SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Bayesian Networks Framework For Predicting Academic Success: A Study With Qassim University Freshman Students In STEM Courses

Elham A. Al-Madhagi

Department of Statistics and operations Research, college of Science, Qassim University, Qassim, Saudia Arabia.
e.almadhagi@qu.edu.sa
Mathematics Department, Faculty of Education, Hodeidah University, Hodeidah,Yemen. Email: elham2110@yahoo.com ORCID:0000-0002-6326-6231

Abstract
Typically, models for predicting students' academic success achieve high classification accuracy but offer poor explainability for administrators and teachers who must interpret and act on their predictions. To predict students' academic success and provide interpretable explanations of their predictions, this study proposes and experimentally evaluates a hybrid machine learning framework that combines an ensemble learner, Bayesian Networks (BNs) structure learning, and post-hoc explainable artificial intelligence (XAI) techniques. The academic success was treated as a binary outcome (final grade G3 ≥ 10 versus G3 < 10) using a subsample of 95 secondary school graduates from the publicly available Saudi Student Performance Dataset. These students were inducted into Qassim University, Saudi Arabia in STEM courses. All classifiers were trained with a 75%/25% train/test split and evaluated using 5-fold cross-validation, reporting accuracy, precision, recall, F1-score, and ROC AUC. LightGBM achieved the best held-out performance (accuracy = 0.875, F1 = 0.897, and ROC-AUC = 0.926). In this small sample, prior academic achievement (represented by first-, second-period grades (G1, G2) and success represented by G3 ≥ 10) was the only statistically significant predictor of success ($p < 0.001$), while lack of school educational support and social engagement outside of school showed suggestive associations but failed to reach the $\alpha = 0.05$ significance threshold. The SHAP and LIME explanations were similar and identified the same key drivers of the model's predictions: prior GPA, social engagement outside of school, and weekly study time. A Bayesian network learned using the hill-climb search with the Bayesian Information Criterion (BIC) score recovered a single, directed edge from prior GPA to academic success, providing both explanatory power and structural sparsity in probabilistic graphical models from a limited dataset. The results illustrate that such a hybrid approach, which integrates ensemble learning with Bayesian network modeling and post-hoc XAI methods, yields an interpretable and structurally sparse probabilistic model despite the small sample size.

Keywords: Explainable Artificial Intelligence; Bayesian Networks; SHAP; LIME; Student Academic Success Prediction.

1. Introduction

Given the extensive availability of digital data on students' activities, learning outcomes, and assessment, prediction of academic performance has received research attention within the fields of educational data mining (EDM) and learning analytics. Further, ensemble models such as Random Forest, XGBoost, LightGBM, and CatBoost have consistently outperformed classical statistical models (Asselman et al., 2023; Liang et al., 2023; Tang et al., 2024). These algorithms are often referred to as "black box" algorithms; however, as users – teachers, counselors, and administrators – must learn to translate a prediction to design an intervention, they are difficult to interpret and therefore may be less suitable for direct use by educators and administrators. Meanwhile, model-agnostic explainability has opened a new direction in research by providing tools for explaining individual predictions through identification of the features that lead to the result, such as the widely used SHAP (SHapley Additive exPlanations) (Ujkani et al., 2024; Lu et al., 2026). Bayesian Networks (BNs), however, offer an alternative, highly interpretable probabilistic framework that can model conditional dependencies between variables and facilitate causal reasoning under uncertainty (Zhang et al., 2023; Hao et al., 2022).

This study proposes a single hybrid framework and empirically evaluates it using a small, authentic sample of grade 10 end-of-course student records (n = 95), representing a realistic institutional or classroom-level setting with limited student data. In practice, scenarios involving small cohorts are common. Recent research using small student samples from educational datasets, ensemble learning approach (Kaur et al., 2024), and Bayesian case-study methods (Looi et al., 2024) suggest that meaningful predictive and explanatory insights can be derived from small datasets. This supports the potential utility of such approaches in the present context.

2. Research Problem and Objectives

Despite there being many academic performance predictive models in vogue, their efficacy for the real stakeholders—namely educators and policy makers —remains extremely limited due to the difficulty of interpreting the data. Moreover, no single factor can account for learners' success (or failure) though one or more factors can be combined to optimize academic performance. These can be potent tools in the educators' hands to facilitate the best learning conditions. Contemporary predictive student performance models, on the other hand, prioritize accuracy over the ability to interpret the model or identify the factors that influence predictions. Consequently, educators do not have a clear roadmap to adopt while designing interventions. The current study fills this gap by developing an accurate, interpretable, and explicit model for determining the most important factors for student academic success.

2.1 Research Objectives

The study addressed five primary research objectives:

- To develop a model to predict the effect of nine different variables on student academic success using real-time data.
- To analyze the predictive ability of six machine learning algorithms (Logistic Regression, Random Forest, XGBoost, LightGBM, CatBoost, Artificial Neural Network).
- To implement Bayesian Networks to model probabilistic relationships among predictor variables.
- To implement Explainable AI (XAI) methods (SHAP, LIME) with statistical and machine learning models to improve interpretability of the model.
- To identify the key factors that support students' academic success.

