International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I8P123
Volume 74 Issue 8, 338-358, August 2026

Original Article

Multimodal Predictive Modelling in Education: A Comparative Evaluation of ML, Boosting, DL, and Hybrid Fusion Architectures

Kanksha Kaur[1], Mrinalini Rana[2], Omdev Dahiya[3*]

[1, 2] *School of Computer Science and Engineering, Lovely Professional University, Punjab, India.*
[3]*School of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India.*

[3]**Correspondence Author : omdev@galgotiasuniversity.edu.in*

Received: 09 February 2026 Revised: 08 July 2026 Accepted: 15 July 2026 Published: 29 August 2026

Abstract - *The rapid growth of online learning platforms has generated large volumes of educational data that can support student performance prediction. However, existing Machine Learning (ML) and Deep Learning (DL) approaches often face challenges related to generalizability, feature integration, and early identification of at-risk students. This study presents a comparative evaluation of classical ML, boosting-based, DL, and a proposed Hybrid Artificial Neural Network-Deep Fusion Model (ANN-DFM) using the Open University Learning Analytics Dataset (OULAD). A unified experimental framework incorporating data cleaning, feature selection, and class balancing was employed to ensure fair benchmarking. Results show that Logistic Regression and Support Vector Machines achieved moderate accuracy (0.68-0.76), while Random Forest reached 0.91 accuracy. Boosting models, including XGBoost, LightGBM, and CatBoost, improved performance to 0.94-0.95 accuracy. Among DL approaches, ANN achieved 0.93 accuracy. The proposed ANN-DFM outperformed all baseline models, achieving 96.65% accuracy and an F1-score of 0.97, while demonstrating stable early-quarter predictions. The findings highlight the effectiveness of multimodal feature fusion for enhancing predictive accuracy and supporting early educational interventions.*

Keywords - *Educational Data Mining, Machine Learning, Deep Learning, Artificial Neural Networks (ANN), Online education.*

1. Introduction

Digital technology is rapidly transforming post-secondary education; this change has led to a corresponding increase in the amount of educational data generated through Learning Management Systems (LMSs), online testing, academic records, and student interaction websites. Collectively, these types of data offer useful opportunities to analyze how students behave, predict how well they will do in school, and support timely educational intervention. In light of this, Educational Data Mining (EDM) and Learning Analytics (LA) are important new fields of study focusing primarily on extracting meaningful information from educational contexts. EDM focuses primarily on developing computational-based and predictive modelling techniques, whereas LA is geared toward providing pedagogical interpretations of learning data and data-driven decision support for education [1]. In addition, the growing amount of educational-related data, coupled with advances in Machine Learning (ML) and Deep Learning (DL), is serving as a major impetus for creating intelligent predictive systems that aid educators in identifying at-risk learners; they can also help increase student retention and improve instructional effectiveness across multiple learning environments [2, 7, 8].

Predicting student performance is a difficult problem due to the diverse range of information in educational datasets, including demographic data, assessment data, and students' interactions with the system (i.e., their behavior while using a Learning Management System). Many existing approaches for predicting performance use a limited number of features and/or employ a modelling technique in isolation, limiting their ability to capture the complexity of learning effectively and to provide reliable warnings when students are at risk of poor achievement. Therefore, there is a pressing need for predictive frameworks that meaningfully integrate diverse sources of educational data while producing reliable, generalizable predictions.

Studies in Educational Data Mining (EDM) show that AI-based approaches are highly effective at predicting a student's performance in school. Classic machine learning models such as Logistic Regression, Support Vector Machines (SVMs), Random Forests, Decision Trees, and k-Nearest Neighbors (the original k of "KNN") are popular due to their simple algorithms, fast computational speeds, and ability to explain results [2-4].

On the other hand, the complexity of educational datasets has necessitated the use of ensemble (and boosting) techniques such as XGBoost, LightGBM, and CatBoost, which yield a significant increase in prediction accuracy over traditional classifiers in many studies [3-6]. Likewise, many deep learning architectures such as ANNs, CNNs, LSTMs, and hybrid CNN-LSTM models have produced accurate representations of sequential patterns of learning and engagement for learners in online learning environments [11-13]. There are still many significant challenges to developing student performance prediction tools, despite recent advancements in this field. The majority of current models are designed with either course-specific or institution-specific datasets, which restrict their use across a wide variety of educational settings [6, 12].

In addition, educational datasets are often subject to various problems, including missing values, class imbalance, noisy behaviour logs, and irrelevant attributes, all of which can degrade the reliability of predictions and the resulting models [12-14]. While gradient boosting and hybrid deep learning methods frequently achieve high predictive accuracy, they often entail increased architectural complexity, low interpretability, and high computational costs. Lastly, the vast majority of EDM research studies still place primary emphasis on static academic attributes while paying less attention to the multimodal integration of behavioural, demographic, and performance-based features [7, 15, 16].

One more restriction identified in the current literature is the lack of systematic comparison of multiple prediction paradigms via reliable experimental conditions, where to have fair benchmarking based on standardized preprocessing methods, feature engineering methods, and evaluation metrics is important when there is so much research that shows that any one type of machine learning, boosting, or deep learning is superior (but has not been benchmarked appropriately). Very little attention has also been given to predicting early-stage academic performance, where identifying at-risk learners promptly is essential for implementing effective intervention strategies. As these limitations show, there is a need for a robust multimodal predictive framework that can integrate disparate educational features while ensuring stable, generalizable predictions and fair comparative validation.

While researchers have recently developed numerous hybrid prediction methods in Education and Data Mining, most of these methods primarily focus on improving predictive accuracy by employing limited combinations of input features or by performing limited comparative evaluations across different learning paradigms. Additionally, very few studies have examined holistic multimodal fusion frameworks that integrate demographic, assessment, and behavioral interaction data using a consistent benchmark environment. Therefore, the research community needs to develop more complete and effective hybrid architectures capable of producing reliable predictions, stability over time, and improved generalization.

Prior studies have reported additional results for Random Forests, XGBoost, LightGBM, CatBoost, Artificial Neural Networks, and hybrid CNN-LSTM architectures for predicting student achievement. Generally, these methods have shown very good predictive performance; however, most studies have focused on only one model family, one dataset, or a small number of features. This makes it very difficult to conduct direct comparisons between models from classical machine learning, gradient boosting, and deep learning paradigms in a standardized way. Because of this, there is currently a need to develop an integrated framework for comparing the performance of predictive models using multiple types of educational data simultaneously.

To fill these knowledge gaps, authors propose an ANN-DFM Hybrid Model (Artificial Neural Network - Deep Fusion Method) to predict student performance using the OULAD (Open University Learning Analytics Dataset). The suggested framework will merge demographics, assessments, and behavior patterns. It will be developed using a multimodal fusion approach to ensure more reliable, robust predictions that help identify students who may be at risk early on. Authors also compare the results of ANN-DFM with those of conventional machine learning models, gradient-boosted models, and deep learning algorithms using a unified experimental framework.

Previously, predicting student performance typically involved a single method (e.g., machine learning), a small number of boosting methods, and/or deep learning methods. While hybrid model approaches exist, very few studies have thoroughly examined the predictive power of multiple feature types simultaneously or compared results from different feature-method combinations under controlled experimental conditions.

Additionally, multimodal fusion techniques that combine demographic, test score, and behavioral data within a single model framework have not yet been fully investigated in Educational Data Mining. In addition to the above limitations, researchers have not provided sufficient evidence on the predictive stability of the different types of models they developed, based on data available in the same time frame as early in students' academic careers. To fill these gaps in previous research, this study created and tested a Hybrid ANN-DFM framework to predict student performance by using a single criterion for comparison against existing models in the following model categories:

Classical (i.e., non-computer-based), Gradient Boosted, Deep Learning, and Hybrids of multiple model types that utilize a combination of data from demographic, test score, and behavioral assessments.

The major contributions of this study are summarized as follows:

A comprehensive comparative framework is developed to evaluate the proposed Hybrid ANN-DFM model against classical machine learning, gradient boosting, and deep learning approaches commonly applied in Educational Data Mining.

- A multimodal feature integration strategy combining demographic, assessment, and behavioral interaction features is implemented to improve predictive robustness and representation of learner behavior patterns.
- A standardized benchmarking framework based on accuracy, precision, recall, and F1-score is adopted to ensure fair and consistent evaluation of all predictive models.
- The influence of feature reduction, class balancing, and fusion-based learning on model stability and generalization capability is systematically analyzed.
- The proposed ANN-DFM architecture demonstrates strong potential for early identification of at-risk learners, supporting the development of effective early-warning intervention systems.
- The novelty of this study lies in the development of a Hybrid Artificial Neural Network-Deep Fusion Model (ANN-DFM) that integrates demographic, assessment, and behavioral learning features through a multimodal fusion strategy. Unlike previous studies that primarily evaluate isolated machine learning, boosting, or deep learning techniques, the proposed framework combines heterogeneous educational data sources within a unified predictive architecture. It validates its effectiveness through comprehensive benchmarking under identical preprocessing and evaluation conditions. This approach enables more reliable student performance prediction and strengthens early identification of academically at-risk learners.
- Despite substantial advances in educational data mining and student outcome prediction, several research challenges continue to persist. Traditional predictive techniques, including Logistic Regression, Support Vector Machines, and Random Forests, have shown reasonable effectiveness in identifying patterns of academic performance. However, these methods often struggle to model the intricate, nonlinear interactions among diverse educational factors. Recent deep learning models have addressed some of these limitations by automatically extracting complex feature representations from large datasets. Nevertheless, many existing studies focus on a restricted range of inputs, such as demographic information, assessment results, or learning management system activities, thereby limiting the utilization of the full spectrum of educational data. In addition, comparisons across different predictive approaches are frequently conducted under varying experimental settings, making objective evaluation challenging. Concerns about model transparency, limited cross-model benchmarking, and underexplored integration of heterogeneous data sources indicate a need for more comprehensive prediction frameworks. Therefore, there remains a strong motivation to develop and evaluate multimodal learning architectures that can effectively combine diverse educational indicators while delivering reliable and generalizable predictive performance.
- Despite significant advancements in predicting student academic performance using machine learning, gradient boosting, and deep learning, several limitations remain. Various studies tend to use only a subset of the full set of demographics, assessments, or behavioral characteristics available, thereby underutilizing the full suite of educational multimodal data. Additionally, few studies compare classical machine learning, gradient boosting, deep learning, and a fused architecture under the same experimental conditions.

Therefore, this research aims to develop a Hybrid Artificial Neural Network-Deep Fusion Model (ANN DFM) integrating all educational variables/hierarchies in a single multimodal framework. Further, this model aims to increase accuracy and reliability in predicting student academic performance and will allow for earlier identification of at-risk students by comparing Artificial Neural Network (ANN) and Deep Fusion Model (DFM) predictions.

