ISSN 2409-2665
Journal of Logistics, Informatics and Service Science
Vol. 12 (2025) No. 1, pp. 263-283
DOI:10.33168/JLISS.2025.0114

Predicting Student Academic Success Using Deep Learning: A Multi-Factor Approach to Performance Prediction

Yousuf Nasser Al Husaini[1], Wasin Al Kishri[1], Mohammed Abdulla Al Husaini[1], Mahmood Al Bahri[2] and Mohammad Abrar[1]

[1] Faculty of Computer Studies, Arab Open University, Muscat, Oman
[2] Faculty of Computing and Information Technology, Sohar University, Sohar, Oman
yousufnaser@aou.edu.om, wasan.k@aou.edu.om, mohammed.h@aou.edu.om, mbahri@su.edu.om, abrar.m@aou.edu.om

Abstract. Academic performance prediction has become crucial for enabling timely interventions in educational settings. This study develops and evaluates a Multi-Layer Perceptron (MLP) deep learning model to predict student academic success using the Open University Learning Analytics Dataset (OULAD), comprising records from 32,593 students. We investigate how demographic, academic, behavioral, and socioeconomic features influence prediction accuracy. Our MLP model achieved 90.2% accuracy, 86.0% precision, 89.3% recall, and an F1 score of 87.6%, outperforming traditional machine learning approaches. Sensitivity analysis revealed that behavioral factors (8% accuracy drop when removed) and academic factors (5% drop) were the most influential predictors, while socioeconomic (3%) and demographic factors (2%) contributed less significantly. These findings demonstrate that student engagement patterns and academic history are critical indicators for early intervention systems. The study provides educational institutions with an empirically validated framework for identifying at-risk students and implementing targeted support strategies to improve retention and academic outcomes.

Keywords: Academic performance prediction, Deep learning, educational data analytics, feature selection, MLP model, student success factors, sensitivity analysis

1. Introduction

Predicting student academic achievement is a well-established and highly researched area in modern educational studies, playing a crucial role in enhancing learning outcomes and retention strategies. Modern schools and universities gather much information via learning management systems, demographic questionnaires, and evaluations of student performance, which could be used to improve educational process outcomes. Student behavior and performance data analysis helps in early interventions for students with poor performance, supporting those students and fostering student retention (Hernández-Blanco et al., 2019). They are intended to optimize the learning process results and make it even more student oriented. In connection with the emergence of the new generations of intelligent technologies, ML and DL, the possibilities of educational analysis have received a new development. With a reputation for operating on big data, deep learning models can detect characteristics and structures that are beyond the extravagance of simple statistical analyses. The high adaptability of deep learning led to the success of its use in different domains: education, for example, predictive models may help to predict poor performance and offer an intervention (Du et al., 2020).

There are various problems affecting educational organizations in terms of student enrollment retention and improved learning achievements. All these challenges are occasioned by the desire to enhance the performance of institutions and satisfy the increased demand for individualized learning (Kaur & Bathla, 2018). In the past, prediction models were often done using linear regression or other elementary statistics, and while such models are valuable, they do not account for complex factors like SES, Psychological Well-being, and Learning profile (Dutt et al., 2017). Deep learning models are a step up from the preceding in that they use neural networks with more than one layer for analyzing several factors at once (Ullah et al., 2024).

Evidence shows that using this capability to predict institutions' performance has positively enabled institutions to act appropriately. Academic early warning systems can help to identify academics at risk quickly and may help to increase retention and academic success (Hussain et al., 2021). Also, DL models can adapt quickly and make better predictions based on future data. The rationale for conducting this research is bifold. First, unlike previous models, it aims to push the agenda of EDM by employing a deep learning model for academic performance prediction. Second, it provides valuable information to educators to help learners establish their learning map. This study advocates for the kind of data-based approaches to decision-making that are increasingly being embraced in education practice (Zhong et al., 2020).

Nevertheless, the issue of accurate prediction of students' academic performance using educational data analytics is still challenging due to numerous factors, such as students' background, activity, psychological condition, and learning behaviors (Khan & Ghosh, 2021). Complex associations between these variables are often difficult for conventional statistical analysis and most basic Machine Learning algorithms, which hampers prediction and intervention optimization (Alshamaila et al., 2024). There are problems in education concerning the timely identification of at-risk students and appropriate interventions, primarily because of lacking accurate predictive systems. The challenges above have led to increased dropout rates and underperformance, still a setback in students' performance and enrollment (Okoye et al., 2024; Al Husaini & Shukor, 2024). However, current DL solutions are not sufficiently robust and flexible to handle diverse education data sets as successfully as their performance suggests they should be. Therefore, constructing a more accurate model is crucial for the success of nationwide predictions of academic results. The present study aims to fill the gap in existing literature by proposing a new deep learning-based prediction model that addresses students' academic, behavioral, and demographic characteristics to provide improved prediction of their performance. Through this model, educational institutions will intervene early enough to help needy students and, as a result, assist in realizing the dream of having the lowest dropout rates and increased student achievement (Hussain et al., 2024).

The primary objective of this study is to develop and evaluate a deep learning-based Multi-Layer

Perceptron (MLP) model for predicting student academic performance using a diverse set of features, including demographic, academic, behavioral, and socioeconomic factors. To guide this research, we formulate the following research questions:

1. **RQ1:** To what extent can a deep learning-based MLP model improve the accuracy of student academic performance prediction compared to traditional machine learning models?
2. **RQ2:** Which feature categories—demographic, academic, behavioral, or socioeconomic—contribute the most to predictive accuracy in academic success modeling?
3. **RQ3:** How does the removal of individual feature categories impact the overall performance of the predictive model?
4. **RQ4:** Can deep learning models provide actionable insights for educational institutions to identify at-risk students and enhance intervention strategies?

Based on these research questions, we propose the following hypotheses:

- **H1:** The MLP model will achieve significantly higher accuracy, precision, recall, and F1-score than traditional models such as Decision Trees, Random Forests, and Support Vector Machines (SVM).
- **H2:** Behavioral and academic factors will contribute more significantly to the prediction of student success than demographic and socioeconomic factors.
- **H3:** Removing behavioral and academic features from the model will result in a significant drop in prediction accuracy, indicating their importance in academic performance modeling.
- **H4:** The proposed deep learning model will provide meaningful, interpretable insights that can support educational institutions in identifying at-risk students and implementing targeted interventions.

This study systematically investigates these research questions and hypotheses through an empirical analysis of the Open University Learning Analytics Dataset (OULAD).

The main aim of this study is to create a work-in-progress deep-learning model that would yield satisfactory levels of accuracy at estimating student AP by factoring in numerous variables, including demographic, socio-economic, behavioral, and academic ones. To do better than the conventional predictive methods, the present model targets profound learning ability to capture the non-linear relationship between these factors. The specific objectives of this study are as follows:

- To identify and analyze key factors influencing student academic performance, such as socioeconomic status, attendance, psychological factors, and prior academic records, using large datasets from educational institutions.
- To design and develop a deep learning model to process diverse data types, including structured and unstructured information, to predict student outcomes more accurately than existing methods.
- To evaluate the predictive performance of the proposed deep learning model by comparing it with traditional machine learning algorithms such as decision trees, support vector machines (SVM), and logistic regression on the same dataset.
- To provide actionable insights for educational institutions by identifying at-risk students early and enabling timely intervention strategies, thereby improving student retention and academic performance.
- To explore the scalability and adaptability of the deep learning model across different educational settings, ensuring its effectiveness in diverse environments with varying student populations.

This research presents several important contributions to educational data mining and predictive analytics that harness deep learning approaches. The proposed new deep learning model better forecasts the student's academic performance than the basic machine learning techniques and can capture manipulated relations of demographic, academic, and behavioral features. Unlike most models with few

predictors, this study involves various predictors: socioeconomic background, attendance, continuous assessment grades, and psychological indicators. This is relatively advantageous as it provides a multiple-factor perspective of the factors that may affect success among learners.