2.2 Research Questions

To address the identified gap regarding predictive academic models, this study examined the following research questions:

- Which everyday behaviors and academic habits are associated with high academic performance among Saudi learners at the threshold of higher education?
- What probabilistic relationships does a Bayesian Network model represent among everyday behaviors, academic habits, and performance at the point of induction into a Saudi public university?
- How do the findings of the Bayesian Network analysis inform educational policymaking and practices in the context of Saudi high school education?

3. Literature Review

3.1 Educational Data Mining and Predictive Modelling

Educational data mining comprises computer-based statistical treatment of educational data, such as student performance and classroom behaviors (e.g., attendance and use of learning management systems), to generate models and predict learning outcomes. As these methods tend to be more effective on tabular, structured data from educational contexts than single classifiers, and are also more computationally tractable than deep learning on moderately sized sample sizes, ensemble methods such as bagging as in Random Forest and boosting as in XGBoost, LightGBM, and CatBoost have gained the most traction in literature. An XGBoost-based model performs significantly better at forecasting student performance than traditional models, as demonstrated by Asselman et al. (2023), whereas Liang et al. (2023) highlight the advantages of gradient boosting regression over other ML models for predicting academic performance from SOPL. This is followed by Tang et al. (2024), who present a deep ensemble of multiple neural networks with an optimization-based feature ranking process, and by Kaur et al. (2024), who present multiple base learners (Naïve Bayes, Support Vector Machines, Multi-Layer Perceptron, and Logistic Regression) combined in an ensemble for classifying students as weak and strong learners. The effectiveness of these approaches is further supported by the empirical benefit of ensemble methods across different educational domains.

3.2 Explainable Artificial Intelligence

The limited transparency of ensemble and neural models has led to the development of a distinct area of research which focuses on XAI to make their predictions more interpretable. In the context of XAI, SHAP and LIME are two widely used explanation methods executed to interpret challenging models. Based on cooperative game theory (Shapley values), SHAP can identify the features that contribute to a prediction and to what extent in addition to providing global explanations such as identifying the most influential feature across an entire dataset. LIME can explain individual predictions by interpreting the complex model around each observation. Chen et al. (2022) and Ujkani et al. (2024) employed DxAI for week-wise early predictions in VL and explainable AI to detect at-risk students to facilitate early intervention in course-completion prediction, respectively. In both contexts, the explainability and performance of various classifiers are contrasted and analyzed. Similarly, Guleria and Sood (2023) worked on an educational data mining application. Nagy and Molontay (2024), on the other hand, developed a framework for predicting dropout explicitly geared toward personalized intervention through XAI. More recently, Lu et al. (2026) noted that different explanation techniques (attribution techniques like SHAP versus counterfactuals like DiCE) could contradict each other in their assessments of the relative importance of features for the same underlying model, making cross-method agreement checks helpful for avoiding undue reliance on a single explainer.

3.3 Bayesian Networks in Educational Contexts

Bayesian Networks consist of a set of variables and their associated conditional probability tables, along with a directed acyclic graph (DAG) that encodes the conditional dependencies among the variables. One of their main advantages over black-box ML models is that the graph itself is a human-readable description of which variables are thought to affect which, and the network supports probabilistic inference (e.g., updating the probability of success based on observed facts such as prior grades). The graph can be obtained automatically from data through a structure-learning algorithm derived using the Bayesian Information Criterion (BIC) (Chickering, 2002) but the resulting graph depends on the sample size. Zhang et al. (2023) used Bayesian networks to quantify the indeterminate relationships among student-, curriculum-, and environment-level features and developed a model explaining how dualistic learning influences the innovation performance of university student teams. Hao et al. (2022) applied a Bayesian network model to predict MOOC performance and quantified the uncertainty in the relationships among student-, curriculum-, and environment-level features. Looi et al. (2024) presented a case study demonstrating the continued application of Bayesian networks to uncover relationships among factors affecting academic performance in online instruction during the COVID-19 era, with data size considered “moderate” for these models.

3.4 Integration Gap

Bayesian Network approaches have been applied to diverse predictive contexts of learners’ academic performance. One of the most relevant recent studies is Suzuki et al. (2026) in Japan. They used academic and non-academic factors to determine learners’ academic performance. However, the Japanese context differs vastly from the Saudi context in terms of the education systems, learners’ needs, and socio-economic circumstances, which warrants inquiry into this palpable research gap.

Closest to the current study in the Saudi context is Almarabeh (2017) which analyzed students’ performance using different data mining classifiers, the focus of the study was clearly on developments in the field of computer science and not education. What is relevant to mention here is that the study amongst the five tools used in this study, Bayesian Network classifier had the highest accuracy of prediction. Ramaano et al. (2026) concluded that the Admission Point Score (APS) metric used to admit students to academic programs in South Africa was not effective in predicting student success since as many as 50% of those admitted to science programs in higher education failed to complete the program during the period 2008-2015. The study aimed to use the Bayesian Network to develop a probabilistic approach as a viable alternative to APS.

