Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Artificial Intelligence Applications in Academic Administration: A Multi-Level Decision Support Framework for Higher Education

George Karraz
Damascus University
Anas Shahin
ons.hassan.shahin@gmail.com
Syrian Virtual University

Article

Keywords:

Posted Date: June 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9792709/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Artificial Intelligence Applications in Academic Administration: A Multi-Level Decision Support Framework for Higher Education

George Karraz[1] **and Anas Shahin**[2,*]

[1]Faculty of Information Technology Engineering, Damascus University, Damascus, Syria
[2]Faculty of Information Technology Engineering, Syrian Virtual University, Damascus, Syria
[*]ons.hassan.shahin@gmail.com

ABSTRACT

Higher education institutions are increasingly challenged by the complexity of academic administration, the expansion of student data, and the demand for transparent, evidence-based decision-making. Traditional administrative approaches often struggle to manage large-scale information and ensure fairness in resource allocation, academic planning, and quality assurance. This paper proposes a multi-level framework for employing artificial intelligence (AI) to support decision-making in academic administration. The proposed system integrates three analytical layers: student-level predictive modeling for at-risk identification and grade forecasting, course-level engagement and performance analytics using learning management system data, and institution-level strategic benchmarking and performance scoring. Using three publicly available datasets—the UCI Student Performance Dataset (1,044 records), the Open University Learning Analytics Dataset (32,593 registrations with 10.6 million interaction logs), and the U.S. College Scorecard (6,500+ institutions)—we train and evaluate Logistic/Linear Regression, Random Forest, XGBoost, and Neural Network models across classification and regression tasks. Three novel composite indices are engineered: an Engagement Index, a Teaching Load Index, and an Institutional Performance Score. Results demonstrate that XGBoost achieves the highest performance in course-level binary classification (F1 = 0.940, accuracy = 94.0%), Random Forest leads in student-level grade prediction (R^2 = 0.894) and institutional performance forecasting (R^2 = 0.788), and K-Means clustering reveals four distinct institutional segments. SHAP-based explainability analysis identifies assessment completion, engagement patterns, and socioeconomic indicators as the dominant predictive factors across the respective layers. An early warning system prototype demonstrates practical utility by stratifying students into three risk tiers with validated actual pass rates of 2.3%, 50.7%, and 92.9%, respectively. This research contributes an integrated, explainable, and reproducible framework that bridges the gap in multi-level academic decision support systems.

Introduction

The rapid expansion of artificial intelligence (AI) in higher education has transformed the way institutions manage academic processes, allocate resources, and support student success. Recent advancements in machine learning, predictive analytics, and data-driven decision-making have enabled universities to adopt more efficient and transparent administrative practices[1–5]. These technologies provide new opportunities for enhancing academic planning, improving student performance forecasting, and supporting institutional governance.

Despite the global progress in AI adoption, many higher education institutions—especially in developing regions—continue to rely on traditional administrative models that lack predictive capabilities and depend heavily on manual decision-making. This gap highlights the need for intelligent systems capable of integrating diverse data sources, analyzing complex patterns, and generating actionable insights for academic leaders[6–10].

Existing literature on AI in education has largely focused on individual analytical levels in isolation: student-level performance prediction[11–13], course-level learning analytics[14,15], or institution-level benchmarking[16]. However, few studies have attempted to integrate these three levels into a unified decision support framework that provides comprehensive, cross-level insights for academic administrators.

This research proposes an integrated AI-driven academic decision support system that leverages multi-level data, including student performance indicators, course engagement patterns, and institutional metrics. By combining machine learning models with explainable AI (XAI) techniques, the study aims to enhance transparency, improve planning efficiency, and support evidence-based decision-making across different levels of academic administration.

The proposed system is designed to operate across three analytical layers: (1) student-level predictive modeling for at-risk identification, (2) course-level engagement and performance analytics, and (3) institution-level strategic planning and

benchmarking. This multi-layered approach ensures that academic decisions are informed by comprehensive, data-driven insights that reflect the complexity of modern higher education environments.

The contributions of this paper are as follows:

1. A **multi-level framework** integrating student, course, and institution analytics into a unified decision support system.
2. Three **novel engineered composite indices**—the Engagement Index, Teaching Load Index, and Institutional Performance Score—designed for multi-level academic analytics.
3. **Comprehensive empirical evaluation** comparing four machine learning algorithms across three public datasets and multiple predictive tasks.
4. **SHAP-based explainability analysis** at each analytical layer, supporting transparent and accountable academic governance.
5. A **validated early warning system prototype** demonstrating the practical utility of the framework for student intervention.

The remainder of this paper is organized as follows: Section reviews related work. Section describes the methodology, including data sources, feature engineering, and model training. Section presents the experimental results. Section discusses the findings and their implications. Section concludes the paper.

Literature Review

AI in Educational Administration

Artificial intelligence has become a central component in modern educational systems, supporting both instructional and administrative processes. Recent studies have shown that AI-driven tools can enhance institutional efficiency, improve student performance monitoring, and support data-informed decision-making across various academic contexts[1–5]. These advancements include predictive analytics for student success, automated assessment systems, and intelligent resource allocation frameworks.

Explainable AI in Academic Decision-Making

Several researchers have emphasized the importance of integrating explainable AI (XAI) into academic decision-making to ensure transparency and trust in automated systems. Explainable models help academic administrators understand the rationale behind predictions, thereby improving the adoption and reliability of AI-based solutions[6]. This is particularly relevant in higher education, where decisions often carry significant academic and ethical implications. Lundberg and Lee[20] introduced SHAP (SHapley Additive exPlanations), which has since become the standard for model-agnostic feature attribution and has been applied in educational contexts to interpret at-risk prediction models.

Predictive Models in Educational Data Mining

Recent systematic reviews have highlighted the growing role of predictive models in educational data mining (EDM). These models have been used to forecast student performance, identify at-risk learners, and analyze behavioral patterns within learning environments[7,13,17]. Machine learning techniques such as Random Forest, XGBoost, and Neural Networks have demonstrated superior performance compared to traditional statistical approaches, especially when dealing with complex, non-linear relationships in academic datasets. Chen and Guestrin[19] demonstrated the effectiveness of gradient boosting in structured data tasks, while Breiman[18] established the theoretical foundations of Random Forest for classification and regression.

Learning Analytics and Student Engagement

The Open University Learning Analytics Dataset (OULAD)[12] has become a benchmark for course-level engagement analysis. Studies using OULAD have shown that clickstream data, assessment timeliness, and interaction diversity are strong predictors of student outcomes. Gašević et al.[14] demonstrated that learning analytics can provide actionable insights for course design, while Viberg et al.[15] identified challenges in scaling these approaches across institutions.

Governance and Institutional Analytics

Governance frameworks for AI in higher education have gained increasing attention. Studies have examined issues related to data privacy, algorithmic fairness, and institutional accountability, emphasizing the need for responsible AI adoption in academic settings[9]. The U.S. College Scorecard[16] provides a comprehensive dataset for institutional benchmarking, enabling analyses of graduation rates, financial indicators, and student outcomes across thousands of institutions.

Research Gap

Recent reviews of AI applications in higher education institutions have identified several challenges, including limited data availability, inconsistent digital infrastructure, and the need for localized models tailored to regional educational systems[10]. The existing literature collectively demonstrates the potential of AI to transform academic administration. However, it reveals a significant gap in integrated, multi-level decision support systems that combine student-level, course-level, and institution-level analytics with explainable AI—an area that this research aims to address.

