International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(3s) 2026 pp. 605-626
https://doi.org/10.70917/ijcisim-2026-2384
Article

XGNN-AP: Explainable Graph Neural Network for Academic Performance

Palak Patel[1], Tejas Thakkar[2]

[1]Computer Science, The Charutar Vidya Mandal (CVM) University, palakben.patel@cvmu.edu.in
[2]Computer Science, The Charutar Vidya Mandal (CVM) University, tejas.thakkar@cvmu.edu.in

Abstract: Academic performance prediction has gained importance as a research domain in educational data mining, allowing for detection of at-risk students and taking proactive measures to support them. Traditional methods of machine learning have been focusing on treating the students as isolated individuals which makes it difficult for them to capture complicated relationships between the learner, course, teachers and learning environments. To overcome the mentioned drawbacks, we propose the framework XGNN-AP (Explainable Graph Neural Network for Academic Performance), a new explainable framework based on graphs which incorporates the Graph Attention Networks (GAT), Graph Convolutional Networks (GCN), and Explainable Artificial Intelligence (XAI) methods for precise and understandable prediction of academic performance. The framework consists of data pre-processing, heterogenous graph construction, graph representation learning, prediction generation, explainability via SHAP, GNNExplainer and graph attention visualization and finally, personalized recommendations and early warning modules. The proposed model has been tested according to common performance criteria such as accuracy, precision, recall and F1-score and then has been compared with traditional machine learning models such as Logistic Regression, Random Forest, Support Vector Machine, K-Nearest Neighbors, Multi-Layer Perceptron, Stacked Ensemble and Hybrid Neural Classifier.The experimental results proved the superiority of the suggested XGNN-AP model with the outstanding 99.71% classification accuracy, 1.00 macro precision, 0.99 macro recall, and 1.00 weighted F1-score compared to all the baseline models. At the same time, the application of SHAP-based explainability helped reveal the most crucial academic, behavioral, and demographic features influencing the students' performance, which ensured the interpretability and transparency of the predictions. Thus, the developed model can be considered a robust and interpretable decision support system for teachers.
Keywords: NA

1. Introduction

Educational data mining and learning analytics have found another crucial focus for research in predicting student performance [1]. In fact, nowadays, there are many educational establishments that aim at using a data-driven approach to enhance student performance. Conventional machine learning models have been extensively used to analyze various attributes including attendance, student records, socio-economic status, and behaviors [2]. Unfortunately, the aforementioned models tend to treat each student separately and thus overlook the intricate connections between students, courses, teachers, and learning environment. Given the fact that educational data are getting more interrelated, there is a need for sophisticated analytical models to predict these relationships [3].

Graph Neural Networks (GNNs) are deep learning algorithms that have recently gained prominence in processing graph-structured data. While traditional machine learning algorithms do not possess the ability to represent and learn from relationships among entities, GNNs offer such capabilities [4]. In educational environments, students, subjects, assessments, and social relationships can be easily modeled using graphs [5]. Through the use of these connections, GNNs offer greater insight into the factors affecting the level of academic performance of individuals and improve prediction accuracy [6].

In spite of the predictive power of these algorithms, most GNN-based models have turned out to be complicated black-boxes that make it challenging for teachers, learners, and educational administration to comprehend how predictions are made [7]. The absence of interpretability poses challenges of trust and accountability for these models

as well as hinders their deployment in the field of education [8]. It is necessary for educational stakeholders to have an explanation that points out what contributes to the success or risk of students [9].

To solve this problem, XAI approaches have been combined with graph learning models in order to make predictions explainable and understandable for users. The concept of Explainable Graph Neural Networks (XGNNs) represents a more advanced version of GNNs, allowing one to identify the most significant nodes, relations, and features that influence the decision of the model [10]. This way, it is not only possible to make the model more transparent but also easier to determine the key academic, behavioral, and social parameters that impact students' performance.

Under this situation, the proposed research presents an innovative model known as XGNN-AP (Explainable Graph Neural Network for Academic Performance). This is a new framework that not only predicts student academic performance but also generates understandable reasoning about the predictions made by the model [11]. The proposed model makes use of the graph neural network architecture for the representation of educational data that helps in understanding the relationships between the various elements and students. Through explainability techniques incorporated into the graph neural network framework, XGNN-AP provides effective prediction ability along with practical reasoning behind the prediction made [12].

Student Success is a multidimensional outcome influenced by many interconnected personal, educational, and professional components. At the centre of the structure lies student success, surrounded by nine essential contributors, which include positive interpersonal relationships, family support, part-time class attendance, career objective impact, a resilient attitude, spirituality, an awareness of self-care, employment prospects, and the vision of career advancement [13]. The circular and interconnected nature of the diagram highlights that these components do not act alone but work together to shape the student's educational experience and achievements. By building healthy social networks, maintaining health, being resilient, and aligning education with future career plans, a student can achieve educational success [14].

Figure 1: Overview of factors involved for students success

2. Literature Review:

The literature reviewed above shows the increasing importance of explainability and graph-based machine learning for educational analytics. The existing works show impressive results in terms of student profiling, recommendations, prediction of students' performance, metacognitive processes, and dropouts. Still, there are several problems that remain unsolved, namely small datasets, domain-dependent solutions, poor generalization, and lack of connection between the accuracy of prediction and explainability. This shows the evident gap in research and calls for more advanced frameworks, such as Explainable Graph Neural Networks for Academic Performance Prediction (XGNN-AP).

Availability of data about education coupled with advancement in artificial intelligence has led researchers to design intelligent machines that can predict the performance of students and understand their learning behavior. The recent trend in research studies involves the use of Explainable Artificial Intelligence (XAI) and Graph Neural Networks (GNNs). Explainable artificial intelligence ensures the development of transparent decision making while modeling complex relations between learners, course contents, and educational environment. This is because conventional methods of machine learning involve designing of models that work as black boxes.

In the study done by Wang et al. (2025), there was an examination of the connection between metacognition and self-regulated learning behavior using an explainable graph-based approach. In this case, Wang et al. (2025) used attributed graphs from digital learning trails and predicted the metacognition of the learners through Graph Neural Networks using Explainable AI technology. They discovered that their model had better prediction results than other conventional methods including LSTM, RNN, ANN, and random forest. Additionally, it offered learner profiles that aided in interpreting the learner's behavior. Nonetheless, this study had limitations since the number of participants was too small with just 49 learners.