Suzuki et al. (2026) applied the Bayesian Network to a large dataset from Japan’s National Assessment of Academic Ability to examine how academic performance is associated with every day and study behaviors. The Bayesian Network brought them to more than one interesting predictor of academic success. First, cutting on screen time and increasing study time on weekends; second, encouraging students to plan their study schedule; and third, encouraging students to study out of school on weekends were the interventions likely to improve academic performance trained Dynamic Bayesian Networks (DBN) and Bayesian Networks (BN) using psychological scores (PHQ-9, GAD-7, PSS-10, CD-RISC) to model psychological risk and academic records to model academic outcomes. The study concluded that the top mental health predictors for academic success were GPA, stress scores, depression scores, and intervention engagement.

Saeedi et al. (2024) noted that the ability to anticipate academic performance has gained considerable attention since contemporary educational institutions seek to enhance student outcomes and reduce the risk of failure. They evaluated existing predictive models against a proposed Bayesian Networks model and concluded that the latter has superior accuracy and robustness not only due to its predictive capabilities but also, for providing a valuable tool for early intervention in educational settings.

Overall, the available studies establish the efficacy of Bayesian Networks as predictive models that are efficient to evaluate a diverse range of academic success predictors and should be used to streamline the admission process into institutions of higher education to ensure long term learner retention and cut down on attrition rates. Moreover, the use of Ensemble Learning, XAI, and Bayesian Networks has been widely investigated separately in the Knowledge Harvesting EDM literature, their joint use in a hybrid predictive-and-explanatory setting has been less extensively investigated. Existing methods generally employ either ensemble learning and XAI together (e.g., Ujkani et al., 2024; Nagy & Molontay, 2024) or Bayesian networks independently (e.g., Zhang et al., 2023; Hao et al., 2022), but few studies combine all three in an ensemble cross-validated model. Using a small but comprehensive educational dataset, this study provides an empirical test of this framework.

4. Research Methodology

4.1 Data Source and Sample

The data students were drawn from the publicly available Saudi Student Performance Dataset which captures demographic, social, and academic features of secondary school students. Based on previous literature, the study included three period grades (G1, G2, G3) ranging from 0-20 for a mandated STEM course (Cortez & Silva, 2008; Cortez, 2014). The full public dataset contained 1112 student records. To mimic a realistic small-cohort analysis under the constraint that the analytical sample does not exceed 100 students, 95 students were selected from the large set 1112 students. The criterion for academic success was adopted as a binary target: Success = 1 when the final grade (G3) was ≥ 10 (on the 0–20 scale of Saudi grades at the final exam), and Success = 0 elsewhere. Success/failure was defined within the analytical sample: 63.2% of students were considered successful, while 36.8% were considered unsuccessful.

4.2 Variable Operationalization

The dataset's native attributes were mapped onto the variable categories specified in the study design as shown in Table 1.

Table 1. Mapping of proposed predictor variables to the source dataset

Proposed Variable	Operationalization in Dataset	Type
Age	age in years (15–22 range in source data)	Continuous
Gender	gender (M/F)	Binary categorical
Prior GPA	= mean of first- and second-period grades (G1, G2), 0–20 scale	Continuous
Study hours	= study time (ordinal, 1 = <2h to 4 = >10h/week)	Ordinal
Absence	= number of recorded absences	Continuous (count)
Social engagement outside of school	= composite of leisure time and going-out frequency	Continuous
Family background	family size	Binary categorical
Internet access	access to internet at home	Binary categorical
Socioeconomic status	SES = mean of mother's and father's education level	Continuous
Assessment scores	G1, G2 (used to build prior GPA); G3 (used to build the outcome)	Continuous

4.3 Analytical Pipeline

Data preprocessing began with a missing-value check (no missing values were found), followed by encoding categorical variables (binary variables were label-encoded and nominal variables were skipped), and finally by data scaling (StandardScaler for scale-sensitive models). This was followed by exploratory data analysis, including descriptive statistics and a bivariate correlation analysis of the continuous predictors. Then, categorical (chi-squared test of independence) and continuous (one-way ANOVA) effects were tested for each predictor with Success, and subsequent adjusted effect sizes were calculated using binary logistic regression models (using statsmodels). The machine learning models trained were Logistic Regression, Random Forest, XGBoost, LightGBM, CatBoost, and an Artificial Neural Network (Multi-Layer Perceptron), with a 75/25 train–test split (n train = 71, n test = 24) and evaluated with 5-fold stratified cross-validation. Another Bayesian network (BN) was built with the variables discretized into tercile bins. Its structure was learned by running the Hill-Climb search with the Bayesian Information Criterion (BIC) in the pgmpy package, and the parameters of the learned network were finally estimated using the Maximum Likelihood technique. Tree Explainer was used to calculate global SHAP values for the best-performing tree ensemble and a local LIME explanation for a prediction on a single test-sample in the test set. Finally, model evaluation was conducted on the held-out test set using accuracy, precision, recall, F1 score, and ROCAUC, with cross-validation to ensure stable results.