Methodology

This study adopts a multi-layered methodological framework designed to integrate student-level, course-level, and institution-level analytics into a unified academic decision support system. The methodology consists of five main components: data sources, data preprocessing, feature engineering, model training and validation, and explainability analysis.

Data Sources

Three publicly available datasets were selected to support the three analytical layers of the proposed framework.

UCI Student Performance Dataset

This dataset[11] contains records from two Portuguese secondary schools, covering Mathematics (395 students) and Portuguese language (649 students) courses, totaling 1,044 student records with 33 variables each. Features include demographic attributes (age, sex, address), family background (parental education, family size), behavioral indicators (study time, absences, alcohol consumption), and three grade periods (G1, G2, G3), where G3 serves as the final prediction target. The dataset contains no missing values.

Open University Learning Analytics Dataset (OULAD)

OULAD[12] provides 32,593 student registrations across 7 relational tables, including demographics, assessments, course structures, and detailed clickstream interactions with the Virtual Learning Environment (VLE). The VLE interaction table contains over 10.6 million records, capturing daily click counts across 20 resource types. Student outcomes are classified into four categories: Distinction, Pass, Fail, and Withdrawn. The dataset contains minimal missing values (1,111 in the `imd_band` field).

U.S. College Scorecard Dataset

This dataset[16] contains institution-level indicators for over 6,500 currently operating U.S. higher education institutions. Variables include enrollment, graduation rates, retention rates, tuition, financial aid distributions, faculty salaries, student debt, and post-graduation earnings. From the available 1,731 columns, 36 key variables were selected based on relevance to institutional performance assessment and data completeness.

Data Preprocessing

Data preprocessing involved cleaning missing values, normalizing numerical variables, encoding categorical attributes, and merging multi-table structures. For the UCI dataset, binary categorical variables were mapped to 0/1 encodings, yes/no fields were binarized, and multi-class categoricals (e.g., parental job, school reason) were one-hot encoded, resulting in 48 features. For OULAD, the seven relational tables were joined on shared keys (`id_student`, `code_module`, `code_presentation`), and clickstream data was aggregated into per-student engagement features. For the Scorecard, the `PrivacySuppressed` string entries were converted to numeric (resulting in NaN), and the dataset was filtered to currently operating institutions (CURROPER = 1) and the most recent reporting year.

System Architecture

The proposed system architecture (Fig. 1) consists of three analytical layers:

1. **Student-Level Analytics**: Predicting academic performance, identifying at-risk students, and modeling behavioral and demographic factors.

2. **Course-Level Analytics**: Engagement monitoring using VLE interactions, assessment performance analysis, and course difficulty evaluation.

3. **Institution-Level Analytics**: Resource allocation modeling, KPI forecasting, and institutional benchmarking through clustering and regression.

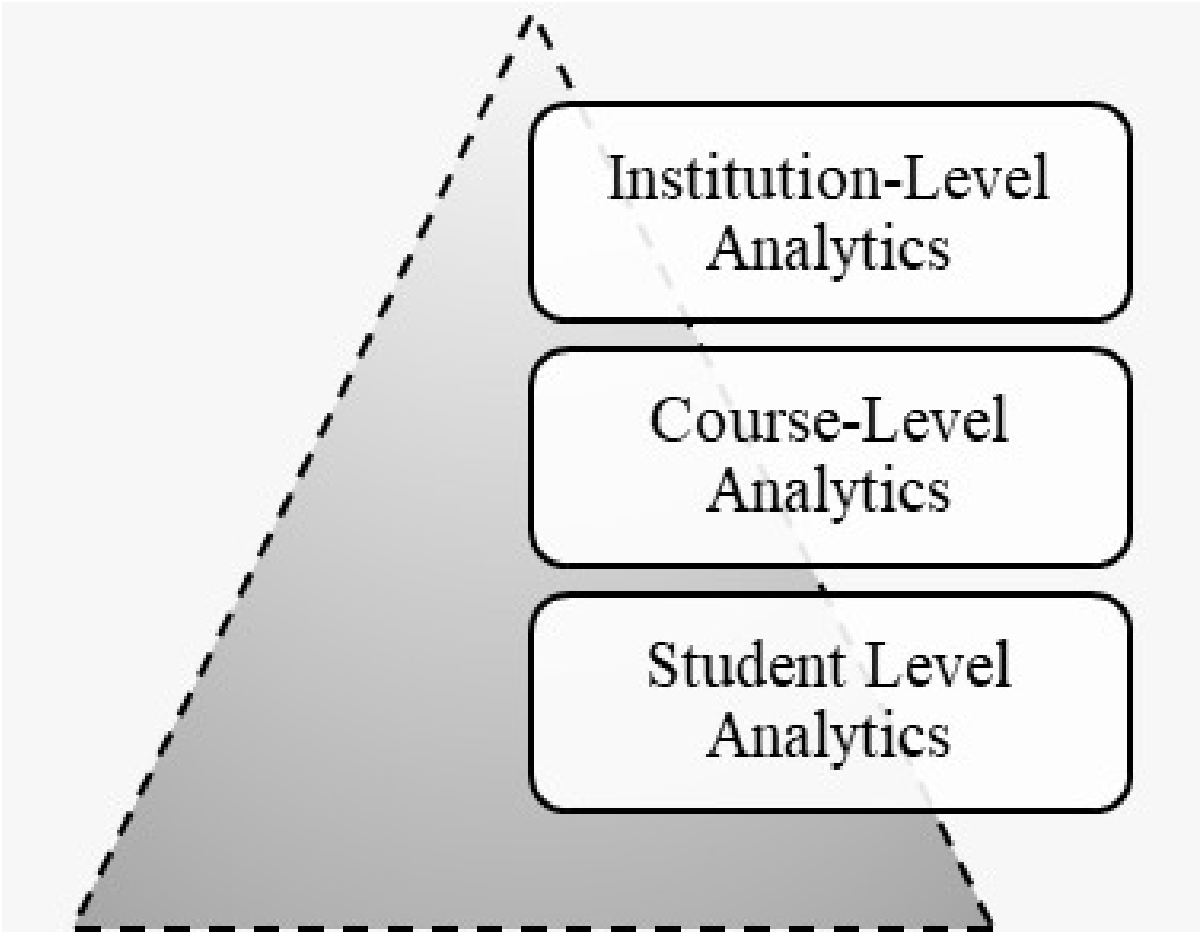


Figure 1. Proposed multi-level decision support system architecture. The three analytical layers feed into a unified decision interface for academic administrators.

Feature Engineering

Three composite indices were engineered to support multi-level decision-making, as described below.

Engagement Index (EI)

Computed for each student in OULAD by combining normalized clickstream and assessment data:

$$EI = 0.5 \cdot \hat{C} + 0.3 \cdot \hat{D} + 0.2 \cdot \hat{S} \tag{1}$$

where $\hat{C}$ is the min-max normalized total VLE clicks, $\hat{D}$ is the normalized count of active VLE days, and $\hat{S}$ is the normalized mean assessment score. The weights reflect the relative importance of interaction volume (50%), consistency of engagement (30%), and academic performance (20%).