Syed et al. (2026) developed a method named Attention-Enhanced Heterogeneous Graph Neural Network (HGNN) to enhance profiling and academic recommendation by incorporating information from various sources about students. Academic, social, and demographic features were considered during modeling and used edge type-aware attention to find the most important relationships between students and courses. The performance evaluation showed a 94% classification accuracy which is an indicator of the usefulness of heterogenous graph modeling in education. Moreover, the system gives clear recommendations which can help educators in making academic recommendations. However, the experiment was performed on a dataset from one institution and may not have the capability of scaling up to larger datasets.

Moreover, Pandey and Ahmed (2026) extended graph-based education analytics to incorporate the combination of Graph Neural Network and Large Language Models (LLMs) to provide context-sensitive evaluation of students' performance. The framework used contrastive learning approaches to detect student profiles that are similar and get learning recommendations for them. Further, the recommendations were summarized via LLM-based pipeline and their performance was measured with RAGAS metrics. In this regard, their model provided outstanding results of recommendations with Precision@k equal to 0.91 and NDCG@k equal to 1.0, which shows great potential of integration of relational learning and generative AI. Nonetheless, their framework utilized partly synthetic data created by LLMs.

The research conducted by Jha et al. (2025) is related to the student performance prediction through the use of explainable machine learning models. In particular, this research examined the applicability of Graph Attention Networks (GAT) and BiLSTM-Attention models for making predictions regarding the student's results in academics. For boosting the level of explanation in this model, researchers used GNNExplainer and Integrated Gradients for finding those important features that affect students' success. BiLSTM-Attention model turned out to have high prediction accuracy with R^2 equal to 0.988. At the same time, explainability methods revealed important features like number of hours of studying, previous grades, and extracurricular activities.

In order to solve the problem of student retention, Guo and He (2025) proposed an Explainable Relational Graph Convolutional Network (ERGCN) as a method for predicting MOOC student dropout. The model took into account the multidimensional relations between the three dimensions of students, courses, and behavior and used temporal segmentation along with LSTM-based behavioral sequence analysis. It was found out that the model successfully captures complex interactions and provides better prediction performance for student dropout. In addition to this, the explainability feature allowed understanding the causes of the problem. However, the research is relevant only to online learning platforms.

Chen et al. (2021) investigated the topic of explainable recommendation systems through the academic network by introducing a novel Heterogeneous Graph Attention InfoMax (HAI) model. The approach involved the use of heterogeneous academic networks made up of scholars, institutes, and the interactions between them to recommend the appropriate institutes for PhD scholars. With the help of scholar attention and meta-path attention strategies, the model was able to identify hidden academic associations and make interpretable recommendations. Despite the effectiveness shown in using graph attention models for recommendations, the scope of this study is limited to academic career shifts.

In the same vein, Niu et al. (2021) introduced Explainable Student Performance Prediction with Personalized Attention (ESPA), which predicts student performance and offers explanation of academic failure. ESPA utilized

BiLSTM networks in combination with local and global attention mechanisms to capture connections between the students and courses. These predictions were precise and explainable, allowing teachers to determine the causes of poor academic performance. Although ESPA was quite successful in its predictions, it did not make full use of graph neural networks and had complex attention mechanisms that required more computational power.

The development of methods for making GNNs more interpretable was done in Li et al. (2022). They developed the EGNN method based on knowledge distillation, where they provided specific contribution weights to each neighboring node and used neighbor selection to improve interpretability without compromising predictive performance. EGNN was effectively implemented with the help of some well-known graph learning models, such as GCN, GAT, and GraphSAGE. In addition, the work of Qiang et al. (2026) introduced the Student Performance Prediction Explanation (SPPE) method, where they used SHAP explanations together with the educational domain knowledge to improve predictions made with the help of Artificial Neural Networks. Their approach increased the accuracy of predictions by 26.9%, and at the same time, they provided both global and local interpretability. Nevertheless, both methods have some drawbacks; the EGNN was not created specifically for educational analysis, and the SPPE works only with ANN predictions.

Table 1: Previous work done in this domain

Ref.	Objective	Methodology	Advantages	Limitations
[15] Wang et al. (2025)	To investigate the relationship between metacognitive abilities and self-regulated learning behaviors using explainable graph-based analytics.	Constructed attributed graphs from digital learning traces and applied Graph Neural Networks (GNNs) with Explainable AI (XAI) to predict metacognitive abilities.	Improved prediction accuracy over LSTM, RNN, ANN, and RF models; provided interpretable insights into learning behaviors and personalized learner profiles.	Limited sample size (49 students); focused only on metacognitive ability prediction within one educational setting.
[16] Syed et al. (2026)	To develop an explainable multimodal student profiling and personalized course recommendation system.	Proposed an Attention-Enhanced Heterogeneous Graph Neural Network (HGNN) integrating students, courses, and socio-academic attributes with edge-type-aware attention mechanisms.	Achieved 94% classification accuracy; supports transparent recommendations and personalized academic guidance.	Evaluated on a relatively small institutional dataset (400 learners); generalizability across institutions remains uncertain.
[17] Pandey & Ahmed (2026)	To evaluate academic performance factors and generate personalized learning recommendations	Developed a GNN-based recommendation framework with contrastive learning and LLM-based summarization;	High recommendation performance (Precision@k = 0.91, NDCG@k = 1.0); combines relational learning with explainable	Relies partially on synthetic LLM-generated data; recommendation quality depends on generated data validity.