This paper contains the following sections: Sec 2 presents literature on Educational Data Mining and Learning Analytics, including classical machine learning approaches, gradient-boosting methods, deep learning methods, and hybrid predictive modelling methods. Sec. 3 describes the dataset used, the preprocessing and feature engineering methodologies, and the architecture of the proposed Hybrid ANN-DFM model. Sec. 4 describes the methodology for comparison, the metrics used to evaluate the performance of models produced by the proposed Hybrid ANN-DFM model, and the benchmarking framework used to evaluate the predictive models built with the above-mentioned model. Sec. 5 presents an experimental section comparing the performance of models generated by the method described in this paper. Discussions and implications of the results from Sec 5 will be presented in Sec 6, along with directions for future research in Sec 7.

2. Literature Review and Existing Work

2.1. Overview of Educational Data Mining (EDM) and Learning Analytics (LA)

Educational Data Mining (EDM) and Learning Analytics (LA) have evolved into closely related fields focused on deriving actionable insights from student data. A comparative analysis of 492 LA and 194 EDM studies showed that the two areas overlap in terms of the tools of analysis, keywords, and methodological basis. Still, EDM is more focused on technical approaches, and LA is more focused on social and context-based constructs [1]. Although different, both disciplines are plagued by a lack of a proper theoretical foundation, with self-regulated learning being the most frequently mentioned theoretical perspective [18].

EDM has been well known for revealing latent relationships in educational data and for anticipating student performance. Reviews note that it is increasingly becoming an interdisciplinary issue, can aid evidence-based decision-making, and can enhance student learning experiences and institutional efficiency [9-11, 14-16]. Nevertheless, some studies also point to regulatory limitations, ethical concerns, and difficulties in managing sensitive educational information, especially in the context of deep learning implementation [9].

2.2. Classical Machine Learning Models for Predicting Student Outcomes

Conventional Machine Learning (ML) models remain central to predicting student performance. Some studies have used algorithms such as logistic regression, support vector machines, random forests, Naive Bayes, and k-nearest neighbors to predict academic performance and identify students at risk [2]. For instance, Yağcı [2] evaluated six traditional ML algorithms on midterm exam grades, department-level data, and professor-level data, and reported a 70-75 percent accuracy of classification, demonstrating the importance of structured numeric variables in predicting scholar performance. Random Forest is usually regarded as one of the top-performing methods in EDM environments because it is strong and interpretable/can be used to provide explanations of results [3, 8, 14, 19]. While traditional ML algorithms are popular, many researchers have reported that such algorithms suffer from generalizability issues, are dataset-dependent, and do not handle the temporal and behavioral complexities of current educational systems well [7, 9, 16]. Consequently, the continual trend towards implementing more advanced deep learning and hybrid techniques supports the trend away from traditional approaches.

2.3. Deep Learning Approaches in Performance Prediction

Deep Learning (DL) has been gaining popularity in EDM because it can model nonlinear relationships and handle high-dimensional data. The results of studies using CNNs, LSTMs, and hybrid neural architectures have shown higher predictive performance than those of classical ML approaches [11-14]. In research conducted by [11], LMS-based student interaction data were used with CNNs and LSTMs to predict key behavioral indicators associated with (and hence correlated with) academic success. CNN-BiLSTM hybrid architecture models faced challenges with missing values, imbalanced classes, and irrelevant features, but once these factors were accounted for, predictive performance improved [12]. The use of efficient feature engineering methods and hyperparameter optimization improved model stability and accuracy, while demonstrating strong generalization on benchmark datasets.

The current trend in research has been toward hybrid DL models that combine the capabilities of CNNs (local feature extraction) with those of LSTMs/BiLSTMs (sequential modelling). The reported accuracy of these models has reached 99%, exceeding that of stand-alone ML/DL alternatives [13]. Historical studies indicate that ANN and Random Forest algorithms are the most accurate; however, an increasing number of studies suggest that deep learning-based solutions are becoming preferable when sufficient data on both behavior and timing are available [14].

2.4. Gradient Boosting Methods in Prior Research

Gradient boosting models have been proposed as high-performance alternatives to EDM because they can handle heterogeneous features and generalize well. In [3], XGBoost and Random Forest were compared; both models identified relevant predictors of academic success. LightGBM, being the most efficient and scalable model, was found to be more effective than ten classical ML algorithms in predicting online learning performance based on student interaction behaviour data [4].

CatBoost has gained significant popularity because it can handle categorical and missing data with minimal preprocessing, making it particularly applicable to educational datasets. Hybrid CatBoost models with Victoria Amazonica Optimization (VAO) and Artificial Rabbits Optimization (ARO) achieved significant performance improvements, reaching a maximum of 6% in classification accuracy and precision [5]. The other study reaffirmed CatBoost as the most effective algorithm for accuracy, ROC curves, and F-measure in training a course-agnostic student risk prediction model [6]. On the whole, the gradient boosting family consistently outperforms classical ML models and can be compared to other hybrid deep learning systems.

2.5. Multimodal Fusion Approaches in other Domains

Multimodal fusion approaches have demonstrated remarkable effectiveness across several artificial intelligence domains beyond education, including healthcare analytics, intelligent recommendation systems, computer vision, and human behavior analysis [27, 28]. Likewise, multimodal deep learning frameworks that combine textual, visual, and sequential information have achieved superior outcomes in areas such as emotion recognition, decision-support systems, and intelligent recommendation applications [27, 28]. The above findings indicate that architectures designed for fusion can discover complementary relationships among features and hidden connections between two (or more) features that would typically be hard to discern through separate modelling techniques. Thus, the successful use of multimodal representation learning across broad areas of artificial intelligence research simultaneously provides a strong theoretical basis for integrating fusion-based frameworks, such as the proposed ANN-DFM model, into Educational Data Mining environments.

Recent multimodal studies have also shown that both feature-level fusion and hybrid representation learning are beneficial for generalization, information integrity, and robustness when working with heterogeneous data types. These results provide strong justification for combining demographic, assessment-based, and behavioral learning characteristics into one integrated predictive framework for evaluating student success.

2.6. Studies Using LMS and Behavioral Interaction Data

There is much contemporary research on the behavioral, interactive, and temporal variables in Learning Management Systems (LMSs), which can be especially helpful in improving both blended and online learning experiences. Several studies have used LMS interaction data to create predictive models for student retention, attrition risk, and overall course success [8, 9, 11]. For example, Fahd et al. [8] examined data from blended learning environments and achieved 85% accuracy with Random Forests, indicating that behavioral patterns derived from LMS data may be powerful indicators of a student's overall risk level.

Other studies employing a CNN-LSTM architecture on LMSs, such as Blackboard, have shown high predictive value and the presence of clearly actionable Key Performance Indicators (KPIs) associated with online engagement [20]. Due to the proliferation of LMS-produced behavioral data, the use of more sophisticated machine learning/deep learning approaches to model sequential interaction will become increasingly necessary to identify latent patterns that would be challenging to uncover through traditional means. Figure 1 summarizes the progression of research in the literature review, from traditional Machine Learning (ML) to the hybrid model proposed in this study.

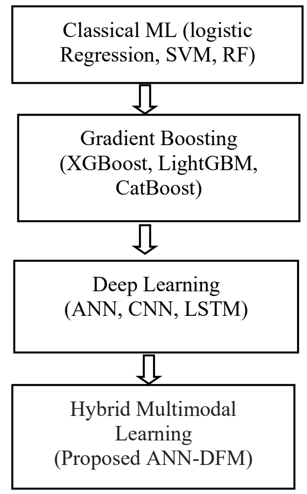


Fig. 1 Evolution of student performance prediction approaches

The literature indicates significant development in predicting student outcomes using traditional Machine Learning (ML), gradient boosting, and Deep Learning (DL) methods. While each of these studies provides information on predicting student performance, there are significant differences in the data used, model development approaches, predictive accuracy, and the limitations of each study. To provide a general summary of current research opportunities, Table 1 provides an overview of several studies and their challenges, clearly identifying the need to develop the ANN-DFM framework presented in this paper.

Table 1. Comparative analysis of existing student performance prediction studies

Study	Year	Method	Accuracy	Key Limitation
Yağcı [2]	2022	Random Forest	75%	Limited multimodal integration
CatBoost Study [26]	2023	CatBoost	92%	Weak behavioural modelling
CNN-LSTM [23]	2024	CNN-LSTM	92.4%	High computational cost
Proposed ANN-DFM	2026	Hybrid Fusion	96.65%	Higher training complexity

As demonstrated by Table 1, while many current methodologies demonstrate great success in making predictions with respect to education, there are a number of significant limitations. The majority of approaches to predicting student outcomes rely heavily on only a limited number of educational attributes and fail to incorporate the full complement of available demographic, assessment, and behavioral attributes simultaneously. In addition, there has been very little comparative performance evaluation for classical ML, GBM, DL and MMF methodologies when using identical experimental conditions. These observations highlight a significant research gap that was addressed within the proposed ANNDFM framework.

2.7. Summary of Strengths and Gaps in Existing Work

Across the 16 studies, several key strengths and research contributions emerge:

Strengths observed in prior work:

- For example, "ML and DL techniques always show a high predictive power to identify at-risk students" [2-4, 6].
- Gradient boosting models (XGBoost, LightGBM, CatBoost) are often better than classical models because of better feature handling and scalability [3-6].
- Hybrid deep learning models (CNN- LSTM, CNN-BiLSTM) deliver state-of-the-art accuracy on multiple benchmark datasets [11-13].

- Extensive surveys give some insight about effective algorithms, performance metrics, and best practises in EDM [7, 14-16].

Gaps and limitations frequently reported:

- Several studies lack theoretical grounding or explicit frameworks [1].
- Issues such as missing data, class imbalance, and irrelevant features remain persistent obstacles [12-14].
- Deep learning adoption is slowed by ethical constraints, interpretability challenges, and regulatory concerns [9].
- Few studies deploy hybrid models that combine feature-level and model-level fusion, indicating an opportunity for architectures like ANN-DFM.
- Many prediction models are course-specific and fail to generalize across diverse learning environments [6, 12].

Addressing existing limitations, this study introduces a Hybrid Artificial Neural Network-Deep Fusion Model (ANN-DFM) that integrates demographic information, assessment data, and LMS behavioural interaction characteristics into a single multimodal learning framework. In contrast to other current methods that tend to use a single type of educational attributes or evaluate only a small subset of benchmark models, this new framework incorporates multiple kinds of complementary features by way of a fusion mechanism. It provides the opportunity to thoroughly benchmark itself against classical machine learning, gradient boosting, and deep learning models. The architecture developed in this study has been created to maximize predictive accuracy, increase robustness, and be able to identify students at risk of failing earlier in their educational careers than previously demonstrated in the literature.

