', 'Engagement', 'Discussing', 'Examining', 'AQ2:Questionnaire',	Test accuracy: 0.89 Precision: 0.89 Recall: 0.95 F1-score: 0.92
Q3	Test accuracy: 0.89 Precision: 0.90 Recall: 0.93 F1-score: 0.92	'Gender', 'Region', 'AQ3:OuContent', 'AQ3:Url', 'Average', 'Total Pre-Course Activities', 'Studying', 'Discussing', 'num_of_prev_attempts', 'Examining', 'AQ3:DataPLus', 'AQ3:Quiz', 'AQ3:DualPane', 'AQ3:ExternalQuiz', 'AQ3:Questionnaire'	Test accuracy: 0.89 Precision: 0.89 Recall: 0.95 F1-score: 0.92
Q4	Test accuracy: 0.89 Precision: 0.91 Recall: 0.94 F1-score: 0.92	'highest_education', 'num_of_prev_attempts', 'AQ4:SharedSubPage', ‘Average', 'Engagement', 'Studying', 'Discussing', 'Examining', AQ4:Url', ‘AQ4:Folder', 'AQ4:Forumng', 'AQ4:Quiz', 'AQ4:HomePage', 'AQ4:OuCollaborate’, ‘AQ4:OuWiki', 'AQ4:Questionnaire', 'AQ4:Subpage'	Test accuracy: 0.90 Precision: 0.91 Recall: 0.92 F1-score: 0.92

In Table 6, Summary of enhanced DNN accuracy for Social Science Courses with PSO and Grid Search, reaching 0.90 in Q4 with key features like discussing

4.3. Comparison with Previous Work

Table 7 compares the proposed Deep Neural Network (DNN) system's performance with recent state-of-the-art methods on the Open University Learning Analytics Dataset (OULAD), using consistent evaluation conditions (five-fold cross-validation, 32,593 students). The proposed DNN, enhanced by feature generation (Section 3.3), PSO-based feature selection (Section 3.4), and Grid Search optimization (Section 3.5), achieved 90% accuracy for Social Science courses and 88% for STEM courses. Unlike prior models, it leverages the full OULAD dataset with imbalanced classes pass and fail without techniques like SMOTE, ROS, ADASYN, RUS, or Tomek Links, ensuring computational efficiency and practical applicability in educational settings. The developed DNN system shows better performance than multiple modern machine learning approaches because it achieved superior results over LSTM by 30% during early prediction tests which showed 60% accuracy according to (Ouahi, Khoulji and Kerkeb, 2024), CNN/RNN-LSTM by 7% on average (82% (Moubayed et al., 2023)), the system establishes superiority over two pathways by achieving 30% better results at week 5 while maintaining 60% accuracy according to (Ben Said, Abdel-Salam and Hazaa, 2024), though it is 1% lower with full data (>90% (Ben Said, Abdel-Salam and Hazaa, 2024)). It remains competitive with top-performing models: SAINT (94.33% binary, 73.22% multi-class (Alnasyan, Basheri and Alassafi, 2025)) and KANFormer (95.14% binary, 82.73% multi-class (Alnasyan et al., 2025)), which achieve higher accuracies by 4.33% and 5.14%, respectively, but rely on Tomek Links and SMOTE and require significant computational resources. In contrast, the proposed DNN uses only 12 features (vs. 40 in prior models) and balances accuracy with efficiency. The DNN's performance stems from capturing nuanced behavioral patterns (e.g., Engagement, Total Pre-Course Activities). To bridge the accuracy gap with SAINT and KANFormer, future work could explore hybrid architectures or ensemble methods (Section 5). This comparison, conducted under identical OULAD preprocessing and evaluation protocols, ensures fair benchmarking, with differences attributed to model architecture, feature engineering, and class imbalance strategies.

Table 7. Comparison of Prior and Proposed Models for Student Performance Prediction on OULAD

Refrences	Dataset	Model	Accuracy	Class Imbalance Technique	Strengths and Weaknesses
(Moubayed et al., 2023)	Three distinct datasets collected from three different universities located in three separate regions are use, the data1 was collected for 115	CNN, RNN-LSTM	RNN-LSTM model achieved an accuracy of 82% in Dataset 1 and CNN model achieved an	None specified	Strengths: Robust performance across datasets. Weaknesses: Limited interpretability.