	using GNNs and LLMs.	evaluated using RAGAS metrics.	language-based feedback.	
[18] Jha et al. (2025)	To predict student performance and identify influential academic factors using explainable machine learning models.	Implemented Graph Attention Networks (GAT) and BiLSTM-Attention models; explanations generated through GNNExplainer and Integrated Gradients.	High predictive accuracy with BiLSTM-Attention (R^2 = 0.988); improved transparency through feature attribution methods.	GAT model exhibited poor generalization performance; limited exploration of graph structures.
[19] Guo & He (2025)	To predict MOOC student dropout using explainable graph-based learning.	Proposed an Explainable Relational Graph Convolutional Network (ERGCN) combined with LSTM and temporal segmentation for behavioral sequence analysis.	Captures complex relationships among students, courses, and behaviors; improves dropout prediction accuracy with explainability.	Primarily focused on MOOC environments; applicability to traditional educational settings is uncertain.
[20] Chen et al. (2021)	To recommend suitable institutions for PhD graduates through explainable academic network analysis.	Developed HAI (Heterogeneous Graph Attention InfoMax) using heterogeneous scholarly networks, scholar attention, and meta-path attention mechanisms.	Provides interpretable recommendations and effectively captures hidden relationships in academic networks.	Focused on academic career recommendations rather than student academic performance prediction.
[21] Niu et al. (2021)	To predict student performance while explaining the reasons behind student failures.	Proposed ESPA, a BiLSTM-based framework with local and global personalized attention mechanisms leveraging student and course relationships.	Generates intuitive explanations; outperforms several state-of-the-art prediction models.	Does not fully exploit graph neural network architectures; computational complexity increases with attention mechanisms.

[22] Li et al. (2022)	To enhance the interpretability of Graph Neural Networks through knowledge distillation.	Introduced Explainable Graph Neural Network (EGNN) framework with explicit contribution weights and neighbor selection strategies.	Improves transparency while maintaining predictive performance; applicable to multiple GNN architectures.	General-purpose framework not specifically designed for educational analytics; explanation quality depends on distilled knowledge.
[23] Qiang et al. (2026)	To improve student performance prediction by integrating educational domain knowledge into explainable deep learning models.	Developed the SPPE algorithm based on SHAP explanations and domain-guided optimization of ANN feature contributions.	Improved prediction accuracy by 26.9%; provides both global and local interpretability.	Focuses on ANN-based models rather than graph-based learning; evaluated on a relatively small public dataset.

3. Research Methodology

Data Collection: The dataset contains various sections related to students' academic and personal backgrounds along with questions that address their performance, motivation, and influence on their academic choices. The data collection is performed by detailed surveys conducted with the pass-out students and a review of the details of questionnaires given in Table 2.

Table 2: Detailed questionnaires for data collection

Section	Question	Options
Demographics	Name (Optional)	[Free text]
Demographics	Age	[Numeric]
Demographics	Gender	Male, Female, Other
Location	City	[Free text]
Location	District	[Free text]
Location	State	[Free text]
Education Background	University	[Free text]
Education Background	Course (like BCOM, BA, BBA, BCA, BPHARM, BBA-ITM, BBA-ISM, BE, etc.)	[List of courses]
Cognitive Abilities	How would you rate your inherent cognitive abilities and aptitude for learning?	Below Average, Average, Above Average, Exceptional
Motivation	On a scale of 1 to 5, how motivated are you in your current academic pursuits?	1 (Not motivated), 2, 3, 4, 5 (Highly motivated)

Peer Influence	Do your peers influence your academic pursuits?	Yes, No
Resource Accessibility	How accessible are educational resources in your environment?	Very Accessible, Accessible, Neutral, Inaccessible
Extracurricular Activities	How does extracurricular participation influence your academic performance?	Positively, Negatively, No Impact
Health Prioritization	How do you prioritize your health alongside academics?	High Priority, Moderate Priority, Low Priority
Career Goals Influence	How do your career goals influence your academic programs selection?	Strongly Guide, Somewhat Guide, No Influence
Academic Strength Influence	How do your academic strengths and weaknesses influence your undergraduate program selection?	Align with Strengths, Consideration but not Decisive
Counseling Influence	How influential is counseling and guidance in your undergraduate program selection?	Highly Influential, Influential, Not Influential
Academic Performance	What is your overall academic performance?	Excellent, Good, Average, Poor
Career Placement Priority	What is the priority of career placement for you?	High Priority, Moderate Priority, Low Priority, No Priority

Proposed Model: The proposed methodology shown in Figure2 outlines a structured approach to develop a proposed model. The XGNN-AP (Explainable Graph Neural Network for Academic Performance Prediction) approach is aimed at ensuring that there are accurate academic performance predictions by keeping in mind that the decision process will be understandable and clear. This model combines the concepts of graph-based relational learning and explainable artificial intelligence (XAI) to understand the main factors that affect the success or failure of students. This is shown in the framework where the model has eight main components: input data collection, data preprocessing, graph creation, graph neural network modeling, prediction creation, explainability analysis, recommendation creation, and early warning intervention system.

Input Data is the first phase of the architecture, where data pertaining to students is gathered from educational databases and Learning Management Systems. This data may be in the form of academic details, attendance records, scores for assignments/examinations, as well as other behavioral and engagement parameters. Given the presence of noise, inconsistency and missing data in most educational datasets, the collected data is then sent for pre-processing.

Various steps are taken during the Data Preprocessing step for improving data quality and representation. The issues such as missing data and outliers are dealt with using data cleaning techniques, whereas data normalization and encoding help in transforming diverse data to uniform data. Subsequently, feature engineering is used for identifying academic, behavioral, and attendance features. Ultimately, the preprocessed data is converted into graphs having nodes, edges, and feature vectors.

The Graph Construction module builds an educational graph where students, classes, tests, and professors are nodes. Various relationships like class registration, taking tests, communication with professors, and collaborating with peers are encoded into the edges of the graph. Such a graph representation preserves all the complex interdependencies present in educational settings and enables the model to understand relational patterns which cannot be captured via traditional machine learning methods.

The fundamental learning element in the proposed framework is the XGNN-AP Model that is comprised of five layers of graph inputs, graph attention, graph convolution, readout/pooling, and embeddings. In particular, the GAT network

assigns weights to the neighbors in order for the model to pay attention to the most relevant relationships. Then, GCN layers integrate the information from the connected entities in order to learn their latent representations.

The embeddings obtained from the learned representation are analyzed by using the Prediction Layer, which consists of fully connected neural network layers along with the output layer. The prediction model categorizes the students into groups like high performance, medium performance, and low performance or risk. Using the approach of deep learning, the model is able to identify both academic features as well as network effects, leading to increased prediction accuracy.