4.4 Software and Tools

Python 3 code was used for all analyses, including logistic regression, random forest, MLPClassifier, evaluation metrics, gradient boosting ensembles (XGBoost, LightGBM, and CatBoost), logistic regression inference (statsmodels), chi-square analysis, ANOVA, structure learning, parameter learning for a Bayesian network (pgmpy), and explainability analyses (SHAP and LIME). This follows the R/Python/SPSS/GeNIe/Netica toolchain specified in the original study design (interchangeable with R/VPython/SPSS/GeNIe/Netica). For this toolchain, the open-source Linux version of Python was used instead of Java, enabling pgmpy to support GeNIe/Netica-style BNs in addition to R/Netica/GeNIe/BN.

5. Results

5.1 Descriptive Statistics

The analytical sample (n = 95) was balanced by gender (48 male, 47 female) and predominantly composed of students living with a family unit larger than three members (GT3, n = 68 vs. LE3, n = 27), with home internet access (n = 77), no social engagement outside of school (n = 63), without additional school support (n = 87), and intending to pursue higher education (n = 93). Mean age was 16.57 years (SD = 1.25). Mean prior GPA (average of G1 and G2, 0–20 scale) was 10.75 (SD = 3.61); mean weekly study-time category was 1.95 (SD = 0.83, on a 1–4 ordinal scale); mean recorded absences was 4.82 (SD = 6.12); mean engagement composite was 4.66 (SD = 1.35); mean socioeconomic status proxy (parental education) was 2.57 (SD = 1.03, on a 0–4 scale.

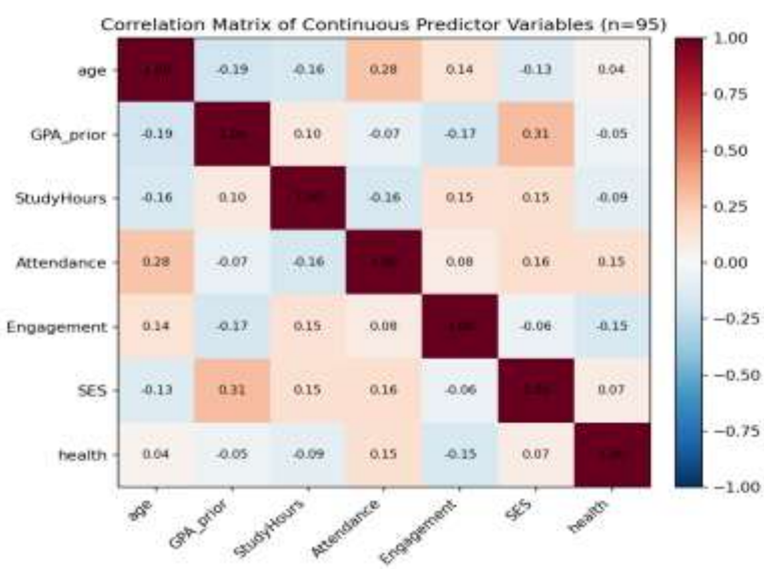

Figure 1. Correlation matrix of continuous predictor variables (n = 95).

5.2 Statistical Analysis

The results of these Chi-squared tests of independence (Table 2) revealed no statistically significant relationships at the conventional α = 0.05 level between the categorical predictors and the success outcome; however, school support and social engagement outside of school showed weak trends at (χ^2 = 3.822, df = 1, p = 0.051) and (χ^2 = 2.788, df = 1, p = 0.095) respectively.

Table 2. Chi-square test results for categorical predictors vs. academic success

Variable	χ^2	df	p-value
Gender	0.863	1	0.353
Family size	0.045	1	0.833
Internet access	0.222	1	0.637
Social engagement outside of school	2.788	1	0.095
School support	3.822	1	0.051
Plan to pursue higher education	0.123	1	0.726

One-way ANOVA among continuous variables (Table 3) also identified that prior GPA was the only variable with a statistically significant mean difference between successful and unsuccessful students ($F(1,93) = 118.55$, $p < 0.001$). No other variable reached significance, though the continuous variable of absence showed a weak pattern (F = 2.865, p = 0.094).

Table 3. One-way ANOVA results for continuous predictors vs. academic success.

Variable	F-statistic	p-value
Age	1.915	0.170
Prior GPA_	118.552	<0.001
Study hours	0.304	0.583
Absence	2.865	0.094
Social engagement outside of school	2.416	0.124
SES	1.342	0.250

After adjustment, prior GPA was the only statistically significant predictor in the binary logistic regression (Table 4) (Success ~ prior GPA_ + Study hours + Absence + Social engagement outside of school+ SES + age). For each one-unit increase in GPA (0-20), the odds ratio was approximately 5.92 (β = 1.779, SE = 0.504, z = 3.53, p < 0.001). See Table 4.