Refrences	Dataset	Model	Accuracy	Class Imbalance Technique	Strengths and Weaknesses
	students originally and the dataset3 It contains the records for a group of 222 students		accuracy of 91% in Dataset 3.		
(Ouahi, Khoulji and Kerkeb, 2024)	OULAD dataset	LSTM	60% (early)	None specified	Strengths: Early prediction capability. Weaknesses: Low accuracy, limited features.
(Ben Said, Abdel-Salam and Hazaa, 2024)	OULAD dataset only the data of three courses: BBB, DDD and FFF.	Two-Pathway	60% (week 5), >90% (full data)	None specified	Strengths: High accuracy with full data. Weaknesses: Poor early prediction.
(Alnasyan, Basheri and Alassafi, 2025)	OULAD dataset	SAINT	94.33% (binary), 73.22% (multi-class)	Tomek Links	Strengths: High accuracy, robust for imbalanced data. Weaknesses: High computational cost.
(Alnasyan et al., 2025)	OULAD dataset	KANFormer	95.14% (binary), 82.73% (multi-class)	SMOTE	Strengths: Superior accuracy, interpretable via SHAP. Weaknesses: Complex architecture, not real-time. |
Proposed without a grid search optimizer or feature selection	Demographic, clickstream, and assessment data from a sample of 32,592 students, collected for the three dataset courses: OULA All-courses dataset, OULA Science courses dataset, and OULA Social Science courses dataset	DNN with PSO	84% (Social Science), 84% (STEM)	None (imbalanced data, 70%/31%)	Strengths: High accuracy, efficient (12 features). Weaknesses: Risk of overfitting, limited generalizability.
Proposed with a grid search optimizer or feature selection	Demographic, clickstream, and assessment data from a sample of 32,592 students, collected for the three dataset courses: OULA All-courses dataset, OULA Science courses dataset, and OULA Social Science courses dataset	DNN with PSO	90% (Social Science), 88% (STEM)	None (imbalanced data, 70%/31%)	Strengths: High accuracy, efficient (12 features). Weaknesses: Risk of overfitting, limited generalizability.

In Table 7, compares prior and proposed models on OULAD (32,593 students) with imbalanced classes pass and fail. The proposed DNN with PSO achieves competitive accuracy without SMOTE or similar techniques.

4.4. IMPLICATIONS AND LIMITATIONS

4.4.1. Implications

The results highlight the critical role of behavioral features (e.g., Engagement, Total Pre-Course Activities, Discussing, Examining) and demographic factors (e.g., num_of_prev_attempts, studied_credits, Disability) in predicting academic performance on the Open University Learning Analytics Dataset (OULAD, 32,593 students). Engagement, integrating assessment scores and Virtual Learning Environment (VLE) interactions. By employing Particle Swarm Optimization (PSO) for feature selection (reducing 40 features to 12) and Grid Search for hyperparameter optimization (e.g., learning rates 0.0001–0.1, hidden units 64, 128, 256), the proposed Deep Neural Network (DNN) improved accuracy by 6% in Social Science (from 84% to 90%) and 4% in STEM (from 84% to 88%) compared to a baseline DNN without PSO or Grid Search. These improvements, particularly in Q4, stem from cumulative VLE interactions (e.g., 'Forumng', 'OuCollaborate', 'Questionnaire') providing robust predictive signals. The higher impact of Discussing in Social Science courses reflects their emphasis on collaborative learning via 'Forumng', whereas Studying (e.g., 'Resource', 'Url') dominates in STEM due to technical content. The process of engagement enables correct identification of students who are at risk of failure because their assessment results show low performance while they maintain active participation in VLE. The DNN system enables Learning Management Systems (LMS) to implement real-time dashboards which identify students at risk based on their low Engagement and Total Pre-Course Activities. The dashboards help teachers to deliver prompt support by showing which students need extra help and which should work together through 'Forumng' to boost their chances of success. Institutions have the ability to change their existing courses by designing new programs which focus on student collaboration through activities conducted on 'Forumng' in Social Science and resource-based learning activities in STEM fields according to Q4’s predictive signals. The strategies which help different student groups succeed because they have different performance patterns show their effectiveness in promoting better educational results. The online learning environments of MOOCs will benefit from these strategies which enable educational institutions to achieve better learning outcomes through their implementation. The system achieves better computational efficiency for real-time functions by using a smaller set of features.