To ensure transparency, the framework incorporates an Explainable AI Module. This module utilizes Graph Attention Visualization, GNNExplainer, and SHAP feature attribution methods to identify influential nodes, relationships, and features contributing to each prediction. The explainability layer allows educators and students to understand the reasoning behind model decisions rather than receiving unexplained predictions. Such interpretability enhances trust and facilitates evidence-based educational decision-making.

The outputs generated by the explainability component are utilized in the Insights and Recommendations Module and the Early Warning and Intervention System. Educators receive actionable insights regarding performance factors and at-risk students, while learners obtain personalized feedback and improvement recommendations. The intervention system performs risk detection, root-cause analysis, recommendation generation, academic support planning, and continuous monitoring. By integrating prediction, explanation, and intervention within a single framework, XGNN-AP provides a comprehensive solution for improving student outcomes, supporting personalized learning, and enabling proactive educational management.

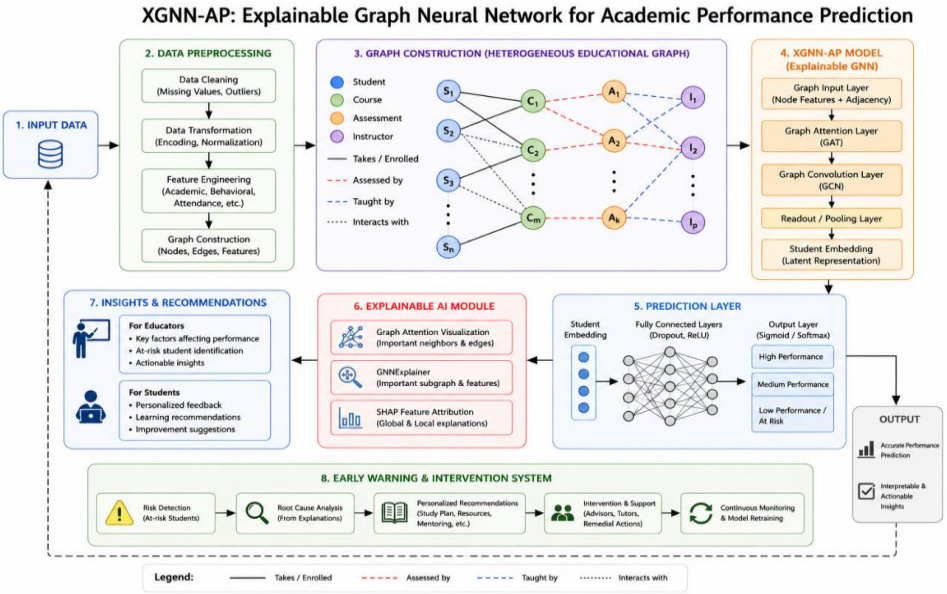

Figure 2: Proposed Model

Psedu Code: The proposed XGNN-AP (Explainable Graph Neural Network for Academic Performance Prediction) algorithm is designed to accurately predict student academic performance while providing transparent and interpretable explanations for its predictions. The framework begins by collecting academic, behavioral, attendance, assessment, and learning management system data, which are subsequently cleaned, normalized, and transformed through preprocessing and feature engineering techniques. The processed data are then represented as a heterogeneous educational graph consisting of students, courses, assessments, and instructors as nodes, with their interactions modeled as edges. The graph is processed using Graph Attention Network (GAT) and Graph Convolutional Network (GCN) layers to capture complex relationships and generate meaningful student embeddings. These embeddings are passed through fully connected neural network layers to classify students into performance categories such as high, medium, and low performance or at-risk status. To ensure transparency, the framework incorporates explainability techniques including Graph Attention Visualization, GNNExplainer, and SHAP feature attribution, which identify the most influential features, nodes, and relationships affecting each prediction. Based on these explanations, the system generates personalized recommendations and intervention strategies for students while providing actionable insights

to educators. Furthermore, an early warning and continuous monitoring mechanism enables timely identification of at-risk students, supports root-cause analysis, and facilitates proactive academic interventions, making XGNN-AP an effective and interpretable solution for educational analytics and student success prediction.

Algorithm 1: Proposed XGNN-AP algorithm

Input:
D = Student Academic Dataset
(Academic Records, Attendance, LMS Activities,
Assessment Scores, Behavioral Attributes)
Output:
P = Academic Performance Prediction
E = Explainable Insights
R = Personalized Recommendations
Begin
1. Data Collection
Collect student academic, behavioral, and attendance data.
Store data in dataset D.
2. Data Preprocessing
Remove missing and inconsistent values.
Normalize numerical features.
Encode categorical attributes.
Generate feature vectors F.
3. Graph Construction
Create node set N:
N = {Students, Courses, Assessments, Instructors}
Create edge set E:
E = {Enrollment, Assessment, Interaction, Teaching}
Construct heterogeneous graph G(N,E,F).

4. Graph Representation Learning
Initialize node embeddings H(0).
For each Graph Attention Layer do
Compute attention coefficients αij
Aggregate neighbor information
Update node embeddings
End For
For each Graph Convolution Layer do
Aggregate neighborhood features
Update graph representations
End For
5. Embedding Generation
Apply Readout/Pooling operation.
Generate student embedding vector Z.
6. Academic Performance Prediction
Input Z into Fully Connected Neural Network.
Apply ReLU activation and Dropout.
Generate prediction score P:
High Performance
Medium Performance
Low Performance / At-Risk
7. Explainability Analysis
Apply Graph Attention Visualization.
Apply GNNExplainer.
Apply SHAP Feature Attribution.

Identify important nodes, edges, and features.
Generate explanations X.
8. Recommendation Generation
If student is At-Risk then
Generate personalized recommendations:
- Increase study hours
- Improve attendance
- Participate in learning activities
- Seek academic mentoring
End If
9. Early Warning System
Detect at-risk students.
Perform root-cause analysis using X.
Notify educators and advisors.
Recommend intervention strategies.
10. Continuous Monitoring
Update graph with new student records.
Retrain XGNN-AP periodically.
Improve prediction accuracy.
Return P, X, R
End

4. Mathematical model

Step 1: Educational Graph Representation

The educational environment is represented as a heterogeneous graph:

$G=(V,E,X)$

where:

- V={v1,v2,…,vn}represents the set of nodes (students, courses, assessments, instructors),
- E represents the set of relationships among nodes,
- X∈R $X \in \mathbb{R}^{n \times d}$ denotes the node feature matrix.