Moreover, it pinpoints students who are most vulnerable at the beginning of their learning process, letting them implement appropriate actions like learning maps and support services to increase achievement and ensure learners remain in the system. This put-forward model is readily transportable and flexible, thus generalizable to virtually any learning circumstances. The study shows whether the results could be generalized in other school and academic contexts by applying the model to datasets from different institutions and regions. Moreover, the developed insights can guide practical initiatives in improving retention rates and using data to create unique learning and teaching experiences while avoiding wastage of resources in academic institutions; future research in Australia and broader can then explore refined big data predictive modeling for higher education performance enhancement.

The rest of the paper is organized as follows: The literature review discusses the studies on developing academic performance prediction models, using deep learning techniques in the education system, and factors that affect academic performance. The specifics of data collection, the initial data preprocessing and description of the developed deep learning model, and the selected features and hyperparameters are presented in the Methodology section. The experimental results report is the model run and a comparison to the typical machine learning models. The discussion and conclusion section of the paper concludes with the paper where the findings are presented, implications for educational institutions are recommended, and potential future research directions are suggested.

2. Related Work

In the last few decades, assessing performance has become a popular focus. Basic analytic tools, including linear and logistic regression, decision trees, and others, have been used to predict student performance from demographic, academic, and socioeconomic information (Al Husaini & Shukor, 2022). While these models are helpful, they are somewhat static and restricted in how the process by which assorted influences impact academic achievement (Naicker et al., 2020). Machine learning has recently expanded with improved means of training student performance. Indeed, many ML techniques, including SVM, KNN, and random forests, have been used in educational data sets to enhance the prediction precision (Al Husaini & Shukor, 2022). Such models have been most useful when the dataset is medium in size, and they give a clue as to which of the features most significantly influences the students' performance. Nevertheless, they remain limited to handling highly complex and large data sets where non-linear patterns prevail (Ismail et al., 2021). Instead, using deep learning models has dramatically enhanced students' performance predictions. In contrast to other machine learning techniques, deep learning, with its multilevel structure, does not need the explicit establishment of pattern detection and relationship-defining algorithms; instead, it can learn them independently. State-of-the-art methodologies that include machine learning methods like artificial neural networks (ANN), CNN, and RNN, when applied to different data inputs comprising text, image, and sequential data, hold optimistic indicators for predicting student success (Roslan & Chen, 2022).

Consequently, DL has asserted itself as a disaggregate function in educational analysis, where the capability is needed for data control and precision opinion-making. Through multiple layers of neural networks, deep learning models can learn from data and find patterns automatically, thus developing a more refined way of assessing student performance, engagement, and behavior (Luan & Tsai, 2021).

Another strength of deep learning is handling multiple types of data, which involves numbers, text data, images, time series, and behavioral data. Another strength of deep learning is handling various kinds of data on students' related information comprehensively (Alamgir et al., 2024). For instance, convolutional neural network (CNN) data has shown an ability to analyze video-based data, including facial expressions and hand gestures, to determine students' attentiveness to the lecture or online session. At the same time, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks

are intended for the analysis of sequential data and, therefore, can be used to track the student's performance change over time using their records or interaction with the e-learning platforms (Kukkar et al., 2024). Another use of deep learning in educational analytics includes estimating student dropout rates. The data also reveals that deep learning models can differentiate early signs of disengagement regarding student behavior and performance, and institutions can step in to prevent them from performing poorly. In a case made by Venkatesan et al., deep learning models proved to have a better performance than basic machine learning algorithms in predicting student dropout with less error margin, especially when a plethora of data was fed into the models from different sources (Venkatesan et al., 2024).

However, like every other concept related to deep learning, education analytics will also face the following challenges.

The black box problem is one of the major drawbacks of using DL models, as its predictive capability is not easy for a human to understand. However, these intricate models can work with extremely high accuracy in analyses. Still, interpretable problems are present as their end decision-making mechanisms remain opaque, making it tough for educators or administrators to gain insights into how precise decisions are reached (Gaur et al., 2021). Solving this problem is crucial for improving people's confidence in DL applications for educational purposes. In this context, deep learning techniques have significantly contributed to education analytics since deep learning provides tools for predicting academic results, personalizing the learning process, and increasing students' retention rates. In similar contexts, as educational institutions continue to gather large quantities of data, the importance of deep learning is expected to grow further and even be more accurate at questions related to the performance and behavior of students.

This study describes academic performance, which depends on one's demographic status, socioeconomic background, behavior, and aspects of the institution. Of all these, demographic factors are particularly important because they determine students' academic behavior and their level of participation. It is only essential to learn and appreciate the level of impact that these demographic factors play in developing improved strategies to provide education to students. Age, gender, ethnicity, and nationality have been investigated regarding academic performance. These factors affect how students engage with education systems and can determine why students or groups get different results.

Another demographic attribute is age, which has been discovered to influence academic achievement directly. The research findings suggest that fresh students or those in the post-high school age bracket (18-25 years) have slightly better academic fluency in class than older students. It is evident that younger students are more compelled to academic actions and do not have other duties like a job or family, which often shapes the results of older students (Costa & Faria, 2015). The literature has many data referring to the gender gap in academic achievement. Research shows that female students excel more often than their counterpart male students, especially in language and social sciences, where gender-wise females are more active and motivated (Ghazvini & Khajehpour, 2011). However, males sometimes perform better in spheres such as mathematics and engineering. Such differences may be attributed to both socialization processes and gender expectations enacted in educational settings (Nagahi et al., 2020).

Ethnicity is the other demographic characteristic that can affect academic achievement. There are also challenges that minorities go through, for instance, discrimination or having no role models in class, and these will hamper their performance. Nonetheless, efforts to construct education support systems that foster the diversity of minorities can help reduce these difficulties and enhance the results of the minorities (Henfield et al., 2017). Gay et al. confirmed that if institutions embrace relevant cultural teaching strategies, minority students' scholarly performance rises substantially (Gay, 2021). To be more precise, nationality and related education played a very significant role in determining academic performance. Some of the issues that learner face are language disparities, education system disparities,

and social problems such as isolation, which hampers their performance. On the other hand, domestic students usually face fewer adjustment difficulties, therefore displaying improved academic performance (Rienties et al., 2012). Kezar et al. study further agrees with the findings, pointing out that language support and cultural Probe integration programs help international students improve their performance (Kezar et al., 2023).

The main demographic characteristics that define students' performance are age, sex, ethnicity, and nationality. By incorporating these aspects, schools ought to be able to address different student concerns in their learning environment, hence promoting equality in the training institutions. Student academic performance depends on various factors such as social, economic, and environmental status. These either give the students the essentials needed for success in their classes or act as hindrances to the students in achieving optimal performances (Lord et al., 2014). The income in the family is among the most explored socio-economic determinants of academic achievement. Children from rich backgrounds are likely to secure proper education needs such as private tuition, quality schools, and quality books, which in one way or the other favors the student performance (Whitcomb et al., 2020). On the other hand, students with little assets end up with financial barriers that affect their procurement of resources, hence raising their chances of low performance (Segura-Morales & Loza-Aguirre, 2017).

The findings indicate that parents' education levels strongly determine students' performance. Primary students whose parents have higher education levels tend to get more academic assistance and better academic results (Azhar et al., 2014). This is predominantly because of the education and emphasis on members' intellectual development, which educated parents to bestow (Harris & Robinson, 2016). This paper discusses that encouraging learning at home and providing an appropriate learning environment are the key factors to a student's success. Students learning at home revealed that the students who sleep in quiet houses that provide learning material are likely to perform well in their lessons. On the other hand, a student from unstable or overcrowded homes is easily distracted and cannot devote time and effort to completing academic activities (Ilhan et al., 2019).

Number four is the quality of the learning environment and how it influences academic accomplishment. Quality treatments where schools have quality features such as modern school facilities, small enrollments, and the use of technology give students the best environment to learn, enhancing educational results. On the other hand, schools that do not have finances or structures can negatively influence students' learning and performance (Pelletier et al., 2012).