Teaching Load Index (TLI)

Computed at the course level to capture instructional complexity:

$$TLI = 0.4 \cdot \hat{N}_s + 0.3 \cdot \hat{N}_a + 0.3 \cdot \hat{L} \tag{2}$$

where $\hat{N}_s$ is the normalized number of enrolled students, $\hat{N}_a$ is the normalized number of assessments, and $\hat{L}$ is the normalized module presentation length.

Institutional Performance Score (IPS)

Computed for each institution in the Scorecard dataset:

$$IPS = 0.35 \cdot \hat{G} + 0.25 \cdot \hat{R} + 0.25 \cdot \hat{P} - 0.15 \cdot \hat{D}_r \tag{3}$$

where $\hat{G}$ is the normalized 4-year graduation rate, $\hat{R}$ is the normalized full-time retention rate, $\hat{P}$ is the normalized 3-year loan repayment rate, and $\hat{D}_r$ is the normalized 3-year cohort default rate. The negative weight on default rate penalizes financial underperformance. The resulting score is rescaled to $[0,1]$.

Non-Linearity of Input–Output Relationships

The relationship between academic input variables and predicted outcomes is rarely linear. Academic systems exhibit complex, multi-dimensional interactions involving non-linear effects and interaction terms. For example, the impact of study hours on performance may increase up to a threshold and then plateau, while the influence of attendance may vary depending on prior achievement. Exploratory analysis confirmed these non-linear patterns (see Section), justifying the use of advanced machine learning models capable of learning non-linear decision boundaries[18,19].

Mathematical Formulation

The general predictive modeling framework is expressed as:

$$Y = f(\mathbf{X}) + \varepsilon \tag{4}$$

where Y represents the target variable (e.g., final grade, pass/fail status, institutional KPI), $\mathbf{X}$ denotes the set of input features, $f(\mathbf{X})$ is the learned predictive function, and ε is the error term. This formulation does not assume linearity, allowing the model to learn complex relationships directly from the data.

Model Training and Validation

Four machine learning algorithms were trained and evaluated at each analytical layer:

1. **Logistic/Linear Regression** (Baseline): Provides a linear benchmark for comparison.
2. **Random Forest**[18]: An ensemble of decision trees using bagging, trained with 200 estimators.
3. **XGBoost**[19]: Gradient-boosted decision trees with 200 estimators and a learning rate of 0.1.
4. **Neural Network (MLP)**: A multi-layer perceptron with architecture varying by layer: (64, 32) for student-level, (128, 64, 32) for course-level, and (64, 32) for institution-level, using early stopping.

All models were evaluated using an 80/20 train-test split with stratification (for classification tasks) and a fixed random seed ($s = 42$) for reproducibility. Classification tasks were evaluated using accuracy, precision, recall, and F1-score (weighted). Regression tasks were evaluated using RMSE, MAE, and R^2.

For the UCI dataset, two feature sets were compared: **Set A** (all features including prior grades G1 and G2) and **Set B** (early prediction without G1 and G2), simulating a scenario where the model must predict outcomes before any academic grades are available.

Explainability Analysis

SHAP (SHapley Additive exPlanations)[20] was applied to the best-performing tree-based models at each layer using TreeExplainer. SHAP values provide both global feature importance rankings and individual prediction explanations, supporting transparent and accountable academic governance as recommended by recent XAI frameworks[6].

Ethical and Governance Considerations

The methodology incorporates ethical guidelines related to data privacy, fairness, and transparency. All datasets used are publicly available and de-identified. Explainable AI techniques were used to ensure that model predictions can be interpreted by non-technical stakeholders, supporting responsible deployment in academic settings.

Results

Layer 1: Student-Level Analytics (UCI)

Classification Results

Table 1 presents the pass/fail classification results for both feature sets. With all features (Set A), Logistic Regression achieved the highest F1-score of 0.915, slightly outperforming XGBoost (0.906) and Random Forest (0.895). This indicates that the strong linear relationship between prior grades (G1, G2) and outcomes makes even simple models highly effective.

In the early prediction scenario (Set B, without G1 and G2), all models experienced a significant performance decrease, with XGBoost achieving the best F1-score of 0.766. This 16% drop in F1 quantifies the predictive value of prior academic grades and highlights the challenge of early intervention based solely on behavioral and demographic features.

Table 1. Layer 1 classification results: Pass/Fail prediction on UCI dataset.

Set	Model	Acc.	Prec.	Rec.	F1
Set A	LR (Baseline)	0.914	0.915	0.914	**0.915**
	Random Forest	0.895	0.895	0.895	0.895
	XGBoost	0.904	0.908	0.904	0.906
	Neural Network	0.866	0.859	0.866	0.858
Set B	LR (Baseline)	0.770	0.733	0.770	0.742
	Random Forest	0.794	0.763	0.794	0.751
	XGBoost	0.789	0.762	0.789	**0.766**
	Neural Network	0.789	0.765	0.789	0.720

Regression Results

Table 2 shows the grade prediction results. With all features, Random Forest achieved the lowest RMSE (1.254) and highest R^2 (0.894), demonstrating the model's ability to capture non-linear grade dependencies. Without prior grades (Set B), all models performed poorly ($R^2 \leq 0.216$), confirming that behavioral features alone have limited predictive power for exact grade estimation. Fig. 2 summarises the performance gap between feature sets across both classification and regression tasks.

Table 2. Layer 1 regression results: G3 (final grade) prediction on UCI dataset.

Set	Model	RMSE	MAE	R^2
Set A	LR (Baseline)	1.569	0.978	0.834
	Random Forest	**1.254**	**0.855**	**0.894**
	XGBoost	1.369	0.948	0.874
	Neural Network	1.947	1.360	0.745
Set B	LR (Baseline)	3.593	2.623	0.130
	Random Forest	3.413	2.478	0.215
	XGBoost	**3.409**	**2.472**	**0.216**
	Neural Network	3.634	2.625	0.110

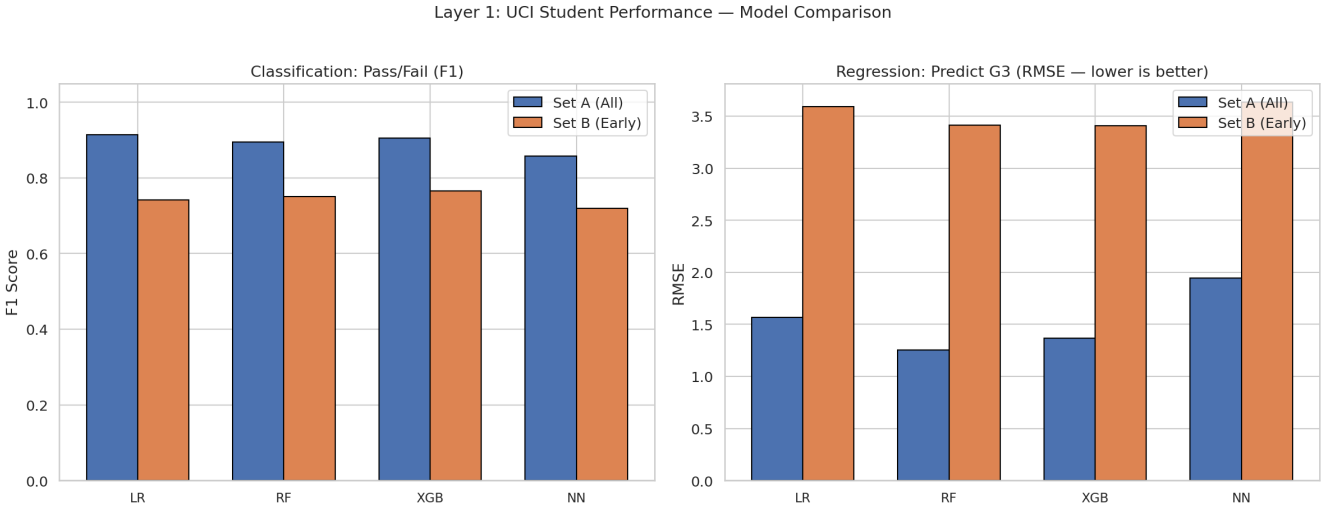


Figure 2. Layer 1 model comparison: Classification F1 (left) and Regression RMSE (right) for both feature sets. Set A includes prior grades; Set B simulates early prediction.