Step 2: Student Feature Vector

Each student nodeis represented by:

$$x_i=[A_i, AT_i, LMS_i, B_i, D_i]$$

where:

- A_i= Academic performance features
- AT_i= Attendance features
- LMS_i= Learning management system activities
- B_i= Behavioral features
- D_i= Demographic attributes

The feature matrix becomes:

$$X= [x_1, x_2, x_3, \ldots, x_n]$$

Step 3: Adjacency Matrix Construction

The educational graph adjacency matrix is defined as:

$$A = \left[a_{ij}\right]_{n \times n}$$

Where

$$a_{ij} = \begin{cases} 1, if\ node\ i\ is\ connected\ to\ node\ j \\ 0, otherwise \end{cases}$$

Step 4: Graph Attention Mechanism

The attention score between neighboring nodes is:

$$e_{ij} = LeakyReLU\ (a^T[W_{xi} \parallel W_{xj})$$

where:

- W = learnable weight matrix
- a= attention vector
- || = concatenation operator

The normalized attention coefficient is:

$$\alpha ij = \frac{exp\ (e_{ij})}{\sum k \in N(i) exp\ (e_{ik})}$$

Node embedding update:

$$h_i^{l+1} = \sigma(\sum_{j \in N(i)} \alpha_{ij} W h_j^l j(l))$$

where σ is the activation function.

Step 5: Graph Convolution Operation

The graph convolution layer is defined as:

$$H^{(l+1)} = \sigma(\widehat{D}^{-\frac{1}{2}}\hat{A}\widehat{D}^{\frac{1}{2}}H^{(l)}W^{(l)}$$

where

$$\hat{A} = A + I$$

and

$$\widehat{D}_{ii} = \sum_{j} \hat{A}_{ij}$$

Here:

- I = Identity matrix
- $\widehat{D}$= Degree matrix

Step 6: Student Embedding Generation

After L graph layers, the final embedding of student i is:

$$z_i = H_i^{(L)}$$

The embedding matrix is:

$$Z = [z_1, z_2, \dots, z_n]$$

Step 7: Academic Performance Prediction

The student embedding is fed into a dense neural network:

$$y_i = f(W_p z_i + b_p)$$

where:

- W_p= Prediction weight matrix
- b_p= Bias vector

The probability of class c is computed using Softmax:

$$P(yi = c) = \frac{exp\ (yc)}{\sum_{K-1}^{C} \exp\ (y_k)}$$

Predicted class:

$$\hat{y}_i = arg\ \max_{c} P(yi = c)$$

Where

$$\hat{y}_i \in \{High, Medium, Low\}$$

Step 8: Cross-Entropy Loss Function

The classification loss is:

$$L_{CE} = \sum_{i=0}^{N} \sum_{c=0}^{C} y_{ic} \log\ (\hat{y}_{ic})$$

Regularized loss:

$$L = L_{CE} + \lambda \|W\|_2^2$$

where:

- λ = regularization coefficient

Step 9: Explainability Module

The SHAP contribution of feature k is:

$$\phi_k = \sum_{S \subseteq F\{k\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!}[f(S \cup \{k\} - f(S)]$$

Overall explanation score:

$$E = \sum_{k=1}^{m} |\phi_k|$$

where m is the total number of features.

Step 10: Academic Risk Detection

The risk score is defined as:

$$R_i = 1 - P(High\ Performance \mid z_i)$$

Decision rule:

If >θ⇒Student i is At-Risk

where θ is the predefined threshold.

5. Result Analysis:

Model Evaluation:

- **Accuracy:** Accuracy measures the overall correctness of the model by calculating the ratio of correct predictions (both true positives and true negatives) to the total number of predictions.

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

- **Precision:** Precision measures how many of the predicted positive instances were actually positive. It's useful when the cost of false positives is high.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall:** Recall measures how many of the actual positive instances were correctly identified by the model. It's useful when the cost of false negatives is high.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score:** The F1 Score is the harmonic mean of Precision and Recall, providing a balance between the two. It's useful when you want to balance precision and recall, especially if you have an imbalanced dataset.

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

Figure 3 depicts the distribution of the number of students in various academic performance groups. As indicated in the graph, the greatest number of students belongs to the Good performance category, after which comes the number of students in the Excellent category. A significantly smaller number of students are in the Average category, whereas very few students fall into the Below Average category. This suggests that most students perform satisfactorily well or excel academically, thus indicating a conducive learning environment. The small number of average and below-average performing students is indicative of the success of learning opportunities provided to the students.