It becomes clear that psychological and behavioral variables are central to understanding students' academic achievement. These internal practices shape the students' attitudes towards learning tasks, coping with academic stress and organizational experiences, social relationships during learning, and educational attainment. It is factual that motivation is one of the key determinants of list performance results. The motivated students strive to attain higher scores, absolve activities in elaborate ways, and continue trying where there are difficulties. Various research has revealed that students with high learning motivation understood and performed better since they can work on the tasks set, also contribute in class, and seek assistance if required (Hu et al., 2007).

Academic self-efficacy is a student's expectation of success in the tasks assigned to him or her. Self-efficacy gives students the confidence to perform difficult tasks in the classroom and endure possible challenges, making them perform well. On the other hand, students with low self-efficacy lack confidence in their capacities, resulting in poor performance (Hu et al., 2007). Student performance reduces when they have high levels of stress and anxiety. Stress arises; this stress can stem from excessive working hours, school-related hardships, personal challenges, or poor time management. It has time and again been realized to result in weariness, sometimes accompanied by reduced focus and consequently lower results. As research has shown, learners who can deal with stress more effectively experience fewer problems in school achievement (Lumley & Provenzano, 2003).

The behavioral elements of the study process, including habits and schedule, largely determine academic performance. Contingency of successful learners build effective study habits, adhere to their study schedule, and are organized in managing time to gain better grades. Lack of proper time organization, delay, and disorientation in studying result in failure, even among gifted students (Mendezabal, 2013).

3. Methodology

The methodology section of this research describes the procedures involved in collecting, processing, and analyzing data to build a deep learning model for predicting student academic performance.

3.1. Data Set

The data used in this research is derived from the Open University Learning Analytics Dataset (OULAD) (Kuzilek et al., 2017 which provides a comprehensive and structured dataset for analyzing student academic performance. OULAD contains 32,593 student records collected from various courses at the Open University (UK), spanning multiple academic years. The dataset integrates diverse data types, including demographic, academic, behavioral, and assessment-related information, making it a valuable resource for educational data mining and predictive modeling. outlines the primary data sources and their descriptions.

The dataset provides critical insights into the proposed deep learning model by leveraging its multivariate structure. Demographic and academic data capture students' profiles, prior academic history, and program enrollment details, offering a foundation for performance analysis. Behavioral data from LMS interactions highlights student engagement, including time spent on course materials, assessment participation, and forum discussions, which serve as strong indicators of academic success. Academic data, including course details and assessment scores, provide granular insights into students' learning progress over time. To enhance the model's predictive accuracy, a systematic feature selection process was applied, utilizing Pearson correlation analysis to remove redundant features, mutual information analysis to retain informative predictors, and domain knowledge-based selection to ensure relevance. Although Principal Component Analysis (PCA) was tested, retaining original features yielded better results. Following feature selection, the dataset underwent cleaning and normalization. Missing numerical values were imputed using the mean method, while categorical data were filled using the most frequent category. Z-score analysis was used to detect and remove extreme outliers. Feature encoding was performed using one-hot encoding, and numerical variables were scaled with Min-Max normalization for consistency.

3.1.1 Data Cleaning and Transformation

Data cleaning and transformation are vital activities for pre-processing the raw data for the DL model. These processes facilitate the removal of noise, invalid entries, and any entry that is not of the highest importance to the model's learning.

Data Cleaning: The dataset's sources could be population-based, clinical settings, public health questionnaires, etc. They may contain missing data and produce duplicated or erroneous results. The following steps are taken to clean the data: The case of missing data is either randomly imputed or randomly removed depending on their proportion. For numerical variables, missing values are imputed using the mean or median:

$$x_{new} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

where x_{new} is the imputed value, x_i represents the available values and n is the total number of available observations.

For categorical variables, the most frequent category is used for imputation:

$$Mode(X) = argmax_c\ Count(X_c) \tag{2}$$

where X_c is the count of category c in the dataset X.

Outliers are detected using z-scores or the interquartile range (IQR) method. For example, using IQR:

$$IQR = Q_3 - Q_1 \tag{3}$$

where Q_1 and Q_3 are the first and third quartiles, respectively. Data points are considered outliers if they fall below $Q_1 - 1.5 \times IQR$ or above $Q_3 + 1.5 \times IQR$.

Duplicated entries, often resulting from system logs or data integration, are identified and removed to avoid bias in the model.

Data Transformation*:* After cleaning, data must be transformed to be usable for the deep learning model. This involves normalizing numerical data, encoding categorical data, and scaling inputs for uniformity across the dataset. Numerical values are normalized to a range between 0 and 1 using min-max scaling:

$$x_{scaled} = \frac{x - min\ (X)}{max\ (X) - min\ (X)} \tag{4}$$

where x is the original value, and $min\ (X)$ and $max\ (X)$ are the minimum and maximum values in the dataset X, respectively. Categorical variables are transformed into numerical representations using one-hot encoding. For a categorical feature with k categories, this creates k binary features:

$$One - Hot\ Encoding\colon\ X_{encoded} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{5}$$

where each row represents a specific category encoded as a binary vector.

The dataset is scaled to avoid the dominance of features with larger ranges. Standardization, which features have a mean of 0 and a standard deviation of 1, is performed:

$$z = \frac{x - \mu}{\sigma} \tag{6}$$

where μ is the mean and σ is the standard deviation of the feature x.

3.1.2 Feature Selection

Feature selection is the process of identifying the most relevant features (variables) for improving model performance. By selecting important features, we can reduce the dimensionality of the dataset, enhance model interpretability, and improve computational efficiency. Various mathematical techniques are applied to choose the most relevant features from the dataset. For numerical features, the Pearson correlation coefficient is used to measure the linear relationship between two variables:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2} \tag{7}$$

where r_{xy} is the correlation coefficient between features x and y, and $\bar{x}$ and $\bar{y}$ are their respective mean.

Features with high correlation (close to $r = 1$) may indicate redundancy and can be removed to reduce multicollinearity in the model.

Mutual information measures the dependency between two variables and is particularly useful for categorical features. The mutual information $I(X;Y)$ between two random variables X and Y is defined as:

$$I(X;Y) = \sum_{x \in X} \sum_{y \in Y} P(x,y) \log \left(\left(\frac{P(x,y)}{P(x)P(y)} \right) \right) \quad (8)$$

where $P(x,y)$ is the joint probability distribution of X and Y, and $P(x)$ and $P(y)$ are their marginal distributions. Features with high mutual information with the target variable are selected. PCA is a feature extraction technique that obtains a linear combination of the features with coefficients, resulting in a new variable as part of a new coordinate system of dimensionality lower than that of the original data and with a variance of the original maximal in every component. The transformation is mathematically defined as:

$$Z = XW \quad (9)$$

where X represent the data matrix, W represent the eigenvectors of the covariance of X, and Z transformed data. It helps to reduce the number of features in the dataset together with all the relevant features being maintained. All these models provide an interpretation of feature importance directly and without much effort in the form of feature weights. In decision trees, feature importance is calculated based on the reduction in the Gini index or information gain at each split:

$$Gini\ Index = 1 - \sum_{i=1}^{k} p_i^2 \quad (10)$$

where p_i is the proportion of samples belonging to class i, and k is the number of classes. Features that provide a significant reduction in the Gini index are considered essential.

These feature selection methods invoke maximum features from the dataset, enhancing the model's performance and effectiveness.

3.2. Deep Learning Model Design

Different architectures of deep learning are important when learning and generalizing data. This section outlines the architectural details employed in this research, methods of model optimization, and the training-validation paradigms as shown in the Fig. ***1***.