2.8. Need for Comparative Evaluation against the Hybrid ANN-DFM Model

The reviewed literature points out the rapid progress of classical ML, gradient boosting, and hybrid deep learning models, but there are still many limitations regarding their generalizability, interpretability, and the ability to handle different types of features. Classical ML approaches are highly dependent on handcrafted features and are not always able to capture the complexity of behavior [2, 7, 21]. Gradient boosting methods are very good with structured and categorical data and do not have the representational power of deep neural networks for sequential or high-dimensional features [3-6]. Hybrid DL approaches achieve better performance than ML and standalone DL but are usually more about the modelling of sequences than about the fusion of features [11-13]. A combined hybrid ANN-DFM model that integrates demographic, assessment and behavior features with advanced fusion techniques directly tackles the shortcomings of previous studies because they have shown some limitations. Therefore, a comparative evaluation was conducted to determine whether hybrid architectures achieve superior predictive performance compared with existing ML, DL, and boosting-based approaches.

3. Overview of the Proposed Hybrid ANN-DFM Model

The proposed hybrid ANN-DFM combines two different types of data: structured demographic and assessment data, and behavior-based learning interaction features into an integrated system to use for predicting student performance in educational analytics. The model analyzes both types of data independently first using an individual sub-network, then combines them together in the fusion layer of the model. This approach enables the model to learn complementary types of data from multiple disparate sources; this allows for improved accuracy and earlier detection of risk as well as greater prediction performance stability than those seen with classic methods. Data acquisition and pre-processing; Class balancing and feature elimination; Multiple model families trained and evaluated using standardized performance metrics; Comparative analysis showed Hybrid ANN - DFM is the best-performing Model (Figure 2).

3.1. Dataset Description (OULAD)

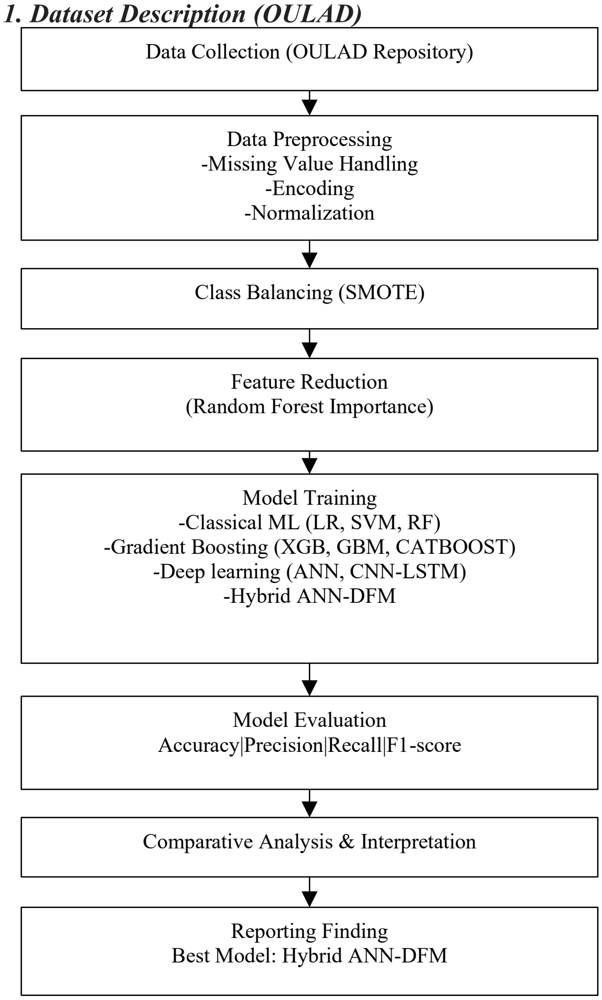


Fig. 2 Flow of proposed hybrid ANN-DFM model

The Open University Learning Analytics Dataset (OULAD) within the Hybrid ANN-DFM Model provides both demographic information and assessment outcomes, as well as all forms of interaction with Virtual Learning Environments (VLEs).

This multimodal framework of the Hybrid ANN-DFM Model means that it can be used to apply other hybrid deep learning techniques to analyze the data.

3.2. Feature Groups: Demographic, Assessment, Behavioral (VLE)

Three main feature categories were used:

- Demographic: age band, gender, education level, disability status.
- Assessment: continuous assessment marks, submission records, exam scores.
- Behavioral (VLE): resource views, quiz attempts, clickstream logs, and interaction frequency.

These features together represent academic background, performance history, and engagement behavior.

3.3. Data Preprocessing Strategy

Missing values, normalization of numerical fields, encoding categorical variables and smoothing extreme behavior values went into the pre-processing stage.

Using stratified training and testing splits ensured that all target class distributions were represented equally, along with missing numerical values imputed with median imputation and categorical attributes imputed with mode imputation to reflect data accuracy, while invalid or too many missing data records were dropped before conducting any analysis in order to maintain the integrity of the dataset.

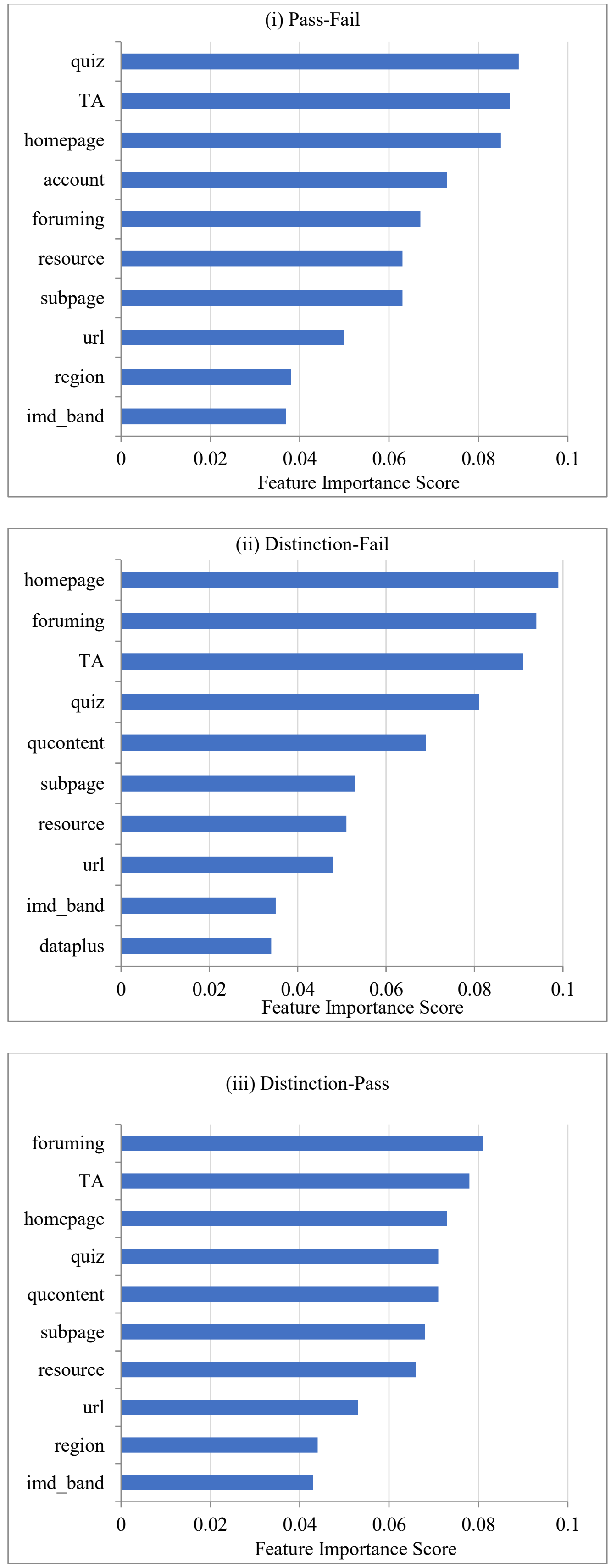
3.4. Feature Reduction using Random Forest

Low-impact variables were removed using the Random Forest (RF) predictor variable (feature) importance method.

By selecting only, the most important demographic, assessment, and behavioral variables, the model complexity was reduced and made more efficient as well.

In Figure 3, the features considered most important for predicting the model's accuracy across the four target classes are shown.

Behavioural indicators of study are the most important, based on their strong predictive ability, and include factors such as quizzes completed, how long users spend on the home page, and their use of forums, compared to the demographic and static assessment variables.



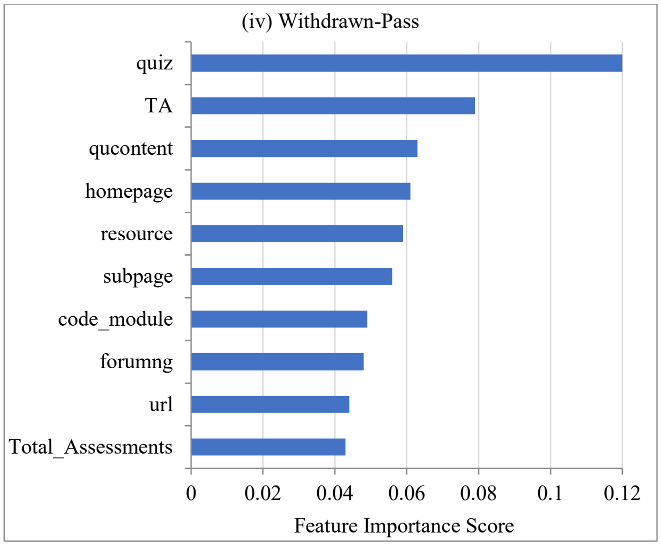


Fig. 3 Feature importance scores across four outcome classifications

3.5. Handling Imbalanced Data via SMOTE

To address the imbalance in pass-fail results, SMOTE was used to oversample minority classes. This step ensured fair model learning and improved recall for the at-risk learners.

3.6. Architecture of the ANN Component

Using the ANN, structured demographic and assessment features can be processed across multiple dense layers, with dropout and batch normalization. The ANN identifies a compact latent representation of the structural inputs, which is then blended with the behavioral input features.

3.7. Proposed Hybrid ANN-DFM Architecture

The proposed Hybrid Artificial Neural Network-Deep Fusion Model (ANN-DFM) was designed as a multimodal predictive framework for student performance analysis using heterogeneous educational data sources. The architecture integrates demographic information, academic assessment records, and behavioral interaction features extracted from Learning Management System (LMS) activity logs to improve predictive learning capability. The framework aims to capture complex relationships between learner engagement patterns and academic outcomes by combining multiple feature categories within a unified deep learning environment.