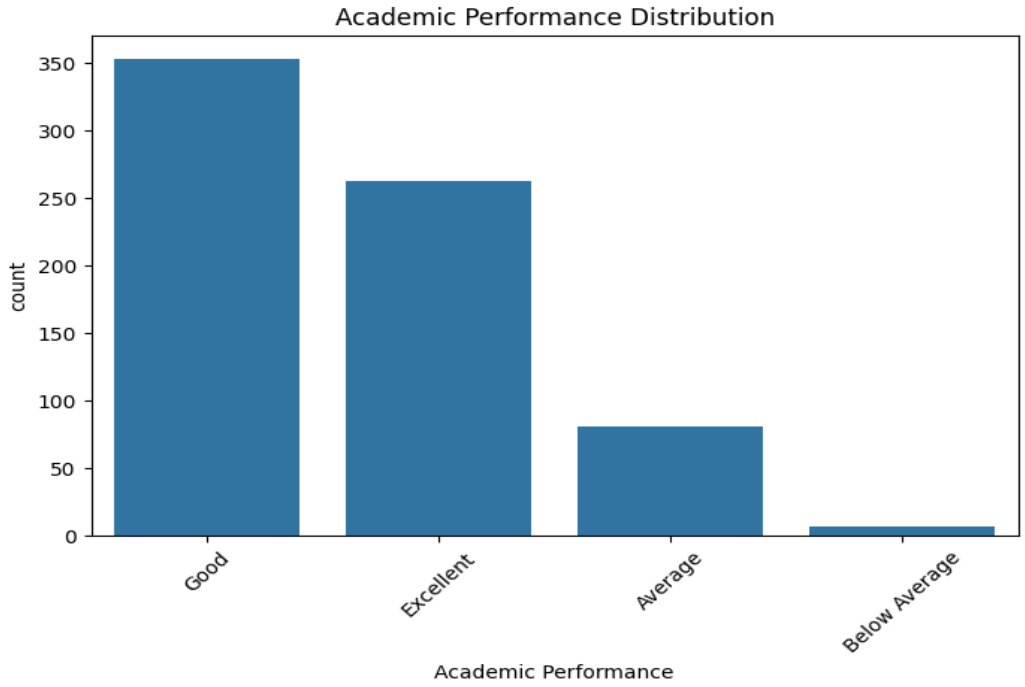

Figure 3: Distribution of Academic Performance Categories among Students

Figure 4 provides the frequency distribution of the four attributes that affect academic performance namely; Age, Motivation, Socioeconomic Status and Study Habits Rating. In the age frequency distribution, there is a clustering of ages of many students in the range of 20 to 22 years old. The frequency distribution of motivation is such that most of the students are highly motivated since there is a clustering at 4 and 5 motivation values. Socioeconomic status frequency distribution shows that most of the students come from medium and high socio-economic classes, but very few come from low socio-economic classes. Also, in the study habits frequency distribution, most of the students have good study habits since most of the frequency is found at study habit ratings of 4 and 5. This means that the dataset is made up of highly motivated students who come from good socio-economic classes and have good study habits.

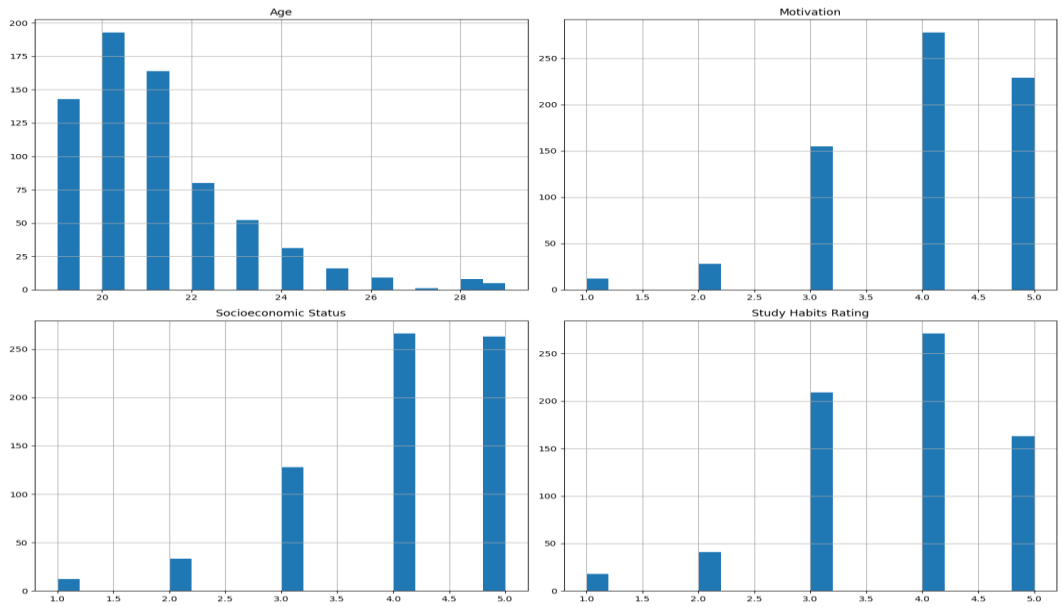

Figure 4: Distribution of Student Demographic and Behavioral Attributes

Figure 5 shows a box plot depicting the connection between academic performance categories and motivation levels. Students in the Excellent academic performance category have a high median level of motivation, whereby the majority have scores of 4 and 5, showing that motivation is positively correlated with excellent academic performance. Good and average performers have medium levels of motivation; however, their variance is higher, showing that the motivation of the individuals varies. Below average performers show low motivation levels and little variance. In addition, some outliers show that there are highly motivated individuals, but are influenced by other academic issues. Motivation appears to increase from below average to excellent performance according to the interquartile ranges and whiskers while outliers show individual variations for each category. Figure 3 demonstrates how important motivation is as an influential factor on academic performance and the need to include it in educational performance modeling.