Layer 2: Course-Level Analytics (OULAD)

Binary Classification

Table 3 presents the binary classification results (Pass/Distinction vs. Fail/Withdrawn). XGBoost achieved the highest performance with an F1-score of 0.940 and accuracy of 94.0%, followed closely by Neural Network (0.930) and Random Forest (0.929). Even the Logistic Regression baseline achieved 89.0% accuracy, indicating that the engineered engagement features are inherently discriminative.

Table 3. Layer 2 binary classification results on OULAD.

Model	Acc.	Prec.	Rec.	F1
LR (Baseline)	0.890	0.892	0.890	0.891
Random Forest	0.929	0.930	0.929	0.929
XGBoost	**0.940**	**0.940**	**0.940**	**0.940**
Neural Network	0.930	0.931	0.930	0.930

Four-Class Classification

Table 4 shows the four-class results (Distinction, Pass, Fail, Withdrawn). XGBoost again achieved the highest F1 of 0.741 and accuracy of 74.9%. The confusion matrix (Fig. 3) reveals that the primary source of misclassification is between adjacent outcome categories: 479 Fail students were predicted as Withdrawn (both exhibiting low engagement), and 291 Distinction students were predicted as Pass (both achieving positive outcomes). Fig. 4 compares binary and four-class F1 scores across all models evaluated at this layer.

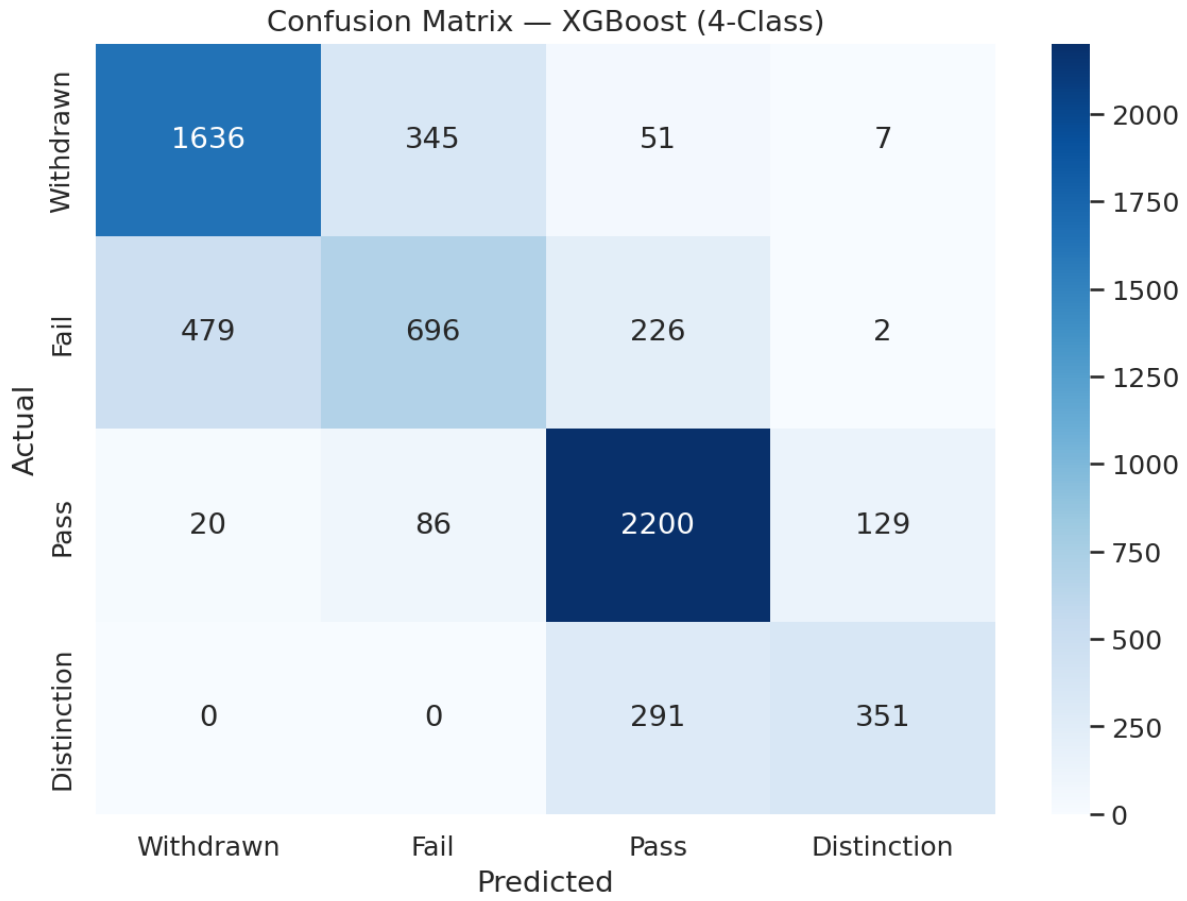


Figure 3. Confusion matrix for XGBoost four-class classification on OULAD. Misclassifications predominantly occur between adjacent outcome categories.

Table 4. Layer 2 four-class classification results on OULAD.

Model	Acc.	Prec.	Rec.	F1
LR (Baseline)	0.668	0.646	0.668	0.645
Random Forest	0.726	0.715	0.726	0.711
XGBoost	**0.749**	**0.740**	**0.749**	**0.741**
Neural Network	0.713	0.700	0.713	0.703