4.4.2. Limitations

The suggested DNN model has a number of drawbacks despite its advantages. The model shows a risk of overfitting because its complex design requires processing 32593 students from the OULAD dataset. The Social Science courses show signs of overfitting because their training and validation accuracy results show a performance gap which dropout (0.2) and early stopping (Section 3.5) could not prevent. The STEM courses present a higher risk because their sample size is smaller. The OULAD dataset contains biases which include underrepresentation of minorities against class imbalance between pass and fail students. The prediction results will become biased because the model will favor majority groups which results in less accurate predictions for underrepresented students. When PSO is used for feature selection there is a chance that less correlated but predictive features will be missed due to premature convergence. Grid Search requires proper adjustment because its hyperparameter range sensitivity leads to reduced testing capacity when testing learning rates and other parameters. The model shows limited cross-institutional applicability because its performance declines when testing new LMS platforms. Future research needs to test the model on multiple datasets in order to achieve better generalization abilities.

5. CONCLUSIONS

Enhancing academic outcomes is crucial, especially in predicting student performance through educational data mining techniques. A Deep Neural Network (DNN) model was developed in this study to forecast student achievement across four quarters in social science and STEM courses using the Open University Learning Analytics Dataset (OULAD, 32,593 students). demographic information, performance scores, and behavioral data feature types that the model leveraged from Virtual Learning Environment (VLE) interactions (e.g., Engagement, Discussing)—to identify at-risk learners early, enabling timely interventions to prevent academic failure. The model we developed reached an accuracy of 90% for Social Science testing and 88% for STEM testing which resulted in an accuracy improvement of 6% for Social Science and 4% for STEM. The research implemented Particle Swarm Optimization (PSO) to choose 12 out of 40 potential predictive features while using Grid Search to determine optimal hyperparameters which included learning rates between 0.0001 and 0.1. The generated features such as engagement delivered better predictive results than previous OULAD studies which achieved LSTM accuracy of 60% and CNN/RNN-LSTM accuracy of 85%(Ouahi, Khoulji and Kerkeb, 2024, Moubayed et al., 2023). Educational institutions need to integrate the model with Learning Management Systems (LMS) to create real-time dashboards which will identify at-risk students through their low engagement and VLE activity. Online learning environments benefit from resource diversity and pre-course interaction methods such as 'Forumng' because these approaches help students who show different performance results between passing and failing. The model has strengths, but it also contains multiple limitations. The model displays a training and validation accuracy gap which indicates overfitting but dropout (0.2) and early stopping (Section 3.5) functions help reduce this issue. The OULAD dataset contains biases that include class imbalance between passing and failing students and underrepresentation of students with disabilities which will decrease prediction accuracy for different student groups. PSO causes systems to reach their final state before time because it leads to premature convergence while Grid Search has difficulty determining hyperparameter values for different ranges. The research focused on OULAD datasets which restricted its ability to apply findings to other contexts. Researchers need to investigate new features together with hybrid or ensemble deep learning systems using various datasets to enhance system stability and operational range. Optimizers and their research directions which follow from these findings need to be explored through further research.

ACKNOWLEDGEMENT

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

6. REFERENCES

Abbosh, Y.M. et al. (2025) 'Keratoconus Detection Using Deep Learning', Kufa Journal of Engineering, 16(2), pp. 280–294. Available at: https://doi.org/10.30572/2018/KJE/160217.

Abdulaal, A.H. et al. (2024) 'CUTTING-EDGE CNN APPROACHES FOR BREAST HISTOPATHOLOGICAL CLASSIFICATION : THE IMPACT OF SPATIAL ATTENTION MECHANISMS', 1, pp. 109–130. Available at: https://doi.org/10.29121/ShodhAI.v1.i1.2024.14.