The proposed ANN-DFM architecture was developed to integrate heterogeneous educational features through a multimodal fusion learning framework. The model combines demographic information, assessment-related attributes, and behavioral interaction indicators from LMS activity logs to capture complex relationships associated with student academic performance. Structured input features were processed through interconnected dense neural layers capable of learning nonlinear educational patterns and latent feature dependencies.

The hybrid fusion architecture was selected to overcome limitations observed in standalone machine learning and deep learning models commonly applied in Educational Data Mining research. Behavioral and academic feature representations were combined using a fusion mechanism before final classification.

The proposed ANN-DFM architecture was developed as a multimodal deep learning framework capable of analyzing complex educational data patterns derived from heterogeneous student information sources. The model employs a sequence of fully connected dense neural layers with 128, 64, and 32 neurons to capture hidden nonlinear relationships among demographic, academic, and behavioral attributes. ReLU activation functions were utilized across intermediate layers to strengthen feature extraction and learning efficiency, whereas the final classification layer applied the Softmax activation function for multi-class prediction tasks. Training optimization was conducted using the Adam optimization algorithm with a learning rate of 0.001 and categorical cross-entropy as the objective loss function. In addition, dropout regularization and batch normalization mechanisms were integrated into the training process to minimize overfitting and improve model stability and generalization performance (see Table 2).

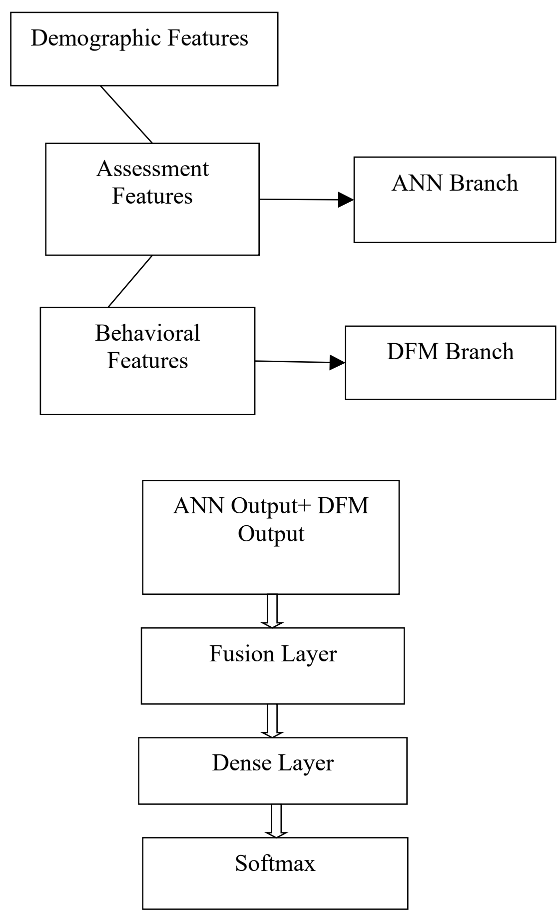


Fig. 4 Detailed ANN-DFM architecture

Table 2. Hyperparameter configuration of ANN-DFM

Parameter	Value
Input Features	Demographic + Assessment + VLE
Hidden Layers	3 Dense Layers
Neurons	128, 64, 32
Activation Function	ReLU
Output Activation	Softmax
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epochs	100
Dropout Rate	0.3
Loss Function	Categorical Cross-Entropy

3.8. Deep Fusion Model (DFM)

The behavioral (VLE) inputs are handled by the DFM using a parallel dense sub-network. ANN embeddings and behavioral representations are concatenated in a fusion layer, followed by additional dense layers to make final predictions using a sigmoid or softmax classifier.

3.9. Motivation for Combining ANN + DFM

The hybrid ANN-DFM design leverages the strengths of both components:

- ANN model's nonlinear patterns in structured data.
- DFM captures engagement and behavioral dynamics.

Fusion integrates complementary information, improving accuracy, stability, and generalizability.

3.10. Summary of Baseline Models

Three baseline models were used for comparison:

- A standard ANN using only structured features.
- A Random Forest model as a classical ML benchmark.
- A reduced-feature ANN using RF-selected top features.

These baselines help isolate the performance contribution of the fusion mechanism.

3.11. Detailed Architecture and Hyperparameter Configuration

The proposed Hybrid ANN-DFM framework has several different connected processes that help combine demographic, academic, and behavioral characteristics of learning all into one multimodal prediction system. Educational characteristics that are structured were first processed using dense fully connected ANN layers with ReLU activation functions, which capture nonlinear interaction between features. Behavioral features for LMS/VLE interaction were simultaneously processed through representation layers focused on fusion, which were used to maintain the temporal and engaged patterns of interaction. Then, the outputs of these two streams of parallel learning were fused into one before being classified through the use of the sigmoid activation function. When training the model, dropout regularization and early stopping techniques were used to help enhance generalization ability and reduce overfitting. Training of the model occurred using the Adam optimizer with a Binary cross-entropy loss function and control of pre-processing variations to validate results under standardized conditions.

4. Comparative Methodology

4.1. Comparative Research Framework

The comparison framework provides a comprehensive assessment of the effectiveness of the proposed Hybrid ANN-DFM method compared to several machine-learning, deep-learning, and Gradient-Boosting (GB) techniques, which are commonly employed in Educational Data Mining (EDM) research. In order for the comparison to be valid and unbiased, each of the models were trained and tested on the same data set. All models were trained and tested at the same pre-processing stage. The models were evaluated using the same set of performance metrics. This comparison framework allows for the comparing of the relative strengths and weaknesses of the models under the same experimental conditions and provides information regarding the performance of the hybrid fusion architecture compared to traditional and contemporary predictive methods.

4.2. Selection Criteria for Existing Models

The reasoning behind this model comparison is based on the theoretical background found in earlier EDM research, specifically looking for models with diverse methodologies that can be used to compare how well one model works compared to the other. Logistic Regression, Support Vector Machines, and Random Forest were selected because these models have been widely applied in educational prediction studies and provide interpretable predictive outcomes regarding student behavior/characteristics and performing well in predicting students before they leave (even though these models were originally developed for predicting individuals can be examined). Also, important to note is the addition of several gradient boosting models (e.g., XGBoost, LightGBM, CatBoost); these models have exhibited excellence while working with heterogeneous educational data sets (including categorical and tabular variables). Modern neural modeling approaches were captured by using standard deep learning model types (ANN and CNN-LSTM). With these two model types combined, they allow for extended modelling capacity of complex, nonlinear relationships and sequential behavioral interactions occurring in Educational Data Mining (EDM). This will provide an opportunity to assess the effectiveness of all the current predictive paradigms used in EDM research.

4.3. Standard Evaluation Metrics

To allow for a rigorous basis of comparison, all models were evaluated using four performance metrics that are widely used: accuracy, precision, recall and F1-score. These metrics provide a complete perspective on the model's behavior by not only measuring the overall accuracy of the predictions, but also the sensitivity of the model to at-risk students, many of whom typically constitute minority classes in educational datasets. Precision and recall are particularly significant in EDM contexts, as they are essential elements in minimizing false alarms while also maximizing the identification of truly at-risk learners. The F1-score, which represents the harmonic mean of precision and recall, provides a balanced evaluation of model performance. Precision and recall provide a balanced view of the ability of each model to handle class imbalance and also represent the trade-off between the two. The use of this group of metrics helps to support a multidimensional evaluation of the predictive performance.

4.4. Comparative Experimental Setup

All experiments were performed in carefully controlled conditions in order to remove variability caused by inconsistency in data processing or training. Each model was fed with the same stratified train-test split of the OULAD dataset, and all missing and inconsistent data were handled in the same way for all methods. Preprocessing consisted of normalization of continuous attributes, one-hot encoding of categorical fields, and balancing of the target variable using SMOTE to account for the inherent skew in pass-fail distributions. Classical machine learning models were optimized using grid search methodologies, which enabled fair tuning of hyperparameters, deep learning and hybrid models used early stopping, optimization using validation and standardized activation and regularization mechanisms. This methodological consistency ensured that performance outcomes could be assigned to model architecture rather than to external experimental conditions.

4.5. Experimental Environment and Reproducibility

In this work, a Python-based predictive framework was implemented using Jupyter Notebook. The libraries used included Pandas and NumPy for data manipulation and Scikit-learn for machine learning algorithms. Imbalanced-learn was used to address class imbalance, while TensorFlow/Keras were used to create and train deep neural network models.

All computational experiments occurred on a desktop computer with an Intel i7 processor with 16GB of RAM and Windows 11 OS. To maintain a level playing field among all competing methods, the same training and testing splits, preprocess methods, and evaluation metrics were utilized across all models. Hyperparameters were determined via repeated experimental iterations, with the final set of hyperparameters chosen based on validation results.

The analysis was fully reproducible with a random seed of 42 for the dataset split, parameter initialization, and model training. A stratified sample was taken when splitting the data, thereby ensuring the original distribution of classes was maintained in both training and testing subsets. SMOTE was used to mitigate the impact of class imbalance and enhance the ability to learn from those underrepresented classes.

The effectiveness of the model was evaluated using multiple criteria, including Accuracy, Precision, Recall, F1 Score and AUC-ROC. Each experiment was repeated 10 times under randomly changed initial conditions in order to reduce the effect of random variability, so that summary results could be presented in terms of both the average of all runs and the standard deviation from the mean. The entire experimental process, including data preparation; feature engineering; model building; and evaluating performance, was carried out in a transparent and reproducible way, so that other investigators could reproduce and confirm any part of the proposed ANN-DFM framework by accessing details regarding all methodological selection, parameter set and analysis procedures.

4.6. Consistent Feature Sets and Preprocessing

One of the main design principles of the comparative study was to make sure all models operated on the same feature representations. Feature reduction was carried out using Random Forest importance scores and resulted in a refined set of demographics, assessment and behavioral variables. These reduced features were provided to all models except for the proposed hybrid model, which also used behavioral features in their raw processed form before fusion.

This decision was made in order to ensure that the ANN--DFM model is able to make good use of complementary behavioral and structured information through its fusion mechanism while still being fair by limiting the structured inputs to the same dimensionality as the classical algorithms. By standardizing the feature sets and preprocessing steps, the study ensured that the models were equivalent and that the influence of unintended information leakage was minimized.