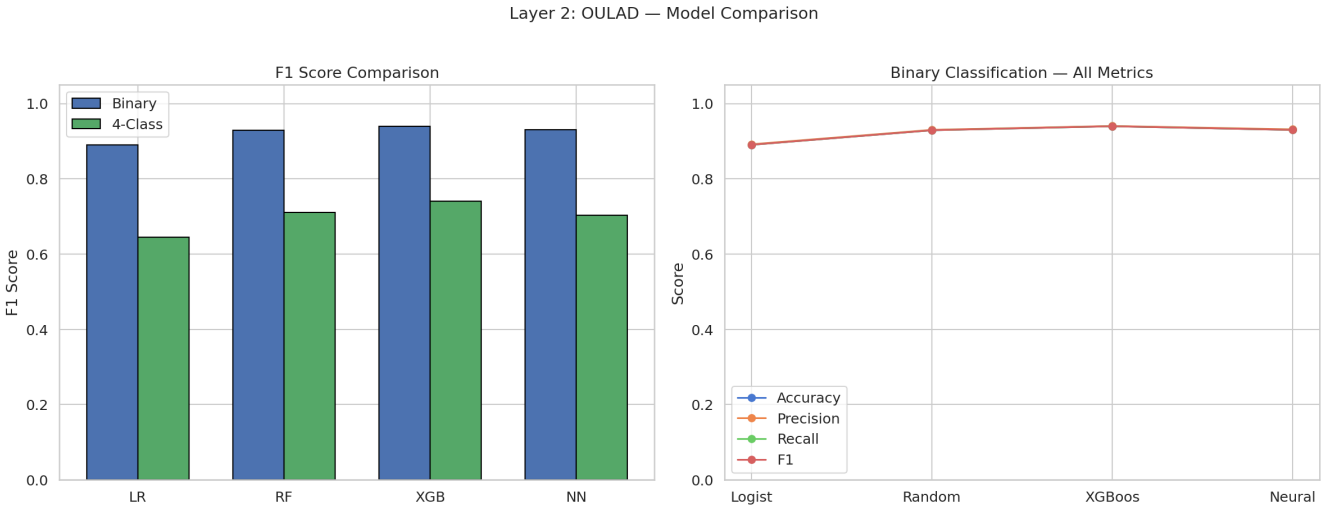


Figure 4. Layer 2 model comparison: Binary vs. 4-class F1 scores (left) and binary classification metrics across models (right).

Layer 3: Institution-Level Analytics (Scorecard)

IPS Regression

Table 5 presents the Institutional Performance Score prediction results. Random Forest achieved the highest R^2 of 0.788, followed by XGBoost (0.776). All models significantly outperformed the linear baseline ($R^2 = 0.706$), confirming non-linear relationships between institutional characteristics and performance outcomes.

Table 5. Layer 3 regression results: Institutional Performance Score prediction.

Model	RMSE	MAE	R^2
LR (Baseline)	0.109	0.081	0.706
Random Forest	**0.093**	**0.067**	**0.788**
XGBoost	0.095	0.069	0.776
Neural Network	0.105	0.080	0.727

Institutional Clustering

K-Means clustering with $K = 4$ (determined by the elbow method) identified four distinct institutional segments, as shown in Table 6 and Fig. 5. The overall IPS distribution and breakdown by institution type are shown in Fig. 6.

Table 6. Institutional cluster profiles ($K = 4$).

	Cluster			
Metric	0	1	2	3
Label	Estab.	Develop.	Needs Imp.	Elite
Mean IPS	0.66	0.60	0.37	**0.84**
Enrollment	1,883	10,183	1,781	4,357
Tuition ($)	27,562	8,245	14,093	41,008
Faculty Sal. ($)	6,614	7,738	4,681	10,062
Med. Debt ($)	19,706	15,330	14,299	20,646
N institutions	605	601	542	208

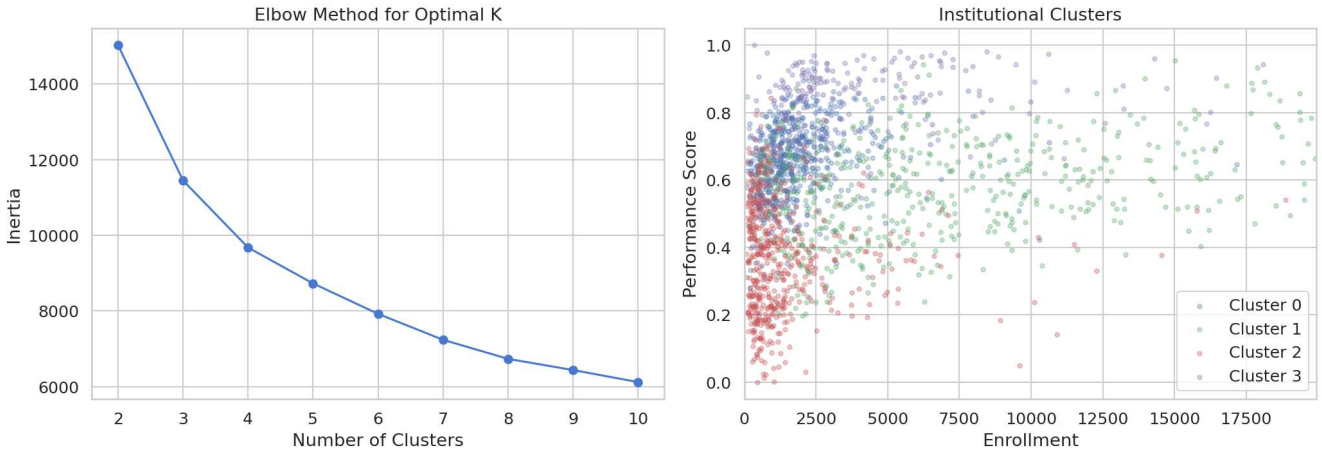


Figure 5. Institutional clustering: Elbow method for K selection (left) and cluster visualization by enrollment vs. IPS (right).

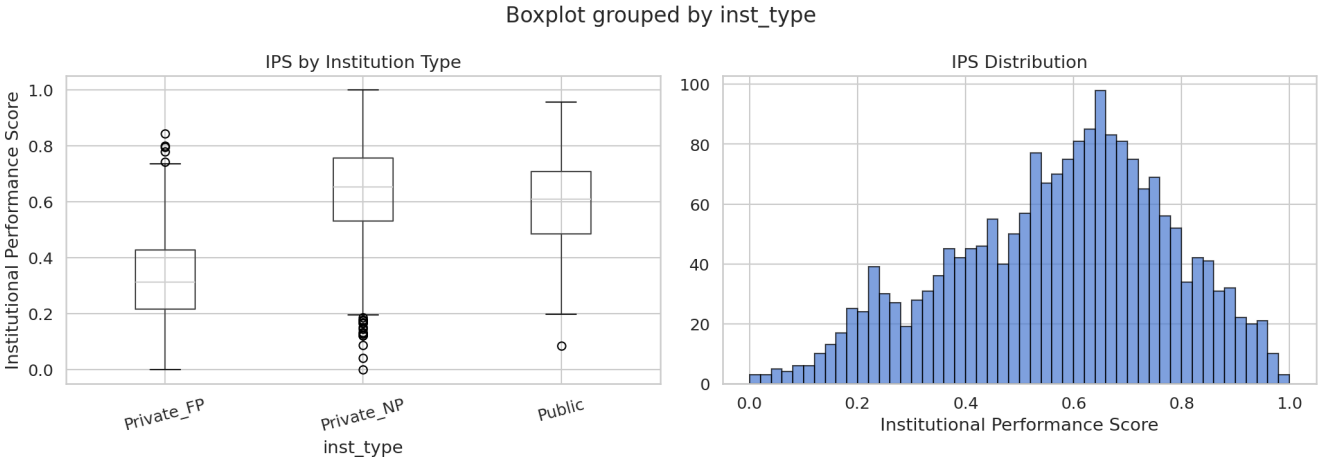


Figure 6. Institutional Performance Score distribution by institution type (left) and overall histogram (right).

Explainability Analysis (SHAP)

Fig. 8 presents the cross-layer SHAP feature importance comparison, revealing that fundamentally different factors drive predictions at each analytical level.