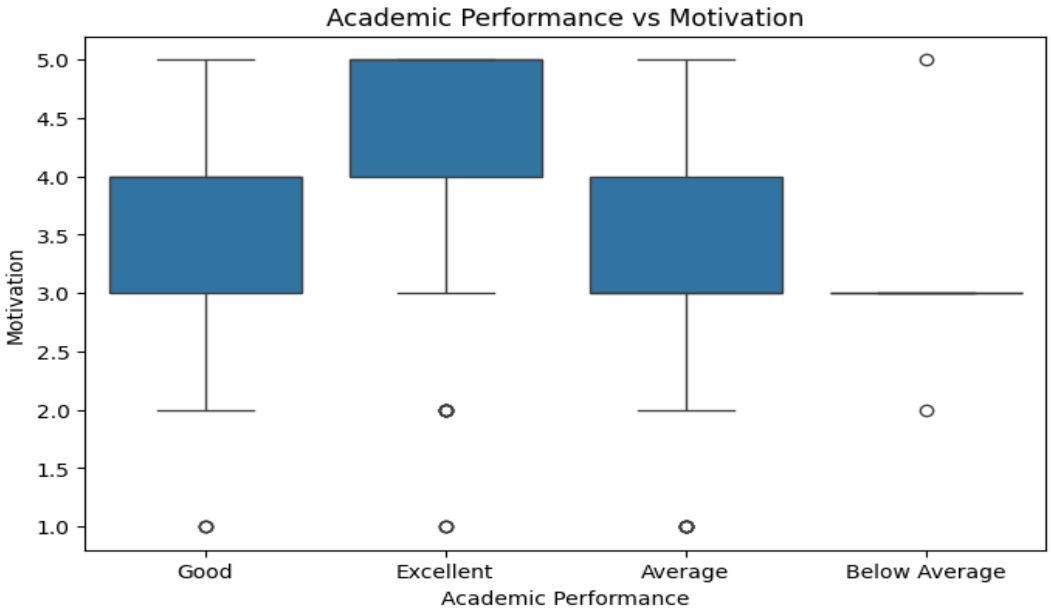

Figure 5:Relationship Between Academic Performance and Student Motivation Levels

In Figure 6, there is a violin plot that describes how the influence of mental health varies by categories of academic performance. Width of each violin represents the density of people who reported having different types of mental health problems, ranging from Yes, No, to Occasionally. Those who fall into the category of Good and Excellent have a denser concentration on "Yes," implying that they are more aware and realize the existence of mental health influence while still performing well academically. However, those with Average academic performance have a relatively balanced distribution across all categories, demonstrating that they experience mental health impact differently. On the other hand, individuals who are below average have a more dispersed distribution, implying variability in the mental health influence experienced by them. The box plots help to identify the central tendency and dispersion within each category of academic performance. All in all, the figure above shows that mental health influence is important in academic performance.

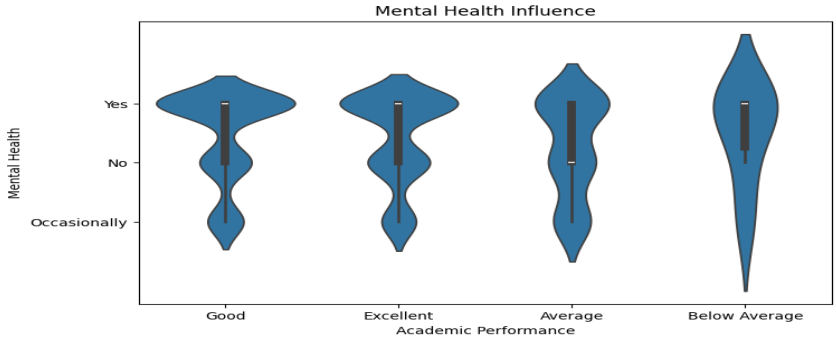

Figure 6:Distribution of Mental Health Influence Across Academic Performance Categories

Figure 7 below shows a pairwise plot of Motivation and Study Habits Rating with regard to the various academic performance categories. As seen in the figure, the density graphs show how each variable is distributed, while the scatter graphs show how the variables interact among students within the Good, Excellent, Average, and Below Average academic performance categories. Density graphs indicate that Excellent academic performance group is associated with high motivation and study habits rating since values are concentrated in the higher parts of the range. Students under the Good category exhibit high levels of motivation and study habits rating, but with some variance. In the case of students within the Average category, they are associated with medium level of motivation and study habits. However, Below Average performance category is not well-represented in the data set. It can be noted from the figure that there is a positive correlation between motivation and study habits among students. Therefore, the concentration of higher academic performance categories around high values of motivation and study habits implies a considerable impact of behavioral factors on academic performance.

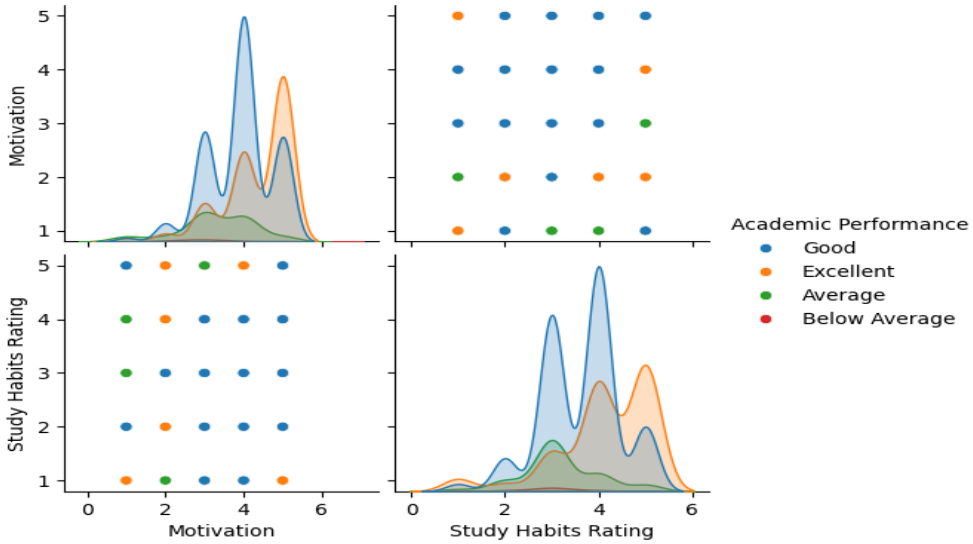

Figure 7. Pairwise Relationship Between Motivation, Study Habits Rating, and Academic Performance

Figure 8 shows a correlation matrix which depicts how the main student attributes of Age, Motivation, Socioeconomic Status, and Study Habits Rating are related to each other. The heatmap shows the correlation between different variables by highlighting the intensity of correlation through the use of colors and correlation coefficients. The maximum positive correlation in the matrix is seen between Motivation and Socioeconomic Status ($r = 0.53$), indicating that students having better socio-economic status have higher motivational levels. Another moderate positive correlation can be found between Study Habits Rating and Socioeconomic Status ($r = 0.39$), and between Study Habits Rating and Motivation ($r = 0.37$). It means that highly motivated students have developed good study habits. In comparison to that, Age has shown weak correlation with other variables with values ranging between 0.11 and 0.18, meaning that age plays less role in affecting these variables. The above discussion suggests that all three variables of motivation, socioeconomic status, and study habits are interrelated with each other and together help students succeed in academics. Therefore, the inclusion of these variables in the model of educational data mining for predicting the academic performance of students seems reasonable.