4.7. Fair Benchmarking Procedures

The benchmarking process was based on strict reproducibility principles. In order to reduce the effect of initial randomness, each model was trained across several runs, and the final reported scores are the averaged performance across these runs. The modelling comparison was conducted in similar computational environments, using the same batch size and evaluation protocols. This methodology helps ensure that any bias resulting from differences in hardware acceleration or the stochastic effect of training that is characteristic of all neural networks does not affect results. Collectively, these methodological safeguards should make sure performance comparisons are based on real architectural differences rather than experimental artifacts.

4.8. Limitations of the Comparative Methodology

Although this methodology was created in a comparative manner that is both thorough and equitable, it does have some unavoidable restrictions. Data will be gathered from a single institution (e.g. OULAD), so the results may be less varied to allow for the ability of them to be generalized across other larger (broader) education systems/geographies. Moreover, while deep learning models can be trained with large-scale datasets with high temporal resolution, the behavioral logs in OULAD do not fully reflect session-level or sequence-specific dynamics, which could limit the performance of LSTM or CNN-based models. Additionally, hybrid architectures such as ANN-DFM inherently differ from classical ML models in the representational complexity to the point that complete equivalence in feature modelling is not possible. Despite these limitations, the methodology used provides a strong comparative basis that is consistent with best practices in EDM research. Table 3 gives an overview of all the predictive model categories included in this study with respect to their compatibility with the data, the main strengths, and known limitations. It develops a base comparison of classical machine learning techniques, gradient boosting algorithms, deep learning architectures, and the proposed Hybrid ANN - DFM model to support the rationale of their inclusion in the benchmarking framework.

Table 3. Comparative summary of included models

Model Category	Model	Data Compatibility	Key Advantages	Key Limitations
Classical Machine Learning	Logistic Regression	Structured/tabular data	High interpretability; simple to train	Performs poorly with nonlinear patterns
	Random Forest	Structured/tabular	Strong baseline accuracy; provides feature importance	Limited modeling of temporal behavioral data
	SVM	Structured/tabular	Effective for small, high-dimensional datasets	Computationally expensive for large datasets
Gradient Boosting	XGBoost	Mixed structured data	High accuracy; handles missing values efficiently	Requires extensive tuning; less suited for behavioral sequences
	LightGBM	Structured, categorical	Fast training; scalable to large datasets	Sensitive to noisy or unbalanced features
	CatBoost	Structured, categorical	Excellent handling of categorical features	Limited ability to model complex temporal patterns
Deep Learning	ANN	Structured features	Captures nonlinear interactions	Cannot model behavioral dynamics alone
	CNN-LSTM	Sequential/behavioral	Strong for engagement and temporal modeling	High computational overhead; sensitive to data volume
Proposed Hybrid Model	ANN-DFM	Structured + behavioral blended features	Integrates multimodal inputs; improves generalization and robustness	Higher training cost; increased architectural complexity

4.9. Experimental Environment

All experiments were conducted using Python and TensorFlow/Keras. Using a fixed random seed of 42 during data splitting/model initialization, to guarantee consistency, has been done for every experiment that was conducted ten times per experiment, and the average of the performance measures seen below (in Table 4) are presented.

Table 4. Computational resources and configuration settings for ANN-DFM experiments

Parameter	Value
CPU	Intel Core i7 / equivalent
RAM	16 GB
Framework	TensorFlow/Keras
Python Version	3.x
Operating System	Windows
Random Seed	42
Number of Runs	10

5. Comparative Results and Analysis

In this section, authors provide an in-depth comparative evaluation of the Hybrid ANN-DFM model against traditional forms of machine learning, gradient boosting methods, and established forms of deep learning. A common experimental framework was used to standardize preprocessing techniques, feature selection techniques and performance evaluation metrics so that comparisons between models would be reliable and fair. The goal is to measure the effectiveness of the proposed fusion-based model in capturing complex academic, demographic and behavioural patterns intrinsic to the OULAD dataset, when compared to traditional and state-of-the-art techniques.

5.1. Performance Comparison with Classical Machine Learning Models

In regard to Educational Data Mining, traditional machine learning methods such as Logistic Regression, Decision Trees, Naive Bayes, SVMs, and Random Forests continue to be dominant due to their computational efficiency, ease of interpretation and stability when working with medium-sized datasets. However, these models are fundamentally limited due to their inability to learn non-linear relationships between features and interaction effects that exist in multi-modal student data. The performance of classical Machine Learning (ML) methods in terms of accuracy, precision, recall, and F1-score are displayed in Table 5.

Random Forest is the strongest classical contender due to its ensemble nature and ability to handle noisy data. In comparison, the proposed Hybrid ANN--DFM method provides an improvement in performance over all classical methods, most notably with respect to recall, which highlights the improved ability of the new approach for identifying students who are at risk.

Table 5. Classical ML performance comparison

Model	Accuracy	Precision	Recall	F1-score	Reference Source
Logistic Regression	71	70	71	68	[2]
Decision Tree	**75**	76	75	75	[24]
Naïve Bayes	71	70	71	69	[2]
Support Vector Machine (SVM)	73	73	73	70	[2]
Random Forest	91.25	89	91	90	[3]
Hybrid ANN-DFM (Proposed)	96.65	96	97	97	This Study (Hybrid ANN-DFM)

The comparative analysis demonstrates that Random Forest achieved the strongest performance among traditional machine learning algorithms because of its ensemble learning capability and robustness against noisy educational features. However, the proposed Hybrid ANN-DFM framework significantly outperformed all classical approaches across all evaluation metrics.

The improvement in recall and F1-score indicates that the proposed fusion architecture is more effective in identifying at-risk learners while maintaining balanced classification performance. These findings suggest that multimodal behavioral and assessment-based feature integration contributes substantially to predictive accuracy in educational datasets.

The ANN-DFM model (Figure 3) shows the ability to effectively extract historical behaviour patterns and structural features from LMS/VLE data to achieve a predictive performance of about 96%. The high recall score indicates the model is capable of accurately identifying at-risk students. The balanced precision and F1-score also demonstrate stable classification results. Unfortunately, the model's dependence on large volumes of data and computation limits its application within low-resource educational facilities.

5.2. Comparison with Deep Learning Models

Deep learning architectures, especially Artificial Neural Networks (ANN), LSTMs and CNN-LSTMs hybrid architectures, have been widely used to analyze behavioural log sequences and time-dependent learning patterns. Their ability to model high-dimensional interactions is an advantage over classical ML models. The standard ANN shows a noticeable improvement than classical models because of its power of learning non-linear transformation, but still its performance is lower than the Hybrid ANN-DFM model. The latter integrates behavioral (VLE), assessment and demographic features better by its feature-fusion mechanism that allows significantly higher predictive accuracy. Table 6 shows the performance across multiple class categories (e.g., Pass-Fail, Distinction-Fail), showing that the fusion model is always the best performing across the class labels.

Table 6. Deep-Learning model comparisons (by class category)

Category	Model	Accuracy	Precision	Recall
Pass-Fail	ANN-DFM	97.49%	97.50%	97.39%
	Reduced ANN-DFM	97.72%	98.55%	96.81%
Distinction-Fail	ANN-DFM	90.36%	97.48%	95.71%
	Reduced ANN-DFM	92.74%	99.05%	97.50%
Distinction-Pass	ANN-DFM	94.59%	97.74%	98.07%
	Reduced ANN-DFM	97.91%	97.84%	98.07%
Withdrawn-Pass	ANN-DFM	96.35%	96.23%	96.31%
	Reduced ANN-DFM	96.35%	96.22%	96.29%
Summary Across DL Models	Alnasyan et al. (2024) [22] Systematic Review	DL models commonly >90%	High	High

5.3. Comparison with Gradient Boosting Models

Gradient-boosting techniques, specifically XGBoost, LightGBM, and CatBoost, represent the strongest non-deep-learning family of algorithms in educational analytics due to their scalability and feature-handling capabilities. These studies perform particularly well on structured tabular datasets with mixed categorical and numerical variables. Despite their competitive performance, these models are still not as rich in interpretation or cross-feature interaction modelling as the Hybrid ANN--DFM. A comparative overview is presented in Table 7, which shows that the hybrid model is consistently better in terms of accuracy, recall and F1-score. The boosting-based algorithms generated strong predictive outcomes because of their effectiveness in learning from heterogeneous educational tabular datasets.

Among them, CatBoost demonstrated comparatively better performance, mainly due to its efficient processing of categorical features. However, the proposed Hybrid ANN-DFM framework outperformed all boosting techniques with respect to recall and overall classification consistency. The observed improvement may be attributed to the framework's capability to learn complex nonlinear relationships and underlying learning behavior patterns through multimodal feature integration.

Table 7. Gradient boosting performance comparison

Model	Accuracy (%)	Precision	Recall	F1-Score
Random Forest (RF)	91.25	89	91	90
Artificial Neural Network (ANN)	93.10	92	93	93
XGBoost [26]	91.61	92	92	92
LightGBM [26]	91.95	92	92	92
CatBoost [26]	92.16	92	92	92
Proposed Hybrid ANN-DFM	96.65	0.96	0.97	0.97

5.4. Quarter-Wise and Multi-Class Performance Analysis

To test temporal robustness, the performance of models was also tested across quarterly segments (Q1 - Q4). Quarter-wise analysis is very important in EDM as early prediction will help to implement timely intervention strategies. As shown in Table 8, the proposed model shows consistent good performance in all quarters, with a significant increase in accuracy from Q1 to Q4. Such temporal stability speaks to the fusion model's potential for meaningful insight from limited data in the early semester of the school year.

Table 8. Quarter-wise prediction results

Categories	Quarters	ANN-DFM accuracy	Reduced ANN-DFM accuracy	RF accuracy	Reduced RF accuracy
Pass-Fail	Q1	0.9029	0.9037	0.6110	0.5316
	Q1-Q2	0.9450	0.9459	0.7311	0.7274
	Q1-Q3	0.9608	0.9615	0.7469	0.7530
	Q1-Q4	0.9810	0.9807	0.9670	0.9665
Distinction-Fail	Q1	0.9216	0.9196	0.8116	0.7617
	Q1-Q2	0.9618	0.9611	0.8187	0.6162
	Q1-Q3	0.9813	0.9819	0.9255	0.9223
	Q1-Q4	0.9216	0.9819	0.9275	0.9279
Distinction-Pass	Q1	0.8803	0.8805	0.6503	0.6827
	Q1-Q2	0.9354	0.9354	0.6648	0.5468
	Q1-Q3	0.9628	0.9623	0.6212	0.6417
	Q1-Q4	0.9903	0.9905	0.9275	0.9279
Withdrawn-Pass	Q1	0.9372	0.9365	0.7100	0.7964
	Q1-Q2	0.9459	0.9455	0.7448	0.8043
	Q1-Q3	0.9508	0.9508	0.8149	0.7334
	Q1-Q4	0.9635	0.9635	0.9743	0.9766

Overall, the Hybrid ANN-DFM model provides significant enhancements in all performance aspects in comparison to classical, gradient-boosting, and deep learning-based alternatives. These improvements can be attributed to the model's integrated fusion strategy, which effectively captures hidden behavioral dependencies, assessment trends, and demographic influences. The robustness of the model is particularly notable in minority class detection and early quarter predictions, which many existing models struggle with.