Table 4. Binary logistic regression results (dependent variable: Success)

Predictor	Coefficient (β)	Std. Err.	z	p-value
Prior GPA_	1.779	0.504	3.528	<0.001
Study hours	-0.744	0.555	-1.340	0.180
Absence	-0.106	0.097	-1.096	0.273
Social engagement outside of school	-0.237	0.318	-0.744	0.457
SES	-0.366	0.476	-0.770	0.441
Age	0.088	0.338	0.261	0.794

5.3 Machine Learning Model Comparison

All six classifiers were trained on the same 71-student training partition and then tested on a held-out test partition of 24 students. With an accuracy of 0.875 and a precision of 0.929 (Table 5, Figures 2, 3), LightGBM had the best overall held-out performance among all models, closely followed by Logistic Regression and CatBoost (F1 = 0.815). The Artificial Neural Network (MLP) had the lowest and most unstable accuracy (0.708) and ROC-AUC (0.719), despite being the logical and natural choice for deep/flexible models, which may require more than the relatively small number of students in these case studies.

Table 5. Model performance comparison on the held-out test set (n = 24) and 5-fold cross-validation (n = 95)

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC	5-fold CV Acc. (Mean ± SD)
LightGBM	0.875	0.929	0.867	0.897	0.926	0.895 ± 0.033
Logistic Regression	0.792	0.917	0.733	0.815	0.919	0.874 ± 0.026
CatBoost	0.792	0.917	0.733	0.815	0.874	0.863 ± 0.071
XGBoost	0.792	1.000	0.667	0.800	0.830	0.895 ± 0.033
Random Forest	0.750	0.909	0.667	0.769	0.874	0.842 ± 0.074
ANN (MLP)	0.708	0.833	0.667	0.741	0.719	0.758 ± 0.108

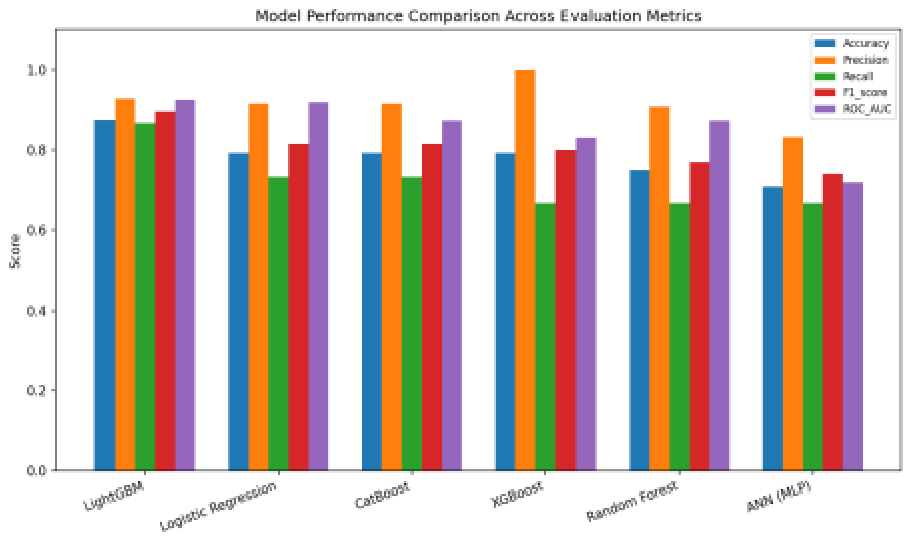

Figure 2. Model performance comparison across evaluation metrics

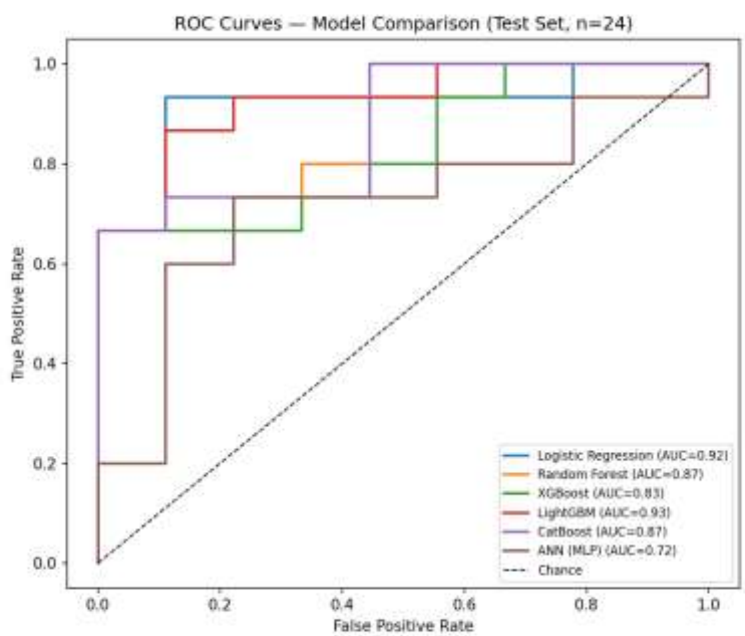

Figure 3. ROC curves for all six classifiers on the held-out test set

5.4 Feature Importance and Explainability (SHAP and LIME)

Prior GPA was by far the most important feature, ranked #1 by both RF and XGBoost native feature-importance scores (0.581 and 0.670, respectively) and by SHAP values computed on the best-fitted tree ensemble. Social engagement outside of school came in second, ahead of weekly study time (mean |SHAP| = 1.086 and 0.485, respectively). Internet access and absence of school support had relatively small, but substantial SHAP attributions compared with the other factors, while socioeconomic status (SES), age, and gender made negligible contributions to individual predictions in this sample. See Table 6.