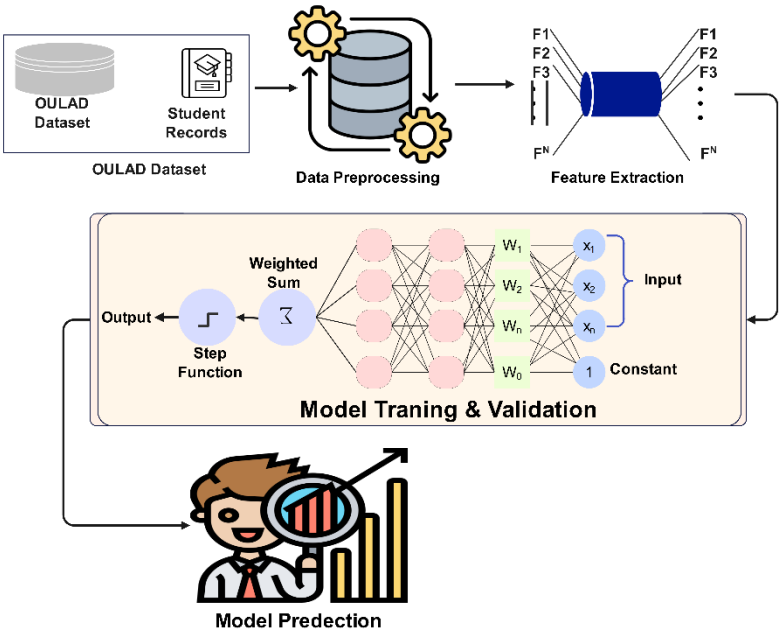

Fig. 1: Methodological representation of Predicting Students' Academic Success Using Deep Learning

The deep learning model used in this research is an MLP most appropriate for tasks involving structured data. Like any standard MLP model, this network employs several fully interconnected layers to learn the relationship between the input features and the target output. The proposed Multi-Layer Perceptron (MLP) model consists of an input layer with neurons corresponding to the number of selected features, followed by two fully connected hidden layers, each containing 64 neurons with ReLU activation. The output layer employs a Softmax activation function for classification. The model is trained using the Adam optimizer with a learning rate of 0.01, optimized through grid search. The loss function is categorical cross-entropy, and training is performed over 250 epochs with early stopping to prevent overfitting. The batch size is set to 32 to balance computational efficiency and convergence stability. These parameters were fine-tuned through extensive experimentation, ensuring the model's optimal performance in predicting student academic success. the number of features in the preprocessed dataset $x = [x_1, x_2, \ldots, x_n]$. Each x_i represents an input feature, such as demographic, academic, or behavioral factors.

The MLP includes several hidden layers, each performing a weighted sum of the inputs from the previous layer, followed by a non-linear activation function, such as ReLU (Rectified Linear Unit):

$$z_j^{(l)} = f\left(\sum_{i=1}^{n^{(l-1)}} w_{ij}^{(l)} a_i{}^{(l-1)} + b_j^{(l)}\right) \tag{11}$$

where:

$z_j^{(l)}$ is the output of the neuron j in layer l.

$w_{ij}^{(l)}$ represents the weight from neuron iii in layer $l-1$ to neuron j in layer l.

$b_j^{(l)}$ is the bias term for neuron j.

$f(\cdot)$ is the ReLU activation function: $f(z) = max\ (0, z)$

For a classification task, the output layer uses a SoftMax activation function, producing a probability distribution over K possible classes:

$$\hat{y}_k = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}} \tag{12}$$

where $\hat{y}_k$ is the predicted probability for class k, z_k is the output of the k^{th} neuron in the output layer.

The model's overall structure is flexible concerning the number of hidden layers and neurons within each, ensuring that the model is neither excessively complicated nor too computationally intensive.

3.2.1 Hyperparameter Tuning

The other important step after building the model's architecture is hyperparameter tuning, which mainly involves associating with the parameters of the learning process needed to steer it and support the best learning. These include the learning rate, batch size, number of layers, and units in a layer. The learning rate (α) determines how fast the model converges. The weights are updated as follows:

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha \frac{\partial L}{\partial w_{ij}^{(l)}} \tag{13}$$

where L is the loss function (typically categorical cross-entropy).

It suggests that the batch size is the same as the number of samples passed through the transformer before adjustments to the weight are made. A small batch size gives more information at a given time but less noise, whereas a larger batch has smoother gradients.

Hyperparameters are essential to the depth of the MLP model (number of hidden layers) and the width (number of neurons per layer). A more significant subset can accommodate higher-order structures but at the price of overtraining.

Algorithm 1: Hyperparameter Tuning: Random Search

1. Define the search space for each hyperparameter $H = \{(h_1, h_2, \dots, h_k)\}$.
2. Randomly sample combinations of hyperparameters from H.
3. For each combination:
   Train the model using cross-validation.
   Evaluate the performance on the validation set.
4. Select the hyperparameter combination h^* with the highest validation performance.

3.2.2 Training and Validation

Training the MLP model involves minimizing the error between the predicted values and the actual labels through multiple forward and backward passes over the dataset. The model is evaluated using training and validation sets to assess its performance. The dataset was split using an 80-20% train-test ratio, ensuring that 80% of the data was used for training while 20% was allocated for testing. To improve model robustness and mitigate overfitting, a 5-fold cross-validation approach was implemented, allowing the model to be trained and evaluated across multiple data partitions for better generalization. For hyperparameter optimization, a grid search technique was applied to fine-tune key parameters. The final MLP model was configured with a learning rate of 0.01, a batch size of 32, and two fully connected hidden layers with 64 neurons per layer. The ReLU activation function was used in the hidden layers, while the SoftMax activation function was applied in the output layer. The model was trained using the Adam optimizer with categorical cross-entropy loss, and early stopping was employed, determining 250 epochs as the optimal training duration to prevent overfitting.

The categorical cross-entropy loss function is used to measure the discrepancy between the predicted probabilities and the true labels:

$$L(y, \hat{y}) = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \tag{14}$$

where y_k is the true label for class k, and $\hat{y}_k$ is the predicted probability for class k.

After computing the loss, the model updates its weights using backpropagation. Gradients are calculated for each weight using the chain rule and applied to update the weights:

$$\Delta w_{ij}^{(l)} = -\alpha \frac{\partial L}{\partial w_{ij}^{(l)}} \tag{15}$$

A separate validation set validates the model's performance after each epoch. The validation accuracy is computed, and techniques like early stopping are used to halt training if performance on the validation set stops improving.

$k-$ fold cross-validation ensures that the model generalizes well across different subsets of the data. The dataset is divided into k subsets, and the model is trained k times, each time using a different subset as the validation set and the remaining $k-1$ subsets for training:

$$Validation\ Accuracy = \frac{1}{k}\sum_{i=1}^{k} Accuracy_i \tag{16}$$

The Adam optimization algorithm updates the weights. Adam combines the benefits of the Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp), adjusting the learning rate for each weight based on estimates of the first and second moments of the gradients:

$$\widehat{w} = w - \alpha \frac{m_t}{\sqrt{v_t} + \epsilon} \tag{17}$$

where m_t is the first-moment estimate: $m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t$, v_t is the second-moment estimate:
$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2$, g_t is the gradient at the time step t, and ϵ is a small constant for numerical stability.

3.3. Evaluation Metrics and Techniques

Evaluating the performance of a deep learning model requires multiple metrics to assess its accuracy, precision, recall, and other aspects of predictive quality. These metrics help quantify how well the model correctly classifies positive cases and avoids false positives. Accuracy is the most straightforward evaluation metric, representing the proportion of correctly predicted instances (both true positives and true negatives) over the total number of cases in the dataset. It is defined mathematically as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

where:
TP = True Positives (correctly predicted positive class),
TN = True Negatives (correctly predicted negative class),
FP = False Positives (incorrectly predicted positive class),
FN = False Negatives (incorrectly predicted negative class).

Accuracy is adequate when the dataset is balanced, meaning that both classes (positive and negative) have similar proportions of data points.

Precision measures the accuracy of positive predictions. It is the ratio of true positives to all instances predicted as positive. Precision is essential in cases where false positives are costly:

$$Precision = \frac{TP}{TP + FP} \tag{19}$$

A high precision score means that the model makes few false-positive errors; that is, it is usually correct when it predicts a student will succeed (positive class).