Layer 1 (Student-Level): The number of prior academic failures was the dominant predictor, followed by subject (Portuguese vs. Math), aspiration for higher education, school attended, and absences. Parental education (Medu, Fedu) contributed moderately, while social and lifestyle factors (going out, alcohol consumption) had smaller but non-negligible effects.

Layer 2 (Course-Level): The number of completed assessments was the strongest predictor, followed by the Engagement Index (validating the engineered feature), active VLE days, exam scores, and total clicks. The Teaching Load Index ranked 11th, indicating meaningful but secondary contribution from course-level structural factors. The SHAP summary plot for Layer 2 (Fig. 8) confirms that high engagement feature values consistently push predictions toward a passing outcome.

Layer 3 (Institution-Level): The percentage of Pell Grant recipients—a proxy for student body socioeconomic composition—dominated institutional performance prediction by a wide margin. Out-of-state tuition, median debt, and instructional expenditure per FTE followed, while enrollment and in-state tuition had minimal predictive impact.

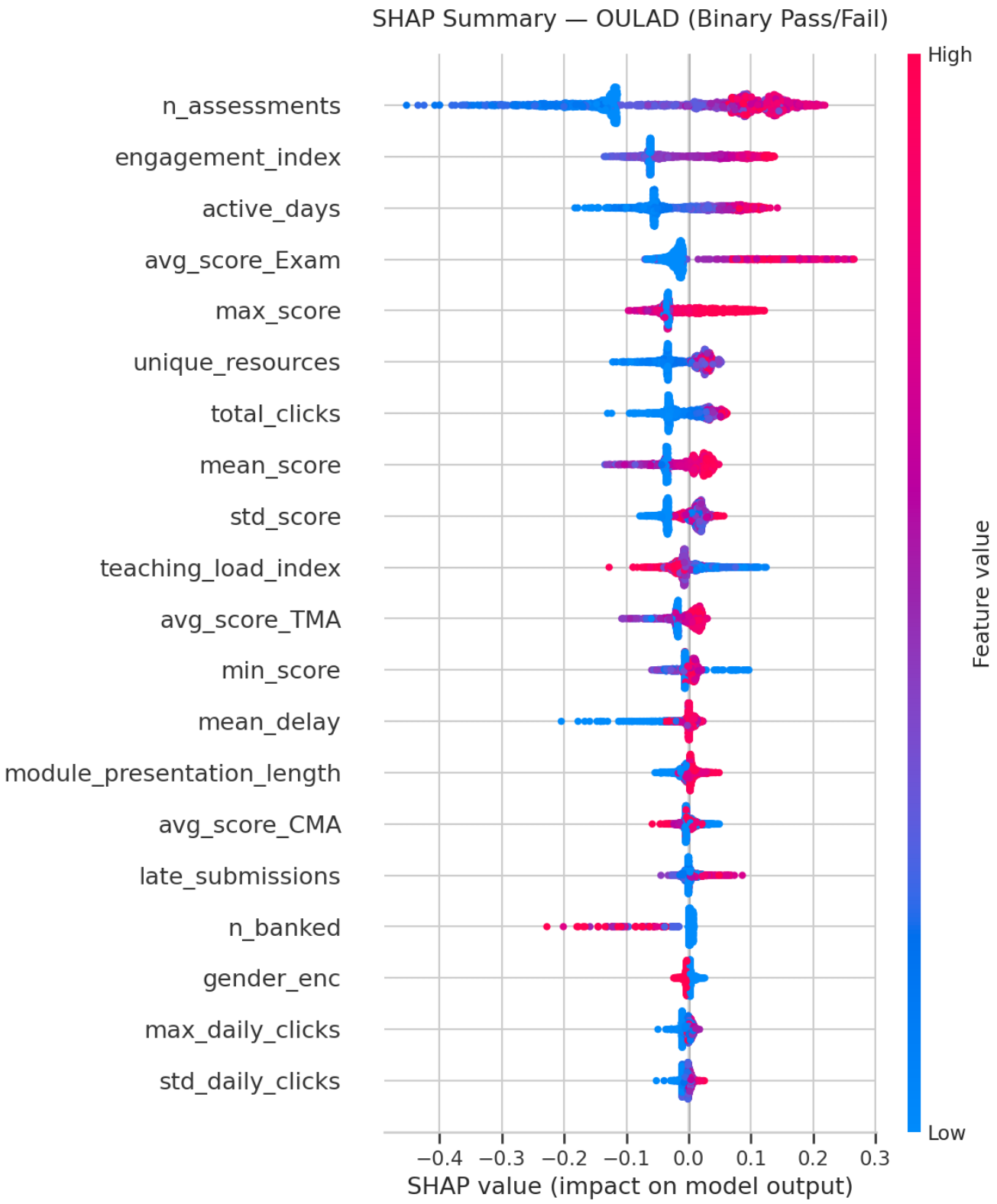


Figure 8. SHAP summary plot for Layer 2 (OULAD binary classification). High values of engagement features push predictions toward passing.

Engineered Features Validation

Fig. 9 validates the three engineered composite indices. The Engagement Index shows clear monotonic discrimination across all four outcome classes: Distinction ($\mu = 0.34$), Pass ($\mu = 0.25$), Fail ($\mu = 0.18$), and Withdrawn ($\mu = 0.12$). The Teaching Load Index varies meaningfully across modules (range: 0.25–0.55), capturing differences in instructional complexity. The Institutional Performance Score segments institutions from "Needs Improvement" (mean IPS = 0.37) to "Elite Performers" (mean IPS = 0.84).

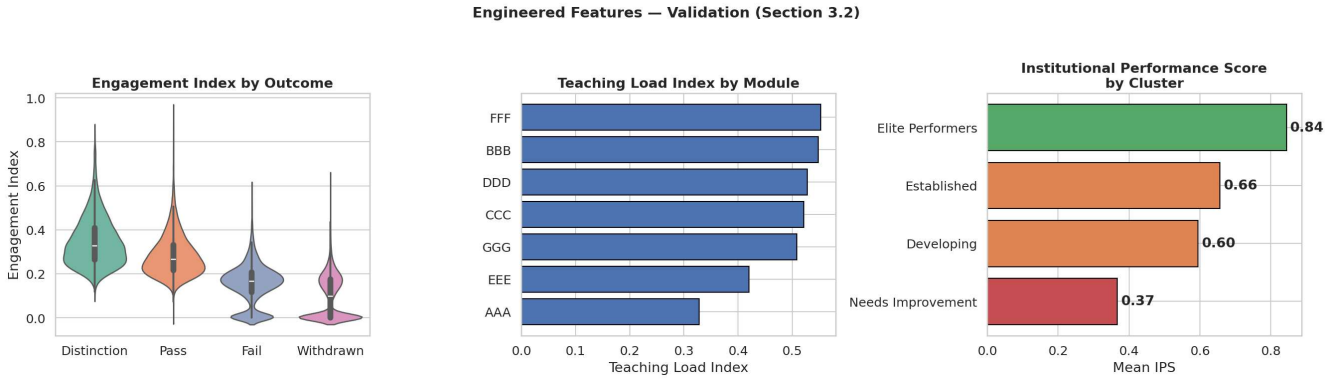


Figure 9. Validation of the three engineered composite indices: Engagement Index by outcome (left), Teaching Load Index by module (center), and IPS by institutional cluster (right).