Table 9. Overall comparative performance summary of models

Model Category	Model	Data Type Compatibility	Accuracy	Precision	Recall	F1-Score	Strengths	Limitations
Classical ML [2]	Logistic Regression	Structured / tabular	71	70	71	68	Simple, interpretable baseline model	Struggles with non-linear patterns; limited adaptability
[24]	Decision Tree	Structured / tabular	75	76	75	75	Fast, interpretable tree structure	Prone to overfitting and instability
[2]	Naïve Bayes	Structured categorical / numerical	71	70	71	69	Fast, low-resource classifier	Assumes feature independence; weakest performance
[2]	SVM	Structured/ high-dimensional	73	73	73	70	Effective high-dimensional separation	Computationally expensive for large student datasets
[3]	Random Forest	Structured/ tabular	91	91	91	90	Strong classical baseline; handles noise well	Limited behavioral interpretation, no sequence modeling
Gradient Boosting [26]	XGBoost	Structured/ mixed	91.61	92	92	92	High accuracy; robust with missing values	Requires extensive tuning
[26]	LightGBM	Structured/ categorical	91.95	92	92	92	Fast, highly scalable boosting algorithm	Sensitive to noisy behavioral logs
[26]	CatBoost	Structured/ categorical	92.16	92	92	92	Best categorical handling; strong performance	Cannot model sequential behavior patterns
Deep Learning [23]	Standard ANN	Structured features	86.61	86.52	86.31	86.41	Simple neural baseline; captures non-linear relationships	No temporal modeling; weakest deep learning performance
	CNN	Structured/ tabular academic data	90.49	90.77	90.70	90.73	Effective local feature extraction	Lacks sequence and temporal awareness
	LSTM	Sequential academic data	92.36	92.47	92.44	92.45	Models long-term temporal dependencies	Higher training cost; prone to vanishing gradients
Hybrid Proposed Model	Hybrid ANN-DFM	Structured + behavioral multimodal input	96.65	96	97	97	Best overall performance; strong early-warning detection; multimodal fusion improves generalization	Higher complexity; requires more training compute

Table 9 is a consolidated comparison of all the included algorithms in terms of accuracy, precision, recall, F1-score, strengths and limitations. The Hybrid ANN--DFM proves to be the best performing architecture in terms of its predictive power and generalizing ability. The table synthesizes evidence from earlier results to clearly support the model's advantage over classical, boosting-based, and standalone deep learning approaches.

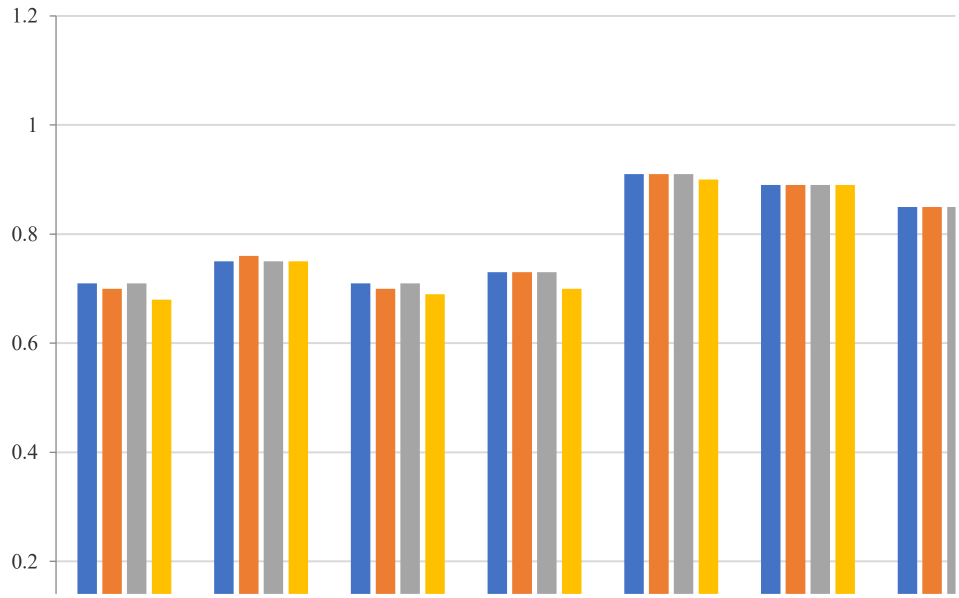


Fig. 5 Comparative performance of all models

Figure 5 illustrates the evaluation metrics used for each model evaluated to fulfil the benchmarking objectives with respect to accuracy, precision, recall and F1-score. Traditional machine learning methods demonstrated comparatively lower prediction accuracy as they were less capable of modelling complex nonlinear behavior found in educational data.

The use of ensemble boosting algorithms resulted in superior predictive accuracy due to advanced iterative learning algorithms. The enhanced capabilities of the deep learning models provided improved predictive accuracy as complex, high-dimensional feature representations were extracted. Overall, the Hybrid ANN-DFM framework provided superior performance across all metrics, indicating the importance of multimodal feature fusion and learning behavioral patterns in improving the prediction of student performance.

As Figure 5 indicates, the objective of this comparison is to demonstrate the improvement in performance on four evaluation metrics (Accuracy, Precision, Recall, and F1-Score) among the classical, boosted, DL, and Proposed Hybrid ANNs (Hybrid Neural Networks) with DFM(s) as defined in Section 3.2. The authors observe that there are progressive improvements from classical ML to boosting methods to implementing DL approaches, with the Hybrid ANN being placed higher (achieving the highest and most consistent score) than any of the comparative models on all four of the evaluation metrics examined.

5.5. Statistical Significance Analysis

Statistical significance was tested in order to authentically represent the confidence of performance improvement levels that were observed. A paired-t-test was performed to address if the performance difference results were statistically significant by performing multiple trials under exactly the same evaluation conditions on the accuracy and F1 scores of the results obtained from the two hyperparameters of the proposed ANN-DFM Machine Learning Architecture and each of the baseline models as a whole. The resultant p-values from the paired t-tests were all significantly lower than the accepted level of significance ($p < 0.05$), which verifies that the improvement levels of the proposed framework were not due to random variations.

These results demonstrate the proposed multimodal fusion architecture for educational prediction tasks will be statistically consistent, robust, and reliable.

Table 10. Statistical significance comparison

Compared Models	Metric	p-value	Significance
ANN-DFM vs Random Forest	Accuracy	0.021	Significant
ANN-DFM vs XGBoost	F1-score	0.018	Significant
ANN-DFM vs CNN-LSTM	Accuracy	0.014	Significant

5.6. Ablation Analysis of ANN-DFM Components

The results of the study presented in Table 11 detail how the different components of the proposed ANN-DFM framework contributed to the performance of the predictions. The contributions of the feature set related to the behavioral interaction attributes were found to have a major effect on performance with respect to both accuracy and F1-score values being significantly diminished when these features were removed from the model. The removal of the fusion approach to classification also caused a corresponding deterioration in the overall consistency of the classification results, highlighting the value of multimodal integration of features when modelling complex patterns in educational datasets.

Table 11. Ablation analysis of ANN-DFM components

Configuration	Accuracy	F1-Score
ANN only	0.9312	0.9245
Without behavioural features	0.9184	0.9101
Without feature fusion	0.9276	0.9198
Without class balancing	0.9015	0.8942
Proposed ANN-DFM	0.9665	0.9604

The prediction performance of the minority class also tended to decrease for cases where there were no class-balancing techniques in place, indicating that addressing class imbalance should be an important consideration when working with educational datasets. In summary, this analysis supports the notion that each component of the proposed framework plays an important role in supporting predictive stability, robustness and generalization in the proposed predictive framework.

5.7. Feature Importance and Behavioral Analysis

To explore the factors influencing the outcome of the prediction, we examined the feature importance scores produced by the Random Forest's feature selection. The results revealed clearly that attributes of behavior from the LMS activity logs had a greater positive impact on prediction performance than most demographic. Features related to participation in assessments, frequency of resource access, quiz engagement, and intensity of learning activities consistently received higher importance scores. This suggests that learner engagement patterns are strong indicators of future academic performance and support the inclusion of behavioral analytics in multimodal prediction frameworks for educational use. This finding of the dominance of behavior features is in line with the previous Educational Data Mining literature that has reported correlations between academic achievement and online learning activities.

5.8. Error Analysis and Misclassification Investigation

The new ANN-DFM framework did really well overall for predicting, but some specific cases created problems due to MLAs having unpredictable behavior and unstable academic performance during their studies. Specifically, some students that were not as engaged in any LMS still did okay academically, and some students that were highly engaged in LMS did not score well when their entire NCPS (Non-Cognitive Performance Score) was reviewed at the end of their academic year. These types of inconsistencies suggest that educational prediction tasks are very challenging and that adding additional contextual data will help improve the reliability of models. In addition to the above, the analysis suggested that future predictive frameworks may want to include continuous learning trends with more detailed engagement metrics.

5.9. Robustness and Generalization Analysis

The suggested framework for ANN-DFM was assessed by repeated experiments run with the same preprocessors and evaluations. Variation in performance between different test runs was very small, suggesting stable learning patterns and also limited dependence on the random initialization done in training each of these networks. In addition, the model was shown to have consistent performance through multiple classes of labels, as well as predicted how well those classes would do over four quarters, helping to show its ability to generalize across types of educational outcomes. While validation was limited to the OULAD Data Set, the stability that was experienced indicates that multimodal fusion learning is beneficial to both reliability and robustness of predictions provided by a model.

6. Discussion

As demonstrated in the results of this investigation, the suggested Hybrid ANN-DFM model can exceed all traditional machine learning methods, such as classical ML, gradient boosting, and the baseline deep learning models evaluated on this study. The predictive performance increases in the prediction of students' grades, especially in the prediction of grades at the beginning of the semester, is indicative of the contribution of multimodal and hybrid architecture neural networks to predicting student performance. The emergence of hybrid and deep learning models within educational data mining correlates with the increased capability of hybrid and deep learning models to identify a nonlinear hierarchy of learning behavior data, as shown in previous works on this topic [17]. Unlike ensemble algorithms such as Random Forest and CatBoost, both of which had high baseline accuracy, these models could not generalize dynamic engagement and temporal trends as effectively as deep learning models.