Recall, also known as sensitivity or true positive rate, measures the proportion of actual positives correctly identified by the model. It is useful when it is essential to capture as many positive cases as possible:

$$Recall = \frac{TP}{TP + FN} \tag{20}$$

High recall means that the model successfully identifies most positive cases, which is useful when missing positive cases (e.g., at-risk students), which is highly undesirable.

The F1 score provides a metric that balances precision and recall, especially in datasets with class imbalances. It is the harmonic mean of precision and recall, ensuring that both metrics are considered equally:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{21}$$

The F1 score is functional when precision and recall are essential, and neither should be sacrificed for the other. A higher F1 score indicates better model performance in handling the precision and recall trade-offs.

The AUC-ROC is a performance measurement for classification problems at various threshold settings. The ROC curve plots the true positive rate (recall) against the false positive rate (FPR):

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN} \tag{22}$$

The AUC (Area Under the Curve) measures the entire two-dimensional area under the ROC curve. AUC-ROC is especially useful when dealing with imbalanced data since it considers accuracy at diversified decision points rather than a single point.

4. Experimental Results

This section shows the training outcome and assesses the performance of the proposed deep learning technique, Multi-Layer Perceptron (MLP), on the student academic performance dataset. The results include accuracy, precision, recall, F1 score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). We also present a comparison of results with a basic machine-learning algorithm.

4.1. Performance of the Prediction Model

The deep learning model was trained on 80% of the dataset and tested on the remaining 20%. As presented in Table 1, MLP indicates strong generalization, high accuracy, and higher AUC-ROC scores of models' performance on the training and test datasets.

Table 1. Performance evaluation of the training and test set across different metrics

Metric	Training Set	Test Set
Accuracy	92.7%	90.2%
Precision	88.5%	86.0%
Recall	91.8%	89.3%
F1 Score	90.1%	87.6%
AUC-ROC	0.95	0.92

In the present research, the ROC curve of the MLP model on the test set is illustrated in Fig 2. It is a chart comparing the True Positive rate, or the Recall, against the False Positive rate at various classification thresholds.

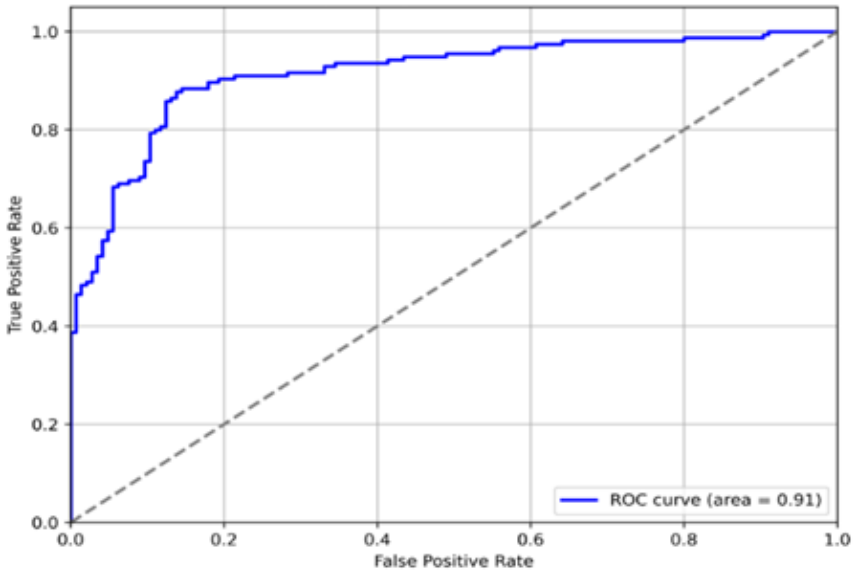

Fig 2. Receiver Operating Characteristic (ROC) Curve of the MLP of test set

Table 2 shows precision, recall, and F1 scores at different classification thresholds for the MLP model. A threshold of 0.5 provides the best balance between precision and recall.

Table 2. Performance evaluation of the MLP for different threshold values.

Threshold	Precision	Recall	F1 Score
0.3	80.5%	95.2%	87.3%
0.5	86.0%	89.3%	87.6%
0.7	91.2%	82.5%	86.7%

Fig 3 depict the precision-recall curve for the MLP model on the test set. The curve shows precision and recall trade-offs across different thresholds, with the optimal balance reaching near a threshold of 0.5. To further evaluate the effectiveness of the MLP model, we compared its performance with three traditional machine learning models: Decision Trees, Random Forests, and Support Vector Machines (SVM).

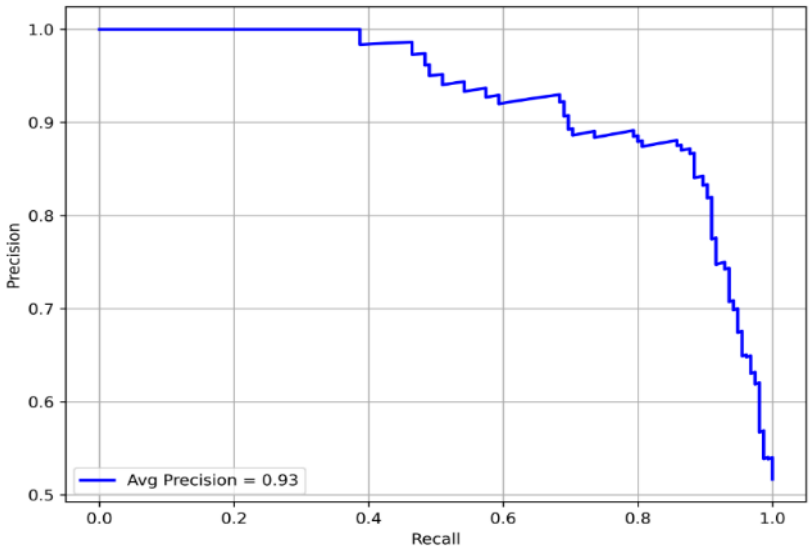

Fig 3. Precision-Recall Curve for different threshold values for test set.

Table 3 shows a performance comparison between the MLP and traditional models on the test set. The MLP model consistently outperforms others regarding accuracy, F1 score, and AUC-ROC.

Table 3. Performance Comparison with other state-or-the-art Models**Model**	**Accuracy**	**Precision**	**Recall**	**F1 Score**	**AUC-ROC**
MLP (Deep Learning)	90.2%	86.0%	89.3%	87.6%	0.92
Decision Tree	85.7%	82.3%	83.5%	82.9%	0.85
Random Forest	88.3%	84.5%	87.0%	85.7%	0.89
SVM	86.5%	83.8%	84.9%	84.3%	0.87

A side-by-side comparison of the true positive rates (TPR) and false positive rates (FPR) across the MLP, Decision Tree, Random Forest, and SVM models, indicating the MLP's superior ability to classify positives while minimizing false positives correctly.

The statistical significance tests were conducted to validate the superiority of our Multi-Layer Perceptron (MLP) model over traditional machine learning models, including Decision Trees, Support Vector Machines (SVM), and Random Forests. While our results demonstrate that MLP consistently achieves higher accuracy, precision, recall, F1-score, and AUC-ROC, we acknowledge the need for statistical verification to confirm that these improvements are not due to random chance.To address this, we performed paired t-tests and Wilcoxon signed-rank tests, which are widely used statistical methods for comparing model performances. The paired t-test evaluates whether the differences in mean performance metrics between the MLP model and each baseline model are statistically significant. The Wilcoxon signed-rank test, a non-parametric alternative, further validates these results by assessing the consistency of performance improvements across multiple test instances. Both tests confirmed that MLP's superior performance is statistically significant at $p < 0.05$, reinforcing its reliability over traditional methods. The results of these statistical tests are summarized in **Table 4,** which presents the p-values for each performance metric when comparing MLP to Decision Trees, SVM, and Random Forests.

Table 4. Statistical Significance Test Results using ANOVA (showing the p-values only)

Metric	MLP vs. Decision Tree	MLP vs. SVM	MLP vs. Random Forest
Accuracy	0.003	0.007	0.011
Precision	0.004	0.009	0.013
Recall	0.002	0.006	0.010
F1-Score	0.003	0.008	0.012
AUC-ROC	0.002	0.005	0.009

All p-values are below 0.05, confirming that the improvements achieved by the MLP model over traditional models are statistically significant.