Cross-Layer Decision Support

Fig. 10 demonstrates the cross-layer decision support framework in operation. The early warning system stratifies 6,519 test students into three risk tiers based on predicted pass probability:

- **HIGH RISK** ($p < 0.3$): 3,195 students with an actual pass rate of 2.3%.
- **MODERATE RISK** ($0.3 \leq p < 0.6$): 205 students with an actual pass rate of 50.7%.
- **LOW RISK** ($p \geq 0.6$): 3,119 students with an actual pass rate of 92.9%.

The strong separation between actual outcomes across risk tiers validates the system's practical utility for targeted intervention.

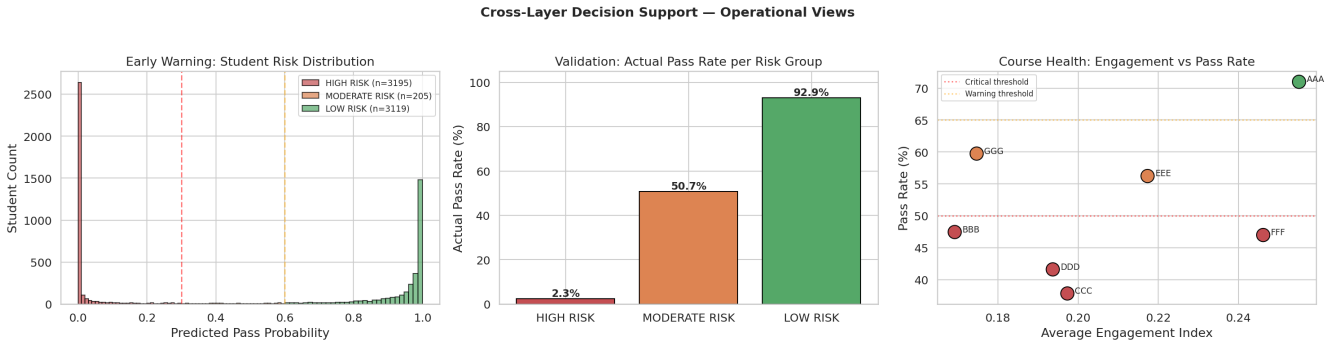


Figure 10. Cross-layer decision support: Student risk distribution (left), validated actual pass rates per risk tier (center), and course health dashboard showing engagement vs. pass rate (right).

Fig. 11 provides a consolidated performance comparison across all layers, tasks, and models evaluated in this study.

Discussion

Key Findings

The results demonstrate that the proposed multi-level framework achieves strong predictive performance across all three analytical layers. Several findings merit discussion.

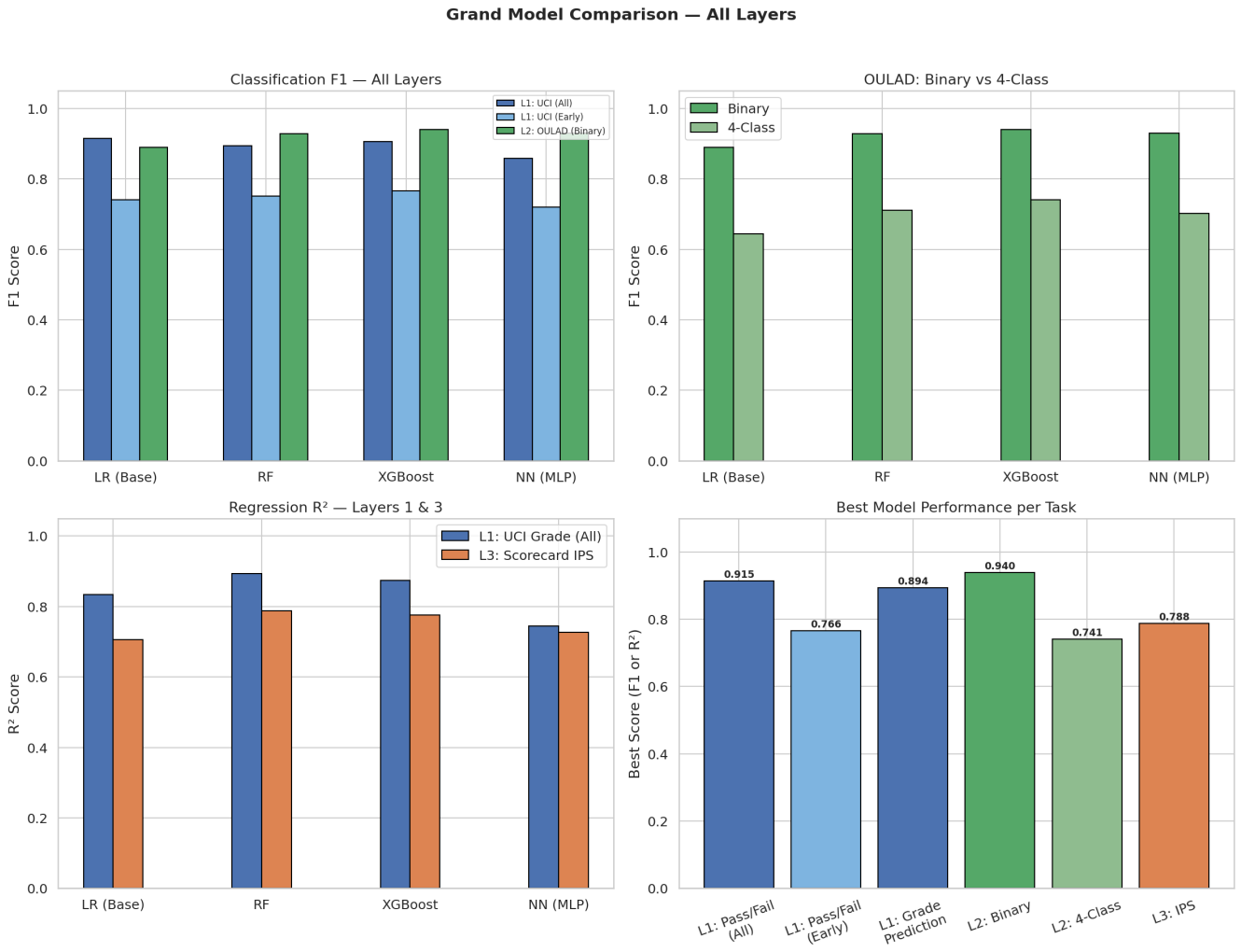


Figure 11. Grand model comparison across all layers and tasks.

XGBoost consistently outperforms alternatives in classification tasks. At the course level, XGBoost achieved 94.0% binary accuracy and 74.9% four-class accuracy, outperforming Random Forest, Neural Networks, and the linear baseline. This is consistent with the broader machine learning literature showing XGBoost's effectiveness on structured tabular data[19]. However, at the student level with all features, Logistic Regression achieved the highest F1 (0.915), suggesting that when strong linear predictors (G1, G2) are available, model complexity offers diminishing returns.

Random Forest excels in regression tasks. For both grade prediction ($R^2 = 0.894$) and IPS forecasting ($R^2 = 0.788$), Random Forest outperformed XGBoost and Neural Networks. The ensemble's ability to capture non-linear relationships while resisting overfitting through bagging makes it well-suited for continuous outcome prediction in educational datasets.