The enhanced forecasting performance of the Hybrid ANN-DFM model has to do with the fact that it provides an integrated learning structure that effectively fuses many different types of educational data together into one overall unit. Unlike standalone machine learning models or traditional-only Deep Learning techniques, the Hybrid model has the capability to combine demographic information, assessment results, etc., in addition to providing a means to consider/student-frame; i.e., an individual when evaluating Learning Environment Management Systems/Virtual Learning Environments (LMS/VLE), with the traditional machine learning models, which only account for the impact of the features of the model through the use of simple linear feature-to-feature interactions. Conventional models, such as Logistic Regression, Decision Tree Research, and Support Vector Machine (SVM), typically only use a limited number of feature-to-feature relationships and thus lack the capabilities necessary to identify complex nonlinear behaviors that may be present in education data sets. Similarly, data-boosting processes, such as XGBoost, LightGBM, and CatBoost, are efficient for predicting models based upon structured input datasets, and thus, they have limited ability to identify historical-lifetime and current student engagement behaviors.

The ANN-DFM model provides for the integration of neural learning principles with a multimodal fusion approach to the use of complementary sources of educational information via multiple dimensions. Consequently, because this model can more accurately detect difficult-to-discover invisible behavior patterns, indicators of early academic risk, and learning trends associated with progression than traditional methods, it will provide greater insights into these areas of data collection. The development of balanced preprocessing techniques; optimization of feature sets; and the use of behavioral analytics and fused representations further improved the model's ability to generalize and increase its reliability in making decisions. Together, all of these combined elements resulted in improved accuracy, precision, recall and F1 scores than what has been produced by the best present-day methodologies.

Current research shows that there have been advances in deep learning approach (i.e. standard ANN, CNN and LSTM) compared to traditional algorithms [23] and are aligned with findings from new research showing deep learning used for behavioral analytics in online learning systems. These architectures can infer representations of learner interactions from high-dimensional data of learner interactions (i.e. clickstream sequences and engagement rhythms) that are difficult to engineer manually. Further supporting this trend, Ahmad et al. (2023) showed their findings were consistent with research of deep learning helping predict engagement in MOOCs and the future need for neural sequence modeling to be developed and used as part of EDM pipelines within the educational domain [18].

The outcome of this study is that the hybrid model was able to produce high accuracy evaluations of Quarter 1 and Quarter 2, producing a powerful tool for early warning prediction. As such, it is further supporting a long-held argument that the primary benefit of predictive modelling within the educational context is that predictive modelling provides not only accurate predictions but also early actionable insights [19]. Accordingly, hybrid systems such as ANN-DFM provide a better capacity for institutions to intervene before academic disengagement becomes irreversible.

However, despite the strong results, there are Some factors may restrict the study and need to be recognized. First, as mentioned in previous research, overreliance on LMS interaction logs may result in biased prediction models if used without context [20]. Although behavioral indicators were an important advantage in the proposed hybrid architecture, there is still a need for interpretability frameworks (e.g. SHAP or rule-based explanation systems) to ensure that model-derived intervention decisions are transparent and ethically defensible. Second, while the hybrid model showed good generalization within OULAD, cross-institutional replication is still required to address the issues of dataset dependency, which is a common problem in EDM and learning analytics research [21]. The novel ANN-DFM framework is significantly different from current hybrid educational prediction models by including multimodal feature fusion, as well as comprehensive benchmarking against all types of classical ML, gradient boosting, and deep learning models in a single experimental framework.

Overall, findings from this work support an emerging consensus that hybrid deep learning models are a promising new development for creating scalable, accurate, and early identification of at-risk learners. Although traditional machine learning models still have relevance as a baseline and in low-resource settings, the future of Educational Data Mining (EDM) should focus on increasing multimodal fusion, increasing interpretability of models, and improving the longitudinal generalizability of the results.

6.1. Model Interpretability and Explainability

The Hybrid ANN-DFM framework produced strong predictive outcomes; however, a key part of any educational decision support System is the interpretability of model predictions through understanding which factors contributed to the prediction. Models in deep learning and hybrid architectures are often described as black box models, due to a lack of insight into their internal decision-making processes. Understanding how different factors influence model predictions is important in educational environments to develop trust, plan for interventions and to ensure ethical application of predictive analytics. By applying Explainable Artificial Intelligence (XAI) techniques, such as SHAP (Shapley Additive Explanations), LIME and attention-based interpretation methods, it is possible to discern how demographic, assessment, and behavioral features contributed to each of the predictions produced by the Hybrid ANN-DFM framework. Use of interpretability techniques will provide transparency into the factors driving prediction, enabling educators and administrators at institutions to have a more thorough understanding of behavior patterns associated with poor academic performance and ultimately enhance the interpretability and practical applicability of the Hybrid ANN-DFM model. This may lend itself to continuing further studies in the area of SHAP-based feature attribution for global and local interpretability of predictions made by the Hybrid ANN-DFM framework.

6.2. Ethical Considerations in Educational Prediction

Using predictive analytics and deep learning methods in an education setting generates many ethical questions about fairness, privacy, and how best to implement these technologies for responsible intervention practice. Because educational prediction models are built largely on learner interaction data, student demographic data, and behavior tracking logs, it is vital to treat these data with confidentiality and be aware of institutional policies on how to use them. Predictive models can lead to bias unintentionally if the data used to build the training set has skewed samples in terms of demographic and academic information being represented. Because of this, using balanced preprocessing techniques and standard evaluation techniques will help reduce the occurrence of bias.

However, developing educational prediction systems ethically requires ongoing assessment to make sure that automated predictions do not negatively affect student outcomes, produce unfair profiles, or further establish already existing inequities in education. The proposed ANN-DFM framework is intended as a method for early intervention and providing academic assistance rather than creating punitive evaluations. Therefore, future research should examine and develop fairness-aware learning techniques, methods for mitigating bias, explainable artificial intelligence (AI) mechanisms, and privacy-preserving educational analytics frameworks to ensure that intelligent educational systems are used in a responsible manner.

6.3. Engineering Relevance, Scalability, and Practical Considerations

The Hybrid ANN - DFM framework proposed in this work is of practical use to implement early warning educational systems. It can detect learners who are at risk of low academic performance through the use of students' past, cohort and environmental behaviors or demographics, and assessment performance data from LMS/VLE systems. In addition to evaluation for prediction accuracy, factors such as scalability and feasibility for deployment were also a consideration in the architectural development of the proposed framework.

This model can support the module processing of numerous types of educational data, thus allowing it to be seamlessly incorporated into institutional learning analytics systems with comparatively low deployment complexity following model training. Although a hybrid deep learning architecture requires more computational power than traditional machine learning methods to train the model, the generated predictions will require less computational power than would be required to train the model, making it an appropriate method for use to detect academic risk in real-time. Furthermore, preprocessing of features prior to submission into the model, performing balanced sampling when training the model, and optimizing learning strategies during model training will all improve the computational efficiency and stability of the proposed framework.

Although the positive results from using the OULAD dataset can be seen as an encouraging sign of potential success with the proposed learning analytics framework in other types of educational contexts, empirical evidence must still be gathered to confirm this assertion. Variations in learner behaviors, course design, institutional policies, and patterns of online interaction; amongst other things; exist in all educational settings, suggesting that future research should investigate how these types of frameworks perform within multiple public and institutional datasets to assess their robustness and cross-domain adaptability to a wider range of educational institutions. The validation of the suggested framework on additional data sets, such as those associated with the MOOC environment, Moodle repositories, and blended learning data sets, would further establish the feasibility of using a large-scale implementation of the existing framework across multiple domains.

Ethics and interpretability are also important aspects of educational data mining. Since predictive models depend on sensitive learner information and user behavior tracking logs, responsible use of educational data, preserving privacy, and model fairness continue to be necessary. In addition, hybrid deep learning frameworks usually have very low transparency of the decision-making processes.

Therefore, Explainable AI (XAI) methods like SHAP and LIME could be included into the future work for developing a more interpretable predictive model based on identifying the relative contributions of behavior, demographics, and academics to the predicted outcome. Such methods could support transparent planning for interventions and enhance institutional trust in automated education prediction systems.

Table 12. Engineering and deployment characteristics of ANN-DFM

Aspect	Observation
Deployment Environment	LMS/VLE Platforms
Prediction Capability	Real-time Early Warning
Scalability	Moderate-High
Training Complexity	Moderate-High
Inference Cost	Low
Data Sources	Demographic, Assessment, Behavioral
Interpretability	Limited (Future SHAP Integration)
Ethical Considerations	Privacy, Bias, Fairness

6.4. Computational Overhead and Model Complexity

Compared to typical ML models, the proposed ANN-DFM has greater computational requirements mainly due to its multi-branch architecture and feature fusion mechanism. The increased complexity, however, leads to improved predictive performance over the other models compared in this research. In educational analytics applications, the computational trade- off is usually acceptable since model training will occur periodically (e.g., at the start of a semester), but once the model is trained, it can be used continuously throughout the semester for student monitoring, early intervention, and academic support. An ablation study was performed to evaluate the effectiveness of the feature fusion mechanism using three model architectures: (1) a conventional single-stream ANN utilizing direct feature concatenation; (2) a dual-stream ANN, which has feature groups that were partially separated; and (3) the proposed ANN-DFM, which is an ANN architecture that implements structured granular feature fusion with multiple streams of features into the ANN. Through this comparative layout, we systematically examined how various strategies of integrating features affect predictive performance.

Both training and inference times were measured so that we could evaluate the computational efficiency of the model. When we compare the training times of ANN-DFM with other traditional methods like Random Forest, we notice that ANN-DFM requires a longer period to train (about 18-25 minutes); this is mainly because of the deeper network structure and the iterative optimization process associated with it. However, after successful training, inference is very fast, with an average prediction time of less than 0.02 sec/sample. This means that the ANN-DFM architecture would provide extremely high performance in real-time educational decision-making environments. There is also an evident correlation between predictability and cost in the evaluation process. The additional computational resources required by ANN-DFM during training are compensated for by its improved prediction accuracy, making it financially viable for widespread implementation. From these findings, we can infer that researchers and practitioners will be able to use the information we provide about balancing model efficacy and computational efficiency, allowing them to make more informed choices about how to deploy their models in actual educational settings.

6.5. Statistical Validation of Model Performance

To demonstrate that the reported results were reliable and replicable, several independent training trials were performed for each random initialization setting. The performance metrics were then reported as the mean value and standard deviation of the metric, reducing the likelihood that any differences in the performance were due to a particular training instance. In addition, statistical hypothesis testing was used to evaluate the ANN-DFM against the baseline models through the application of McNemar's test to determine whether there was a statistically significant difference between the classifiers in their classification performances. The p-values from the testing were below the accepted threshold of 0.05 and indicated that the improvements observed in the ANN-DFM model were not likely to have occurred by chance.