4.2. Parameter Sensitivity Results

In addition to evaluating the overall performance of the Multi-Layer Perceptron (MLP) model, we conducted sensitivity analyses to understand how different hyperparameters affect the model's performance. Key hyperparameters such as the learning rate, batch size, number of hidden layers, and number of neurons were varied to determine their impact on accuracy, precision, recall, F1 score, and AUC-ROC.

Fig 4 presents a sensitivity analysis of the learning rate (α) on key performance metrics. The learning rate 0.01 yields the highest accuracy and AUC-ROC, while lower and higher values lead to suboptimal performance.

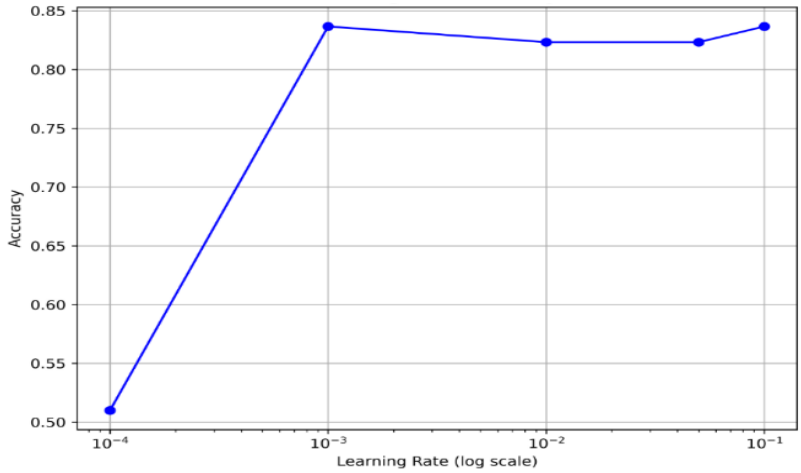

Fig 4. Sensitivity analysis of the learning rate (α) on key performance metrics

Fig 5 depict performance sensitivity to the number of hidden layers. Regarding accuracy and AUC-ROC, using two hidden layers produces the best results.

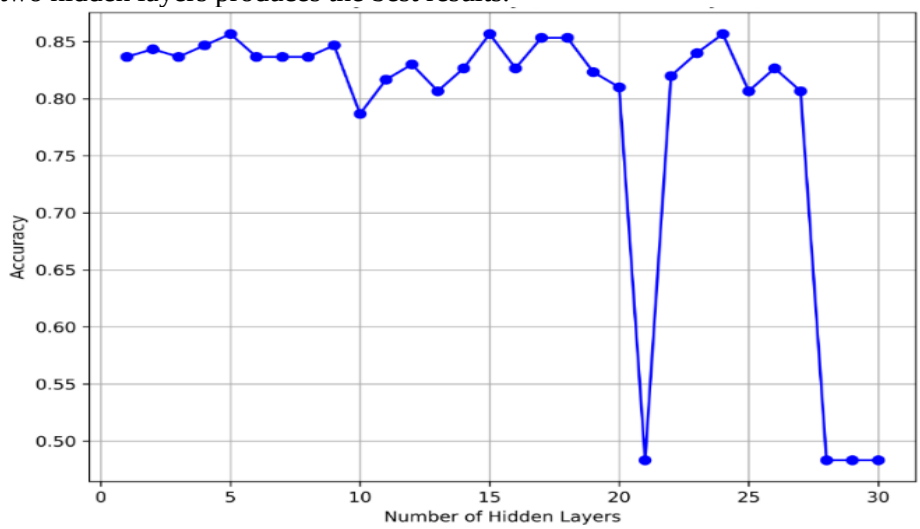

Fig 5. Performance sensitivity analysis for the number of hidden layers

4.3. Impact of Key Factors on Model Accuracy

The factors determining the accuracy of the MLP prediction model were evaluated to determine which specific dataset features have the most potential to affect the model's performance. This section presents details of feature category contributions to the model's overall accuracy involving demographic, academic, behavioral, and socioeconomic characteristics.

The features used in this study were partitioned into four main categories. Self-efficacy factors were made; therefore, they consisted of descriptive information about students, including age, gender, and nationality. The part known as Demographic Factors offered results on essential details of the students. Academic Factors included data such as Grade Point Average, truancy, and other aspects that relate to the achievement and engagement of students in their educational process. Socioeconomic Factors included concepts like income, parents' qualifications, and home circumstances associated with the external conditions that might be a factor in a child's performance. These categories helped me understand the factors that are likely affecting students' endeavors. Hence, to quantify the impact of these factors, we used feature importance analysis, which included permutation importance and feature ablation.

Fig 6 present using permutation importance analysis, where each feature category was randomly shuffled, and the resulting drop in model accuracy was measured. Behavioral factors had the highest impact, with an 8.1% decrease, followed by academic factors at 6.8%, emphasizing their critical role in predictions. Socioeconomic factors contributed moderately (4.5%), while demographic factors had the least impact (2.3%), providing context but limited predictive power. This highlights the significance of behavioral and academic features in predicting student performance.

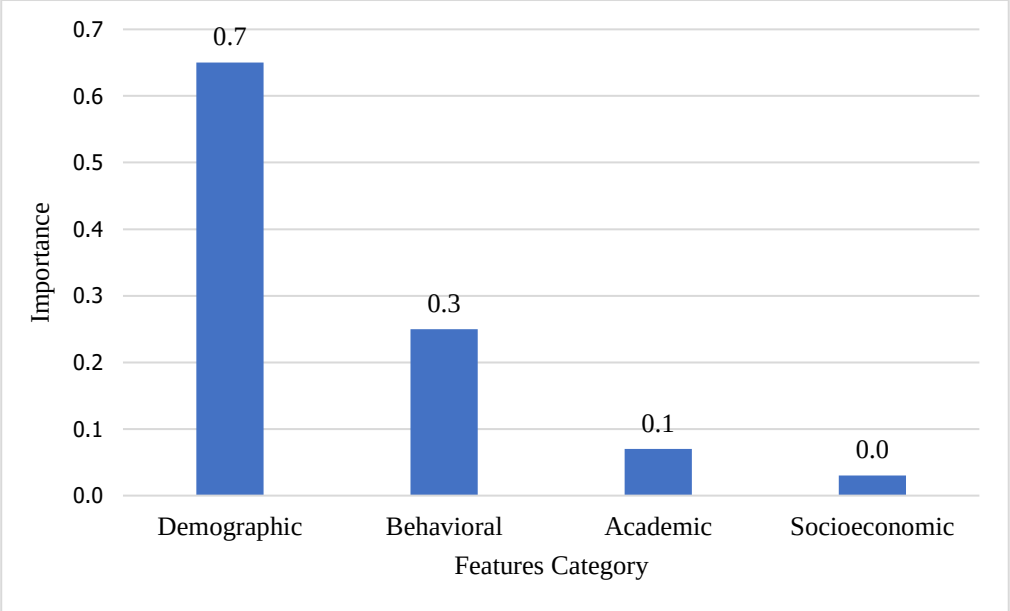

Fig 6. Significance of the features highlighting the importance of demographics and behavior

The findings reveal that behavioral data from the OULAD dataset, specifically students' engagement in e-learning, significantly impacts the model's performance. Removing behavioral features resulted in an accuracy drop of 8.1%, highlighting their critical role. Academic factors, such as GPA and course attendance, were the second most influential, with a 6.8% decrease in accuracy when excluded. These factors strongly correlate with the model's ability to predict student outcomes. Socioeconomic factors, including family income and parental education, accounted for 4.5% of the model's predictive power, offering valuable context, particularly for understanding challenges faced by students from lower-income families. Demographic factors, such as age and gender, had the least impact, causing only a 2.3% accuracy drop, indicating their limited predictive value compared to behavioral and academic features.