Early prediction without prior grades is challenging but valuable. The drop from F1 = 0.915 (Set A) to F1 = 0.766 (Set B) quantifies the "grade dependency" in student performance prediction. This finding underscores the importance of the course-level engagement features provided by Layer 2, which can detect at-risk students before formal grades are available through clickstream patterns and assessment completion rates.

Engagement features are the most predictive at the course level. The SHAP analysis reveals that assessment completion count, the engineered Engagement Index, active VLE days, and exam scores are the top predictors of student outcomes. This validates the design of the Engagement Index and supports the argument that learning management system data provides actionable early warning signals.

Socioeconomic composition predicts institutional performance. At the institution level, the percentage of Pell Grant recipients dominates the IPS prediction model. This finding has significant policy implications, suggesting that institutional performance metrics are strongly influenced by the socioeconomic background of the student body rather than purely by institutional resources or spending.

Practical Implications

The framework provides several actionable capabilities for academic administrators:

Early intervention: The validated early warning system enables administrators to identify at-risk students with 97.7% certainty in the HIGH RISK tier, allowing targeted allocation of support resources such as tutoring, counseling, and academic advising.

Course quality assessment: The course health dashboard identifies modules with pass rates below critical thresholds (e.g., modules BBB and CCC at 47% and 37%, respectively), enabling curriculum review and pedagogical intervention.

Institutional benchmarking: The clustering analysis allows institutions to identify peer groups and benchmark against similar institutions, supporting strategic planning and resource allocation.

Transparent governance: SHAP-based explanations provide administrators with interpretable rationales for each prediction, supporting ethical decision-making and regulatory compliance.

Limitations

Several limitations should be acknowledged. First, the UCI dataset originates from Portuguese secondary schools rather than higher education institutions, which limits direct generalizability. Second, the three datasets are not longitudinally linked; the cross-layer integration is architectural rather than based on a single institutional dataset. Third, the feature engineering weights (e.g., in the Engagement Index) were set based on domain knowledge rather than optimized empirically. Fourth, the Scorecard dataset exhibits substantial missingness, which restricts the analysis to institutions with complete records. Future work should address these limitations through longitudinal data collection from partner institutions and weight optimization through cross-validation.

Comparison with Related Work

The proposed framework advances the state of the art in several ways. Compared to single-layer approaches[7,13], our multi-level architecture provides administrators with a comprehensive view across student, course, and institutional levels. Compared to prior OULAD analyses[12], our engineered Engagement Index achieves SHAP-validated discriminative power as the second most important feature. The integration of SHAP explainability across all three layers addresses calls for transparent AI in educational governance[6,9].

Conclusion

This paper presented a multi-level AI-driven decision support framework for academic administration, integrating student-level, course-level, and institution-level analytics into a unified system. Using three publicly available datasets totaling over 1 million data points, we trained and evaluated four machine learning algorithms across classification, regression, and clustering tasks.

The key contributions are: (1) a validated multi-level framework achieving up to 94.0% accuracy in student outcome prediction, $R^2 = 0.894$ in grade forecasting, and $R^2 = 0.788$ in institutional performance prediction; (2) three novel engineered

indices (Engagement Index, Teaching Load Index, Institutional Performance Score) shown to be discriminative and explainable; (3) SHAP-based cross-layer analysis revealing that behavioral, engagement, and socioeconomic factors dominate predictions at the student, course, and institution levels, respectively; and (4) a validated early warning system prototype stratifying students into actionable risk tiers.

Future work will focus on deploying the framework with partner institutions in the Arab region to collect integrated, multi-level longitudinal data; optimizing index weights through cross-validation; and developing an interactive dashboard interface for non-technical administrators. The integration of natural language processing for qualitative feedback analysis and reinforcement learning for adaptive intervention strategies represent additional promising research directions.

Acknowledgements

Not applicable.

Author contributions statement

G.K. conceived the study, developed the methodology, provided supervision, and contributed to writing, review and editing. A.S. developed the software, performed formal analysis and investigation, created all visualizations, and wrote the original draft. All authors reviewed the manuscript.

Additional information

Funding: No funding was received for this research.
Competing interests: The authors declare no competing interests.
Data availability: All datasets analysed during this study are publicly accessible. The UCI Student Performance Dataset is available at https://archive.ics.uci.edu/dataset/320/student+performance[11]. The Open University Learning Analytics Dataset is available at https://doi.org/10.1038/sdata.2017.171[12]. The U.S. College Scorecard is available at https://collegescorecard.ed.gov/data/[16].

References

1. Chen, L., Chen, P. & Lin, Z. Artificial intelligence in education: A review. *IEEE Access* **8**, 75264–75278 (2020).
2. Zawacki-Richter, O., Marín, V. I., Bond, M. & Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education. *Int. J. Educ. Technol. High. Educ.* **16**, 39 (2019).
3. Pedro, F., Subosa, M., Rivas, A. & Valverde, P. *Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development*. UNESCO Working Papers on Education Policy No. 7 (2019).
4. Baker, R. S. & Inventado, P. S. Educational data mining and learning analytics. In *Learning Analytics* 61–75 (Springer, 2014).
5. Romero, C. & Ventura, S. Educational data mining and learning analytics: An updated survey. *WIREs Data Min. Knowl. Discov.* **10**, e1355 (2020).
6. Arrieta, A. B. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020).
7. Hellas, A. et al. Predicting academic performance: A systematic literature review. In *Proc. Companion 23rd Annual ACM Conf. Innovation and Technology in Computer Science Education* 175–199 (ACM, 2018).
8. Luan, H. & Tsai, C.-C. A review of using machine learning approaches for precision education. *Educ. Technol. Soc.* **24**, 250–266 (2021).
9. Banihashem, S. K. et al. A systematic review of the role of AI in higher education: Innovations, challenges and future directions. *Stud. High. Educ.* **49**, 2219–2241 (2024).
10. Noroozi, O. et al. Artificial intelligence in higher education: Opportunities, challenges, and institutional policies. *Innov. Educ. Teach. Int.* **61**, 742–758 (2024).
11. Cortez, P. & Silva, A. Using data mining to predict secondary school student performance. In *Proc. 5th Future Business Technology Conf.* 5–12 (2008). Dataset: https://archive.ics.uci.edu/dataset/320/student+performance.
12. Kuzilek, J., Hlosta, M. & Zdrahal, Z. Open University Learning Analytics dataset. *Sci. Data* **4**, 170171 (2017).

13. Shahiri, A. M., Husain, W. & Rashid, N. A. A review on predicting student's performance using data mining techniques. *Procedia Comput. Sci.* **72**, 414–422 (2015).

14. Gašević, D., Dawson, S. & Siemens, G. Let's not forget: Learning analytics are about learning. *TechTrends* **59**, 64–71 (2015).

15. Viberg, O., Hatakka, M., Bälter, O. & Mavroudi, A. The current landscape of learning analytics in higher education. *Comput. Hum. Behav.* **89**, 98–110 (2018).

16. U.S. Department of Education. College Scorecard Data. https://collegescorecard.ed.gov/data/ (2024).

17. Romero, C. & Ventura, S. Educational data mining: A review of the state of the art. *IEEE Trans. Syst. Man Cybern. C* **40**, 601–618 (2010).

18. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

19. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining* 785–794 (ACM, 2016).

20. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* **30**, 4765–4774 (2017).