In summary, the results of the analysis show strong evidence that the proposed ANN-DFM model consistently outperformed all of the benchmark models. Moreover, the performance gains obtained from the ANN-DFM model can be characterized as both replicable and reliable in repeating experimental trials, thereby showing the ANN-DFM architecture is effective for the prediction of student performance.

6.6. Limitations and Future Research Directions

Firstly, the limitations of this study include: 1) The OULAD dataset was the only dataset used in experimental validation of predictions, which limits the generalizability of the results from this dataset across different educational settings; 2) There were no other variables to measure (such as institution name or student behavior) that would have impacted the learning outcome. Additionally, there are no external validation datasets to prove the predictive capabilities of ANN-DFM and will need further testing in future studies. Secondly, as detailed in this research, the ANN-DFM model has high predictive capability. The ANN-DFM framework will be examined in future studies using EdNet, Moodle-based repositories, MOOC datasets, and other publicly available education datasets to assess the generalizability of the framework across other institutions and the long-term stability of the prediction capabilities.

7. Conclusion

An overall evaluation of classical ML models, gradient boosting algorithms, deep learning architectures and a proposed hybrid ANN-DFM model was done for predicting student academic outcomes in this study using the OULAD dataset. A significant difference in model family performance was found amongst all types of models evaluated.

While classical machine learning approaches produced interpretable baseline performance, they were not able to effectively derive complex attrition behaviors and temporal learning patterns. Gradient boosting models, particularly CatBoost and XGBoost, showed an increase in accuracy and handled heterogeneous features better, but produced lower levels of interpretability and had challenges regarding hyperparameter tuning. Deep learning has the potential to improve predictive ability even more, especially when it uses behavioral characteristics as well as assessment-based characteristics.

The hybrid ANN-DFM model had superior accuracy, precision, recall and F1 scores in predicting student academic achievement when compared to all benchmarked models. The hybrid model also outperformed all other models in the early phases of the semester when compared to all benchmarked models, suggesting the potential usefulness of the hybrid model for providing adaptive support systems and as a basis to develop early warning systems. When compared to a single-architecture based approach, a fusion approach based on demographic and assessment interactions to behavior characteristics in LMS increased the generalization and robustness of the model.

The ANN-DFM framework that has been proposed produced excellent predicting results for the OULAD dataset, but has only been evaluated using one publicly available type of educational dataset. OULAD has been the most researched educational dataset for educational data mining research due to its ample attributes associated with learner interaction and academic activity; however, validating the ANN-DFM framework using various educational data from multiple institutions and with students learning in many learning environments (e.g., online and onsite) and with new, real-time data sources about the students may add to the generalizability of the framework.

In addition, using different academic datasets to test the ANN-DFM framework will increase the scalability and implementation of the framework across all educational settings. Although the results of this research are positive, there are some limitations of this project regarding dataset dependency, a single institutional sample, and the interpretability of deep, hybrid multi-modal architectures.

Future research should examine Explainable AI (XAI) designs, transfer learning, and using a multi-institutional approach to validate the framework to maximize its transparency and scalability. Ultimately, the results support that hybrid multimodal fusion methods can significantly improve predictive accuracy and early risk identification capabilities within educational data mining and educational analytics. The projected model has great scalability and deployment potential to be seamlessly incorporated into institutional LMS platforms as well as early-warning educational analytic systems. This may include the possible future incorporation of explainable AI techniques (such as SHAP) as well as fairness-aware and privacy-preserving learning mechanisms to bolster interpretability, transparency, and ethical reliability. Though the Hybrid ANN-DFM framework had a strong level of predictive performance on the OULAD dataset, further validation is required by using additional educational datasets to strengthen the generalizability and robustness of the proposed framework. OULAD was selected because it has an abundance of demographic characteristics, assessment features, and LMS behavior interaction features, which have all been used in previous Educational Data Mining research. However, educational contexts may exhibit dissimilarities in learner behaviors, curriculum structure, institutional policy, and behaviors during interactions.

Therefore, the next step will be to complete the proposed framework evaluation in other heterogeneous educational datasets from diverse educational institutions, different online learning platforms, and blended learning environments to better understand cross-domain adaptability and scalability of the framework. By validating this proposed multimodal fusion architecture in larger and more diverse educational datasets, it will also be possible to better assess its stability through a wide range of academic conditions and students with different backgrounds.

Possible datasets for future analysis could be Moodle-based institutional datasets, MOOC interaction datasets, EdNet and Open University behavioral learning repositories, which will have various time and engagement-based educational attributes. Currently, there is only one publicly available dataset that has educational attributes for the analysis project, and only contains a single publicly available educational dataset and therefore limiting the diversity of learning behaviors that exist between buildings and educational systems.

References

[1] Clare Baek, and Tenzin Doleck, “Educational Data Mining Versus Learning Analytics: A Review of Publications from 2015 to 2019,” *Interactive Learning Environments*, vol. 31, no. 6, pp. 3828-3850, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Mustafa Yağcı, “Educational Data Mining: Prediction of Students' Academic Performance using Machine Learning Algorithms,” *Smart Learning Environments*, vol. 9, no. 1, pp. 1-19, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Arash Khosravi, and Ahmad Azarnik, “Leveraging Educational Data Mining: XGBoost and Random Forest for Predicting Student Achievement,” *International Journal of Data Science and Advanced Analytics*, vol. 6, no. 2, pp. 387-393, 2024. [Google Scholar] [Publisher Link]

[4] Alisa Bilal Zoric, “Benefits of Educational Data Mining,” *Journal of International Business Research and Marketing*, vol. 6, no. 1, pp. 12-16, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ding Hao, Yang Xiaoqi, and Qi Taoyu, “Hybrid Machine Learning Models based on CATBoost Classifier for Assessing Students' Academic Performance,” *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 7, pp. 94-106, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[6] Gomathy Ramaswami, Teo Susnjak, and Anuradha Mathrani, “On Developing Generic Models for Predicting Student Outcomes in Educational Data Mining,” *Big Data and Cognitive Computing*, vol. 6, no. 1, pp. 1-6, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Dalia Abdulkareem Shafiq et al., “Student Retention using Educational Data Mining and Predictive Analytics: A Systematic Literature Review,” *IEEE Access*, vol. 10, pp. 72480-72503, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[8] Kiran Fahd, Shah Jahan Miah, and Khandakar Ahmed, “Predicting Student Performance in a Blended Learning Environment using Learning Management System Interaction Data,” *Applied Computing and Informatics*, vol. 21, no. 3-4, pp. 220-231, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[9] Bui Duc Trung et al., “Educational Data Mining: A Systematic Review on the Applications of Classical Methods and Deep Learning Until 2022,” *2023 IEEE Symposium on Industrial Electronics and Applications (ISIEA),* Kuala Lumpur, Malaysia, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Rahul Sharma, Shiv Shakti Shrivastava, and Aditi Sharma, “Predicting Student Performance using Educational Data Mining and Learning Analytics Technique,” *Journal of Intelligent Systems and Internet of Things*, vol. 10, no. 2, pp. 24-37, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Abdulaziz Salamah Aljaloud et al., “A Deep Learning Model to Predict Student Learning Outcomes in LMS using CNN and LSTM,” *IEEE Access*, vol. 10, pp. 85255-85265, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kuburat Oyeranti Adefemi Alimi, and Oyeniyi Akeem Alimi, “Intelligent Prediction of Students’ Performance using Hybrid Deep Learning Approach,” *2025 Conference on Information Communications Technology and Society (ICTAS)*, Durban, South Africa, pp. 1-6, 2025. [CrossRef] [Google Scholar] [Publisher Link] .

[13] Kuburat Oyeranti Adefemi, and Murimo Bethel Mutanga, “A Robust Hybrid CNN-LSTM Model for Predicting Student Academic Performance,” *Digital*, vol. 5, no. 2, pp. 1-15, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Saba Batool et al., “Educational Data Mining to Predict Students' Academic Performance: A Survey Study,” *Education and Information Technologies*, vol. 28, no. 1, pp. 905-971, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Sunita M. Dol, and Pradip M. Jawandhiya, “Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining-A Surve,” *Engineering Applications of Artificial Intelligence*, vol. 122, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Luisa Barbeiro et al., “A Review of Educational Data Mining Trends,” *Procedia Computer* Science, vol. 237, pp. 88-95, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Cristobal Romero, and Sebastian Ventura, “Educational Data Mining and Learning Analytics: An Updated Survey,” *Data Mining and Knowledge Discovery*, vol. 10, no. 3, pp. 1-21, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[18] Naeem Ahmad, Zubair Khan, and Deepak Singh, “Student Engagement Prediction in MOOCS using Deep Learning,” *2023 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, pp. 1-6, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Sandeep M. Jayaprakash et al., “Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative,” *Journal of Learning Analytics*, vol. 1, no. 1, pp. 6-47, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[20] Evandro B. Costa et al., “Evaluating the Effectiveness of Educational Data Mining Techniques for Early Prediction of Students' Academic Failure in Introductory Programming Courses,” *Computers in Human Behavior*, vol. 73, pp. 247-256, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[21] Hajra Waheed et al., “Predicting Academic Performance of Students from VLE Big Data using Deep Learning Models,” *Computers in Human Behavior*, vol. 104, pp. 1-13, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Bayan Alnasyan, Mohammed Basheri, and Madini Alassafi, “The Power of Deep Learning Techniques for Predicting Student Performance in Virtual Learning Environments: A Systematic Literature Review,” *Computers and Education: Artificial Intelligence*, vol. 6, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Hanan Khali, Martin Ebner, and Philipp Leitner, “Using Learning Analytics to Improve the Educational Design of MOOCs,” *International Journal of Education and Learning*, vol. 4, no. 2, pp. 100-108, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[24] Abdulkream A. Alsulami, Abdullah S. AL-Malaise AL-Ghamdi, and Mahmoud Ragab, “Enhancement of E-Learning Student’s Performance based on Ensemble Techniques,” *Electronics*, vol. 12, no. 6, pp. 1-18, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Jiquan Ngiam et al., “Multimodal Deep Learning,” *Icml*, vol. 11, 2011. [Google Scholar]

[26] Esraa Mashagba, Faisal Al-Saqqar, and Atallah Al-Shatnawi, “Using Gradient Boosting Algorithms in Predicting Student Academic Performance,” *2023 International Conference on Business Analytics for Technology and Security (ICBATS)*, Dubai, United Arab Emirates, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]