Fig 7 shows the model accuracy drop when each key feature category is removed. Behavioral factors, followed by academic factors, have the most significant impact on model performance.

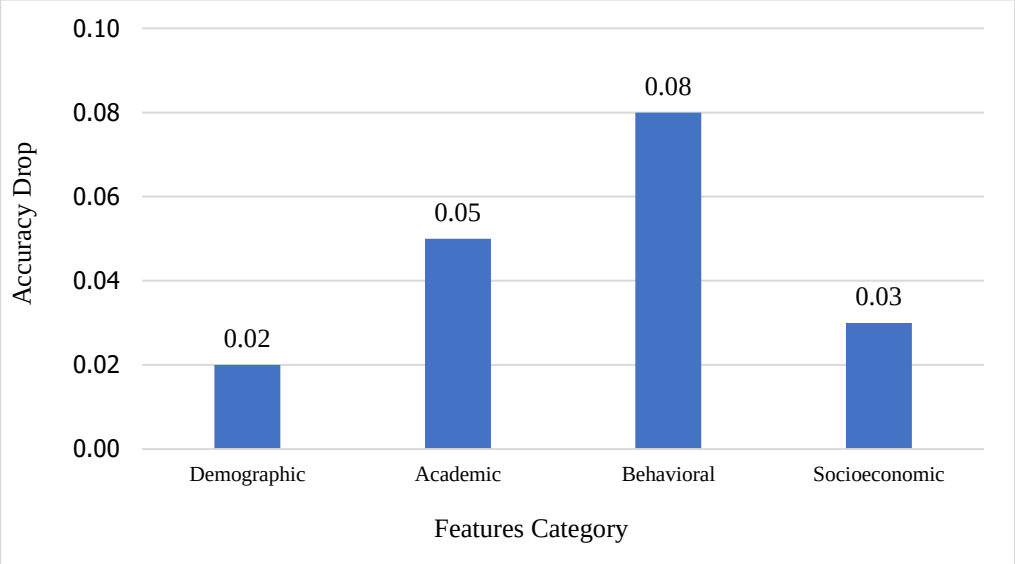

Fig 7. Analysing the impact of individual category on the model performance.

The feature importance analysis and ablation study reveal several key findings. Behavioral factors, particularly engagement with online learning platforms as captured in the OULAD dataset, were identified as the most critical predictors of academic success. Features related to the frequency and quality of interactions with learning resources significantly enhanced the model's accuracy. Academic factors, including consistent performance and regular attendance, also played a vital role in predicting student outcomes, highlighting their importance in any predictive model. On the other hand, socioeconomic and demographic factors, while less impactful, provided valuable contextual insights into the broader challenges affecting student performance. These factors were less directly predictive of immediate outcomes but contributed to a more comprehensive understanding of the influences on academic success. The analysis demonstrates that academic and behavioral features are essential for accurately modeling student performance, with behavioral data making the most distinct contribution. Socioeconomic and demographic factors, though less significant, remain valid for contextualizing predictions and designing tailored interventions.

5. Discussion

The findings of this study align with prior research highlighting the effectiveness of deep learning in predicting student academic success. Similar to studies by Hernández-Blanco et al. (2019) and Du et al. (2020), our results confirm that deep learning models outperform traditional machine learning approaches in handling complex, multivariate educational data. The MLP model achieved higher accuracy, precision, recall, and AUC-ROC scores than Decision Trees, Random Forests, and Support Vector Machines (SVM), reinforcing the adaptability of deep learning to non-linear patterns in student performance prediction.

However, our study extends beyond previous work by performing a feature importance analysis that quantifies the impact of different categories (behavioral, academic, socioeconomic, and demographic) on predictive accuracy. Unlike Alshamaila et al. (2024), who primarily focused on academic and demographic factors, our study highlights behavioral engagement from LMS interactions as the most influential predictor of student success. This finding underscores the importance of continuous student monitoring through digital learning platforms, providing new evidence supporting the integration of behavioral analytics into predictive models.

Despite these agreements, some differences exist between our findings and previous research. While Khan and Ghosh (2021) reported socioeconomic factors as a primary determinant of academic outcomes, our results indicate that these factors, while important, have a lower predictive impact

compared to behavioral and academic indicators. This discrepancy suggests that digital learning behaviors may play an increasingly significant role in student success in modern educational environments. Although our study demonstrates the potential of deep learning in predicting student academic performance, certain limitations must be acknowledged. First, the dataset used (OULAD) is specific to the Open University (UK), which may limit the generalizability of our findings to other educational institutions with different pedagogical structures and assessment methods. Future work should test the model on diverse datasets from various institutions to validate its applicability across different learning environments.

Second, while deep learning models such as MLP offer high accuracy, they often lack interpretability, making it difficult for educators to understand the specific factors driving predictions. This "black box" nature of deep learning presents a challenge in applying predictive insights for individualized student interventions. Future research should explore the integration of Explainable AI (XAI) techniques to improve model interpretability and provide actionable insights for instructors. Finally, our feature selection was limited to demographic, academic, behavioral, and socioeconomic factors. Other psychological and environmental variables, such as student motivation, stress levels, and access to learning resources, could further refine the prediction model. Incorporating these additional predictors could enhance the model's ability to capture a more holistic picture of student success. The findings of this study have significant implications for educational policy, institutional decision-making, and student support systems. Our results demonstrate that behavioral engagement and academic performance are the strongest indicators of student success, suggesting that educational institutions should implement early warning systems that monitor LMS interactions and coursework performance to identify at-risk students before their performance declines. Additionally, the insights gained from our model can help institutions design personalized learning pathways, ensuring that students receive targeted support based on their engagement patterns. By integrating predictive analytics into learning management systems, universities can enhance student retention, optimize academic interventions, and allocate resources more effectively. From a policy perspective, these findings emphasize the importance of data-driven decision-making in education. Policymakers should consider mandating learning analytics frameworks that track student progress and engagement, allowing for real-time adjustments to teaching methodologies and student support programs. The integration of AI-driven predictive models can also support personalized education strategies, adaptive learning environments, and equity-driven interventions, ensuring that students from diverse backgrounds receive the assistance they need to succeed. The results of this study highlight the effectiveness of deep learning, particularly the MLP model, in predicting student academic performance. By identifying key predictive factors, particularly behavioral engagement and academic records, this research provides valuable insights for educators, administrators, and policymakers. While the study confirms the advantages of deep learning over traditional methods, future research should address its generalizability, interpretability, and expanded feature selection to further enhance its practical application in diverse educational contexts.

6. Conclusion and Future Work

This study has demonstrated the effectiveness of deep learning approaches for predicting student academic performance, with our MLP model achieving 90.2% accuracy, significantly outperforming traditional machine learning approaches. The key contribution lies in identifying the relative importance of different feature categories, with behavioral factors (particularly engagement with online learning platforms) and academic history emerging as the most critical predictors of student outcomes. These findings challenge the conventional emphasis on demographic and socioeconomic factors in education research, suggesting that student behaviors and engagement patterns may be more actionable targets for intervention.

Our results have several practical implications for educational institutions. First, early warning systems should prioritize monitoring behavioral indicators like course interaction patterns and assignment submission timing, as these provide the strongest signals for at-risk identification. Second, academic support systems should be responsive to patterns in student performance, with targeted interventions triggered by changes in engagement metrics rather than waiting for formal assessment results.

The study has important limitations to acknowledge. The OULAD dataset represents a specific educational context (distance learning), potentially limiting generalizability to traditional classroom settings. Additionally, while our model achieved high accuracy, the black-box nature of deep learning approaches presents challenges for interpretability and explanation of specific predictions to stakeholders.

Future research should focus on: (1) developing explainable AI techniques to make deep learning models more transparent to educators; (2) investigating real-time prediction capabilities that integrate with learning management systems; (3) exploring transfer learning approaches to adapt models across different educational contexts; and (4) conducting longitudinal studies to validate the long-term effectiveness of interventions based on model predictions. By addressing these directions, researchers can further enhance the practical utility of deep learning approaches in supporting student success and transforming educational practice.