Table 6. Comparison of feature-importance rankings across Random Forest, XGBoost, and SHAP.

Feature	RF Importance	XGBoost Importance	Mean |SHAP value|
Prior GPA	0.581	0.670	3.355
Social engagement outside of school	0.048	0.069	1.086
Study hours	0.014	0.023	0.485
Absence	0.061	0.042	0.007
Internet access	0.009	0.169	0.080
School support	0.035	0.025	0.007
SES	0.045	0.000	0.000
Age	0.036	0.000	0.000

Case-level interpretability is illustrated by a local LIME for a randomly selected test-set student: a lower prior GPA (on the 8–11 scale, compared to 12–15) was detrimental (adjusted predicted successful outcome pS by −0.135). However, high social engagement outside of school (>5) was similarly negative (−0.080). By contrast, low study time, moderate socioeconomic status, and moderate absence, were all negative or positive, with small magnitudes ranging from 0.016 to 0.053. This exemplifies the process of translating the same global drivers found by SHAP into an individualized, case-specific rationale for a one-to-one advising conversation.

5.5 Bayesian Network Structure

A Bayesian Network was induced from 6 variables from the VECs (prior GPA, study hours, absence, social engagement outside of school, socioeconomic status, success outcome), which were first discretized into 3 terciles. Under this scoring criterion (where more complex models, with more parameters for a given sample size, receive fewer points), the search narrowed to a single directed edge, namely prior GPA → academic success, as all other candidate edges were too poorly supported given the small sample. Interestingly, students in the high-GPA tercile are predicted to be successful 100% of the time, and students in the medium and low terciles are predicted to be successful 71% and 19% of the time, respectively. See Figure 4.

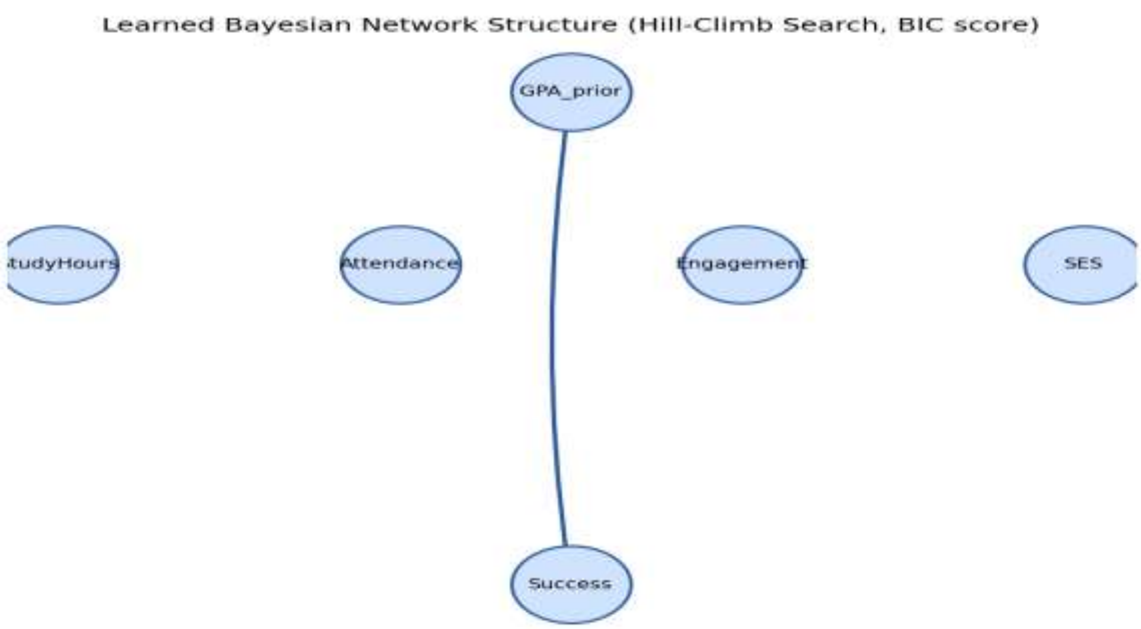

Figure 4. Learned Bayesian network structure (Hill-Climb search, BIC score, n = 95).