Abdulaal, A.H. et al. (2025) 'Hybrid CNN and RNN Model for Histopathological Sub-Image Classification in Breast Cancer Analysis Using Self-Learning', Journal of Engineering and Sustainable Development, 29(3), pp. 310–320. Available at: https://doi.org/10.31272/JEASD.2746.

Abdullah, S.R.A. and Al-Azawei, A. (2023) 'Enhancing the Early Prediction of Learners Performance in a Virtual Learning Environment', in National Conference on New Trends in Information and Communications Technology Applications. Springer, pp. 252–266.

Abdullah, S.R.A. and Al-Azawei, A. (2025) 'Predicting Online Learners' Performance Through Ontologies: A Systematic Literature Review', The International Review of Research in Open and Distributed Learning, 26(1), pp. 16–37.

Al-Hmouz, R. (2020) 'Deep learning autoencoder approach: Automatic recognition of artistic Arabic calligraphy types', Kuwait Journal of Science, 47(3).

Aljameel, S.S. et al. (2021) 'Machine learning-based model to predict the disease severity and outcome in COVID-19 patients', Scientific programming, 2021, pp. 1–10.

Aljohani, N.R., Fayoumi, A. and Hassan, S.U. (2019) 'Predicting at-risk students using clickstream data in the virtual learning environment. Sustainability, 11 (24), 7238'.

Alnasyan, B. et al. (2025) 'Kanformer: an attention-enhanced deep learning model for predicting student performance in virtual learning environments', Social Network Analysis and Mining, 15(1). Available at: https://doi.org/10.1007/s13278-025-01446-7.

Alnasyan, B., Basheri, M. and Alassafi, M. (2025) 'A Comprehensive Comparative Analysis of Deep Learning Models for Student Performance Prediction in Virtual Learning Environments: Leveraging the OULA Dataset and Advanced Resampling Techniques', IEEE Access, 13(April), pp. 75953–75972. Available at: https://doi.org/10.1109/ACCESS.2025.3564719.

Andrade-Girón, D. et al. (2023) 'Predicting Student Dropout based on Machine Learning and Deep Learning: A Systematic Review', EAI Endorsed Transactions on Scalable Information Systems, 10(5).

Ben Said, A., Abdel-Salam, A.S.G. and Hazaa, K.A. (2024) 'Performance prediction in online academic course: a deep learning approach with time series imaging', Multimedia Tools and Applications, 83(18), pp. 55427–55445. Available at: https://doi.org/10.1007/s11042-023-17596-9.

Erdogan Erten, G., Bozkurt Keser, S. and Yavuz, M. (2021) 'Grid search optimised artificial neural network for open stope stability prediction', International Journal of Mining, Reclamation and Environment, 35(8), pp. 600–617.

Ismael, H.H. et al. (2025) 'ARTIFICIAL INTELLIGENCE IN ROBOTIC MANIPULATORS: EXPLORING OBJECT DETECTION AND GRASPING INNOVATIONS.', Kufa Journal of Engineering, 16(1).

Kandula, N. and Kumar, R. (2023) 'A Deep Dive into Academic Excellence: Using Deep Learning to Evaluate and Improve Engineering Students' Performance'.

Kareem, M.I. and Jasim, M.N. (2022a) ‘Fast and accurate classifying model for denial-of-service attacks by using machine learning’, Bulletin of Electrical Engineering and Informatics, 11(3), pp. 1742–1751.

Kareem, M.I. and Jasim, M.N. (2022b) ‘Machine learning-based DDoS attack detection in software-defined networking’, in International Conference on New Trends in Information and Communications Technology Applications. Springer, pp. 264–281.

Kriegeskorte, N. and Golan, T. (2019) ‘Neural network models and deep learning’, Current Biology, 29(7), pp. R231–R236.

Kuzilek, J., Hlosta, M. and Zdrahal, Z. (2017) ‘Open university learning analytics dataset’, Scientific data, 4(1), pp. 1–8.

Liu, Y. et al. (2023) ‘A novel student achievement prediction method based on deep learning and attention mechanism’, IEEE Access [Preprint].