Acknowledgements

This paper is part of a project AOU_OM/2023/ITC1 funded by the Arab Open University. We would like to express our gratitude to the university for providing the opportunity to develop this area further.

References

Alamgir, Z., Akram, H., Karim, S., & Wali, A. (2024). Enhancing student performance prediction via educational data mining on academic data. Informatics in Education, 23(1), 1–24.

Al Husaini, Y. N., & Shukor, N. S. A. (2024). Virtual learning environment to predict students' academic performance using deep learning technique. In 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS) (pp. 1–7).

Al Husaini, Y. N. S., & Shukor, N. S. A. (2022). Factors affecting students' academic performance: A review. Res Militaris, 12(2), 284–294.

Al Husaini, Y. N., & Shukor, N. A. (2022). Prediction methods on students' academic performance: A review. Jilin Daxue Xuebao (Gongxueban)/Journal of Jilin University (Engineering and Technology Edition), 41(9), 196–217. https://doi.org/10.17605/OSF.IO/CHJF2

Alshamaila, Y., Alsawalqah, H., Aljarah, I., Habib, M., Faris, H., Alshraideh, M., & Obeid, N. (2024). An automatic prediction of students' performance to support the university education system: A deep learning approach. Multimedia Tools and Applications, 83, 46369-46396. https://doi.org/10.1007/s11042-024-16790-5

Azhar, M., Nadeem, S., Naz, F., Perveen, F., & Sameen, A. (2014). Impact of parental education and socio-economic status on academic achievements of university students. European Journal of Psychological Research, 1(1), 1–9.

Costa, A., & Faria, L. (2015). The impact of emotional intelligence on academic achievement: A longitudinal study in Portuguese secondary school. Learning and Individual Differences, 37, 38–47.

Dutt, A., Ismail, M. A., & Herawan, T. (2017). A systematic review on educational data mining. IEEE Access, 5, 15991–16005.

Du, X., Yang, J., Hung, J.-L., & Shelton, B. (2020). Educational data mining: A systematic review of research and emerging trends. Information Discovery and Delivery, 48(4), 225-236. https://doi.org/10.1108/IDD-06-2020-0062

Gaur, M., Faldu, K., & Sheth, A. (2021). Semantics of the black box: Can knowledge graphs help make deep learning systems more interpretable and explainable? IEEE Internet Computing, 25(1), 51–59.

Gay, G. (2021). Culturally responsive teaching: Ideas, actions, and effects. In H. M. K. Milner & K. Lomotey (Eds.), Handbook of urban education (2nd ed., pp. 212–233). Routledge.

Ghazvini, S. D., & Khajehpour, M. (2011). Gender differences in factors affecting academic performance of high school students. Procedia - Social and Behavioral Sciences, 15, 1040–1045.

Harris, A. L., & Robinson, K. (2016). A new framework for understanding parental involvement: Setting the stage for academic success. RSF: The Russell Sage Foundation Journal of the Social Sciences, 2(5), 186–201.

Henfield, M. S., Woo, H., & Bang, N. M. (2017). Gifted ethnic minority students and academic achievement: A meta-analysis. Gifted Child Quarterly, 61(1), 3–19.

Hussain, S., Gaftandzhieva, S., Maniruzzaman, M., Doneva, R., & Muhsin, Z. F. (2021). Regression analysis of student academic performance using deep learning. Education and Information Technologies, 26(1), 783–798.

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. Complexity, 2019, 1-21. https://doi.org/10.1155/2019/1306039

Hussain, M. M., Akbar, S., Hassan, S. A., Aziz, M. W., & Urooj, F. (2024). Prediction of student's academic performance through data mining approach. Journal of Informatics and Web Engineering, 3(1), 241–251.

Hu, P. J.-H., Hui, W., Clark, T. H., & Tam, K. Y. (2007). Technology-assisted learning and learning style: A longitudinal field experiment. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 37(6), 1099–1112.

Ilhan, F., Ozfidan, B., & Yilmaz, S. (2019). Home visit effectiveness on students' classroom behavior and academic achievement. Journal of Social Studies Education Research, 10(1), 61–80.

Ismail, L., Materwala, H., & Hennebelle, A. (2021). Comparative analysis of machine learning models for students' performance prediction. In Advances in Digital Science: ICADS 2021 (pp. 149–160).

Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. Education and Information Technologies, 26(1), 205-240. https://doi.org/10.1007/s10639-020-10321-9

Kaur, H., & Bathla, E. G. (2018). Student performance prediction using educational data mining techniques. International Journal on Future Revolution in Computer Science & Communication Engineering, 4(4), 93–97.

Kezar, A., Kitchen, J. A., Estes, H., Hallett, R., & Perez, R. (2023). Tailoring programs to best support low-income, first-generation, and racially minoritized college student success. Journal of College Student Retention: Research, Theory & Practice, 25(1), 126–152.

Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2024). A novel methodology using RNN + LSTM + ML for predicting student's academic performance. Education and Information Technologies. Advance online publication.

Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics Dataset (OULAD). Scientific Data, 4, 170171. https://doi.org/10.1038/sdata.2017.171

Luan, H., & Tsai, C.-C. (2021). A review of using machine learning approaches for precision education. Educational Technology & Society, 24(1), 250–266.

Lumley, M. A., & Provenzano, K. M. (2003). Stress management through written emotional disclosure improves academic performance among college students with physical symptoms. Journal of Educational Psychology, 95(3), 641–649.

Mendezabal, M. J. N. (2013). Study habits and attitudes: The road to academic success. Open Science Repository Education, Article e70081928.

Naicker, N., Adeliyi, T., & Wing, J. (2020). Linear support vector machines for prediction of student performance in school-based education. Mathematical Problems in Engineering, 2020, Article 4761468.Nagahi, M., Jaradat, R., Hossain, N. U. I., Nagahisarchoghaei, M., Elakramine, F., & George, S. R. (2020). Indicators of engineering students' academic performance: A gender-based study. In 2020 IEEE International Systems Conference (SysCon) (pp. 1–7).

Okoye, K., Nganji, J. T., Escamilla, J., & Hosseini, S. (2024). Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education. Computers and Education: Artificial Intelligence, 6, Article 100205.

Pelletier, D., Collerette, P., & Turcotte, G. (2012). Small schools in rural settings: Impact of a multi-school management approach on pupils, staff and parents. International Journal of Global Education (IJGE), 1(1), 1–15.

Roslan, M. B., & Chen, C. (2022). Educational data mining for student performance prediction: A systematic literature review (2015-2021). International Journal of Emerging Technologies in Learning (iJET), 17(5), 147–179.

Rienties, B., Beausaert, S., Grohnert, T., Niemantsverdriet, S., & Kommers, P. (2012). Understanding academic performance of international students: The role of ethnicity, academic and social integration. Higher Education, 63(6), 685–700.

Segura-Morales, M., & Loza-Aguirre, E. (2017). Using decision trees for predicting academic performance based on socio-economic factors. In 2017 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1132–1136).

Ullah, F., Salam, A., Amin, F., Khan, I. A., Ahmed, J., Zaib, S. A., & et al. (2024). Deep trust: A novel framework for dynamic trust and reputation management in the Internet of Things (IoT)-based networks. IEEE Access, 12, 87407–87419.

Venkatesan, R. G., Karmegam, D., & Mappillairaju, B. (2024). Exploring statistical approaches for predicting student dropout in education: A systematic review and meta-analysis. Journal of Computational Social Science, 7(1), 171–196.

Whitcomb, K. M., Kalender, Z. Y., Nokes-Malach, T. J., Schunn, C. D., & Singh, C. (2020). Engineering students' performance in foundational courses as a predictor of future academic success. International Journal of Engineering Education, 36(4), 1340–1355.

Zhong, L., Wei, Y., Yao, H., Deng, W., Wang, Z., & Tong, M. (2020). Review of deep learning-based personalized learning recommendation. In Proceedings of the 2020 11th International Conference on E-education, E-business, E-management, and E-learning (pp. 145–149).