6. Discussion

There is strong consensus across the three lines of evidence in this study regarding a substantive conclusion: prior academic achievement denoted by prior GPA is, by a wide margin, the best available predictor of subsequent academic achievement in this sample. It was by far the highest-ranking variable across all explainability approaches (native ensemble importance, SHAP, and the learned structure of the Bayesian network). It is also the only variable that is statistically significant in both unadjusted ANOVA and adjusted logistic regression. This multi-method approach (statistical, machine-learning, and probabilistic-graphical) significantly increases confidence in the result, despite the relatively small sample size, and follows the reported pattern of prior achievement being the overwhelming predictor in other ensemble-learning work (Asselman et al., 2023; Tang et al., 2024). Social engagement outside of school and study hours ranked behind prior GPA in the machine-learning analyses, although neither reached statistical significance in the conventional inferential tests.

From a model-comparison perspective, LightGBM was the most favorable, on the one hand, for its predictive accuracy and, for its cross-validated stability, since it slightly beat the simple, fully interpretable logistic regression model in terms of ROC-AUC (0.919 vs. 0.926). This observation has practical implications: For small educational cohorts, the accuracy improvement of the most sophisticated model (which, sadly, appears and feels intractable to humans) over the simple, easily understood one remains relatively small, further strengthening the case for this study's approach to hybrid explanations: an ensemble model that generally performs better than the simplest, most understandable model, combined with individual explanations produced using SHAP/LIME, and a learned Bayesian Network for providing easy-to-understand explanations at the institution level. Şahin and Erol (2024) and Kaur et al. (2024) also concluded that simpler or hybrid ensemble models perform competitively with more complex models on moderately sized educational datasets. Finally, the fact that SHAP, the native feature importance of ensemble methods, and Bayesian Network structure learning, all point to prior GPA as the most important academic success predictor, their shared agreement is more useful for informing educators' decisions. This strategy of triangulation addresses concerns about providing explanations with due consideration to disagreements, as outlined by Lu et al. (2026), and is similar to multi-technique explainability approaches considered for educational applications by Guleria and Sood (2023) and Nagy and Molontay (2024).

7. Conclusion

This study proposes and empirically validates an adaptive, explainable machine learning model using Bayesian Networks to predict academic success and explain student performance. Applied to an authentic, small (n = 95) subset of the Saudi Student Performance Dataset, the model indicated that prior academic achievement was the strongest and statistically significant predictor of success across most commonly used analytical techniques. Secondary, non-significant trends involving study time, social engagement outside of school, school support, and socioeconomic status also emerged and could be investigated with larger datasets. Its main methodological contribution is a layer-wise framework for researchers and practitioners to build a competitive ensemble of classifiers (LightGBM gradient-boosting model, F1 = 0.897, ROC-AUC = 0.926), establish a transparent statistical baseline (logistic regression, ROC-AUC = 0.919), cross-validate against it, and train with post-hoc explainability methods (SHAP, LIME). It can also be trained independently with a Bayesian probabilistic modeling structure, yielding predictions that are not only accurate but also defensible for educators or institutional decision makers. Future research needs should include the ability to expand this framework with larger, multi-institutional sample sizes, allow for the inclusion of learning-management-system engagement logs, and examine the practical applicability of this framework in an actual early-warning intervention context.

8. Future Directions

Further avenues of research should be pursued to build on this work: First, the framework needs to be extended to larger multi-institutional cohorts building on the leading work of Hao et al. (2022) and Zhang et al. (2023). The framework should also be expanded with multi-institution and multi-subject replications (as in Liang et al., 2023; Asselman et al., 2023) to examine the degree of generalization of the prior-achievement-dominance phenomenon in Saudi secondary education settings other than STEM fields. In addition to these methodologic extensions, the framework must be tested over time within an actual interactive early-warning intervention pilot, examining the student outcomes of an advising session.

9. Ethical Considerations

This analysis uses a publicly available benchmark dataset (Saudi Student Performance Dataset), a de-identified version, obtained on the premise that the original author(s) consented to its public release with their institution's approval. No new data collection from human participants was undertaken for this study. Nonetheless, this study was reviewed and approved by the Research Ethics Committee of Qassim University, Saudi Arabia (approval number: [insert approval number]).

10. Limitations

The sample size in this study (n = 95, 71 training, 24 test) was modest to detect small-to-moderate effects among secondary predictors. Another limitation is class imbalance and the use of a single train/test split: experiments were optimized to reduce the impact of class imbalance, and confidence intervals for held-out test-set metrics derived from data for just 24 students are broad, so these metrics should be interpreted with caution. Lastly, recent research in XAI emphasizes that post-hoc explanations are not certain but rather a ground truth. Note that there is disagreement across different explanation techniques (e.g., SHAP vs. counterfactuals) on what matters most for the same given model, as pointed out by Lu et al. (2026); this study focused only on comparing SHAP to LIME and did not attempt to quantitatively measure agreement among the different explainers that could be used, a limitation which future research may address.