M.Steinbach, P.Tan and V.Kumar (2006) “Introduction to data mining”. Pearson Education, Inc.

Mansour, H.S. et al. (2024) ‘A NOVEL DEEP 2D-CNN MODEL FOR ECG-BASED ARRHYTHMIA DIAGNOSIS WITH SELECTIVE ATTENTION MECHANISM AND CWT INTEGRATION’ Kufa Journal of Engineering, 16(2), pp. 423 -444.

Moubayed, A. et al. (2023) ‘A Deep Learning Approach Towards Student Performance Prediction in Online Courses: Challenges Based on a Global Perspective’, 2023 24th International Arab Conference on Information Technology, ACIT 2023 [Preprint], (Dl). Available at: https://doi.org/10.1109/ACIT58888.2023.10453917.

Nawang, H., Makhtar, M. and Hamzah, W. (2021) ‘A systematic literature review on student performance predictions’, International Journal of Advanced Technology and Engineering Exploration, 8(84), pp. 1441–1453.

Ouahi, M., Khoulji, S. and Kerkeb, M.L. (2024) ‘Predictive assessment of learners through initial interactions with encoding techniques in deep learning’, Journal of Autonomous Intelligence, 7(4), pp. 1–17. Available at: https://doi.org/10.32629/jai.v7i4.1443.

Poirion, O.B. et al. (2021) ‘DeepProg: an ensemble of deep-learning and machine-learning models for prognosis prediction using multi-omics data’, Genome medicine, 13(1), pp. 1–15.

Probst, P., Wright, M.N. and Boulesteix, A.L. (2019) ‘Hyperparameters and tuning strategies for random forest’, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(3). Available at: https://doi.org/10.1002/widm.1301.

Q. Li et al (2022) ‘A Survey on Text Classification: From Traditional to Deep Learning’.

Qasrawi, R. et al. (2021) ‘Predicting school children academic performance using machine learning techniques’, Adv Sci Technol Eng Syst J, 6, pp. 8–15.

Rani, P. et al. (2023) ‘PCA-DNN: A Novel Deep Neural Network Oriented System for Breast Cancer Classification’, EAI Endorsed Transactions on Pervasive Health and Technology, 9.

Saranya, G. and Pravin, A. (2023) ‘Grid Search based Optimum Feature Selection by Tuning hyperparameters for Heart Disease Diagnosis in Machine learning Abstract ’:, pp. 1–13. Available at: https://doi.org/10.2174/18741207-v17-e230510-2022-HT28-4371-8.

Siahkali, M.A. (2022) ‘Pirates Swarm Search Algorithm : A Meta-Heuristic Search Algorithm for optimization problems’.

Tarek, Z. et al. (2023) ‘Soil Erosion Status Prediction Using a Novel Random Forest Model Optimized by Random Search Method’, Sustainability (Switzerland), 15(9). Available at: https://doi.org/10.3390/su15097114.

Waheed, H. et al. (2023) ‘Early prediction of learners at risk in self-paced education: A neural network approach’, Expert Systems with Applications, 213, p. 118868.

Wotaifi, T.A. and Dhannoon, B.N. (2022) ‘Improving Prediction of Arabic Fake News Using Fuzzy Logic and Modified Random Forest Model’, Karbala International Journal of Modern Science, 8(3), pp. 477–485. Available at: https://doi.org/10.33640/2405-609X.3241.

Wotaifi, T.A. and Dhannoon, B.N. (2023a) ‘An Effective Hybrid Deep Neural Network for Arabic Fake News Detection’, Baghdad Science Journal, 20(4), pp. 1392–1401. Available at: https://doi.org/10.21123/bsj.2023.7427.

Wotaifi, T.A. and Dhannoon, B.N. (2023b) ‘Attention Mechanism Based on a Pre-trained Model for Improving Arabic Fake News Predictions’, Iraqi Journal of Science, 64(11), pp. 6041–6054. Available at: https://doi.org/10.24996/ijs.2023.64.11.45.

Wu, O. et al. (2019) ‘Big data approaches to phenotyping acute ischemic stroke using automated lesion segmentation of multi-center magnetic resonance imaging data’, Stroke, 50(7), pp. 1734–1741.