References

1. Almarabeh, H. (2017). Analysis of students' performance by using different data mining classifiers. International Journal of Modern Education and Computer Science, 9(8), 9–15. https://doi.org/10.5815/ijmecs.2017.08.02
2. Asselman, A., Khaldi, M., & Aammou, S. (2023). Enhancing the prediction of student performance based on the machine learning XGBoost algorithm. Interactive Learning Environments, 31(6), 3360–3379. https://doi.org/10.1080/10494820.2021.1928235
3. Chen, H.-C., Prasetyo, E., Tseng, S.-S., Putra, K. T., Kusumawardani, S. S., & Weng, C.-E. (2022). Week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence. Applied Sciences, 12(4), 1885. https://doi.org/10.3390/app12041885
4. Chickering, D. M. (2002). Optimal structure identification with greedy search. Journal of Machine Learning Research, 3, 507–554.
5. Cortez, P. (2014). Student performance [Data set]. UCI Machine Learning Repository. https://doi.org/10.24432/C5TG7T
6. Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. In A. Brito & J. Teixeira (Eds.), Proceedings of the 5th Annual Future Business Technology Conference (pp. 5–12). EUROSIS.
7. Guleria, P., & Sood, M. (2023). Explainable AI and machine learning: Performance evaluation and explainability of classifiers on educational data mining inspired career counseling. Education and Information Technologies, 28(1), 1081–1116. https://doi.org/10.1007/s10639-022-11221-2
8. Hao, J., Gan, J., & Zhu, L. (2022). MOOC performance prediction and personal performance improvement via Bayesian network. Education and Information Technologies, 27, 7303–7326. https://doi.org/10.1007/s10639-022-10926-8
9. Kaur, H., Kaur, T., Bhardwaj, V., & Kumar, M. (2024). An ensemble deep learning model for classification of students as weak and strong learners via multiparametric analysis. Discover Applied Sciences, 6(11), 1–14. https://doi.org/10.1007/s42452-024-06274-6
10. Liang, G., Jiang, C., Ping, Q., & Jiang, X. (2023). Academic performance prediction associated with synchronous online interactive learning behaviors based on the machine learning approach. Interactive Learning Environments, 32(6), 3092–3107. https://doi.org/10.1080/10494820.2023.2167836
11. Looi, Z. N., Song, P. C., Lim, H. T., & Looi, S. Y. (2024). A case study via Bayesian network: Investigating factors influencing student academic performance in online teaching and learning during COVID-19 pandemic. In Y. Bee Wah, D. Al-Jumeily, & M. W. Berry (Eds.), Data Science and Emerging Technologies (DaSET 2023) (Lecture Notes on Data Engineering and Communications Technologies, Vol. 191). Springer. https://doi.org/10.1007/978-981-97-0293-0_23
12. Lu, Y., Rajendra, A. S., Zhang, J., & Zhao, T. (2026). Explainable machine learning for student performance prediction. AI in Education, 2(2), Article 17. https://doi.org/10.3390/aieduc2020017

13. Nagy, M., & Molontay, R. (2024). Interpretable dropout prediction: Towards XAI-based personalized intervention. International Journal of Artificial Intelligence in Education, 34, 274–300. https://doi.org/10.1007/s40593-023-00331-8
14. Ramaano, T., Jadhav, A., & Ajoodha, R. (2026). Developing a Bayesian network model to predict students' performance based on the analysis of their higher education trajectory. International Journal of Modern Education and Computer Science, 18(1), 21–47. https://doi.org/10.5815/ijmecs.2026.01.02
15. Saeedi, S., Božanić, D., & Safa, R. (2024). Strategic analytics for predicting students' academic performance using cluster analysis and Bayesian networks. Education Science and Management, 2(4), 197–214. https://doi.org/10.56578/esm020402
16. Şahin, S., & Erol, Ç. (2024). Prediction of secondary school students' academic achievements with machine learning methods and a sample system. Cybernetics and Systems, 55(4), 940–960.
17. Suzuki, T., Morita, N., & Okuda, T. (2026). Bayesian network analysis of academic performance and its behavioral determinants among junior high school students in Japan. Discover Education, 5(1). https://doi.org/10.1007/s44217-026-01222-5
18. Tang, B., Li, S., & Zhao, C. (2024). Predicting the performance of students using deep ensemble learning. Journal of Intelligence, 12(12), 124. https://doi.org/10.3390/jintelligence12120124
19. Ujkani, B., Minkovska, D., & Hinov, N. (2024). Course success prediction and early identification of at-risk students using explainable artificial intelligence. Electronics, 13(21), 4157. https://doi.org/10.3390/electronics13214157
20. Zhang, X., Liu, J., Zhang, C., Shao, D., & Cai, Z. (2023). Innovation performance prediction of university student teams based on Bayesian networks. Sustainability, 15(3), 2335. https://doi.org/10.3390/su15032335

Appendix A: Reproducibility Notes

All results reported in this manuscript were generated from a single, fixed random subsample (random_state = 42) of the publicly available Saudi Student Performance Dataset. Source data, the analysis script, and all intermediate result files (descriptive statistics, chi-square and ANOVA tables, logistic regression output, model evaluation metrics, feature-importance rankings, SHAP values, LIME explanation, and Bayesian network structure/parameters) are retained alongside this manuscript for verification and re-analysis.