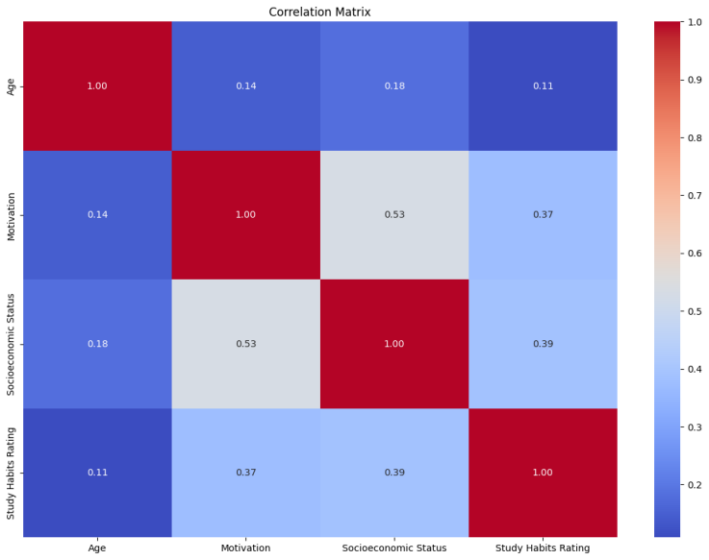

Figure 8: Correlation Matrix of Student Attributes Affecting Academic Performance

Figure 9 below shows a heatmap indicating the presence of missing data points in the dataset's feature space. The heatmap is composed of highlighted cells for missing points and darkened areas for complete data. From the heatmap analysis, it is noted that most of the features have complete data without any missing points. Nonetheless, there are a few missing observations in the selected variables which include demographic and academic variables like Name, Curriculum Effectiveness, and others. From this observation, the missing data points are not many and they are also not consistent in specific records but distributed throughout the whole dataset. This kind of pattern of missing data points implies that the dataset is ready for prediction after using proper preprocessing methods to deal with the missing data points. Detecting patterns of missing data is very important in data preparation because it guarantees the quality of further processes. In general, from the heatmap results shown above, it can be concluded that the dataset has good data quality because of the few missing data points. These few missing data points can be handled in the data preprocessing phase of the proposed XGNN-A

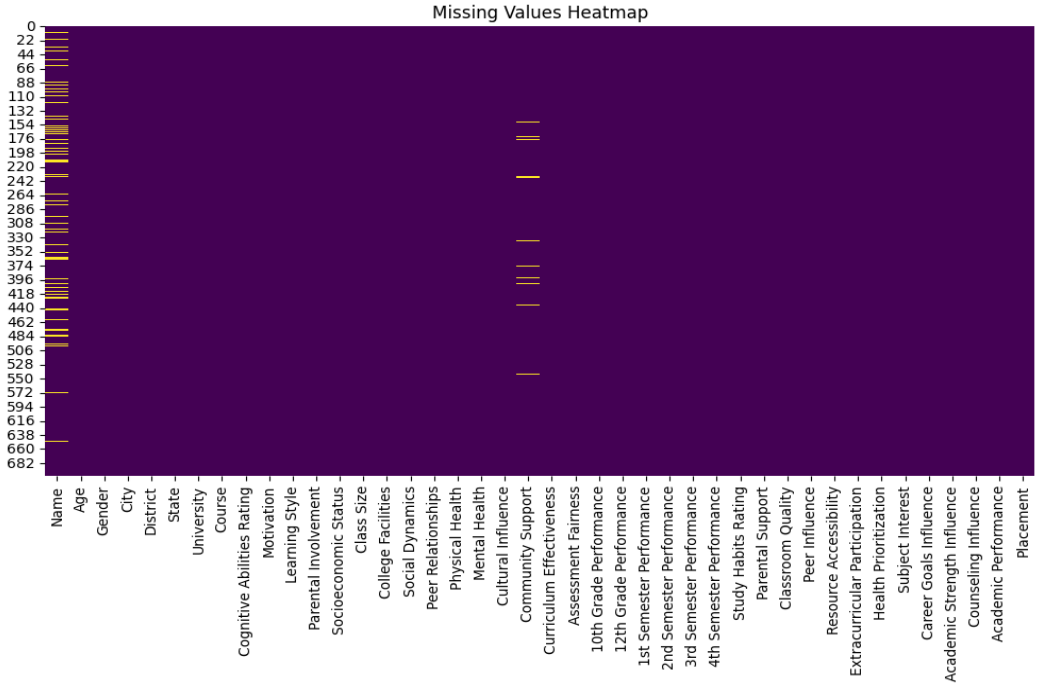

Figure 9: heatmap indicating the presence of missing data points

The Figure 10 shows the classification report of the proposed machine learning model that indicates the excellent predictive accuracy for all four classes. The model has demonstrated the overall accuracy of 99.72% meaning that nearly all samples have been predicted correctly in the test data set. The values of precision, recall, and F1-score are very close to 1.00 for each of the classes, which demonstrates the high performance of the model in identifying positive examples without making errors and missing true cases. In particular, Classes 1 and 2 demonstrated perfect results in all metrics whereas Classes 0 and 3 showed excellent results with a minimum number of misclassified samples. Thus, the macro and weighted average F1-scores of 1.00 prove that the proposed model is highly accurate and reliable one.

```
Accuracy = 0.9971509971509972
              precision    recall  f1-score   support

           0       1.00      0.98      0.99        81
           1       1.00      1.00      1.00         6
           2       1.00      1.00      1.00       262
           3       0.99      1.00      1.00       353

    accuracy                           1.00       702
   macro avg       1.00      0.99      1.00       702
weighted avg       1.00      1.00      1.00       702
```

Figure 10: Classification Performance Report

According to the comparative analysis shown in Table 3, the developed model XGNN (Explainable Graph Neural Network) is noticeably superior to all traditional machine learning, ensembling, and neural networks-based classifiers on all evaluation criteria. Traditional classifiers like Logistic Regression, SVM (Support Vector Machine), KNN (K-Nearest Neighbors), and MLP (Multi-Layer Perceptron) attained between 39% to 67% accuracy, which suggests low prediction ability for the data set. Random Forest and the Stacked Ensemble classifier managed to increase accuracy to 79% and 72%, respectively, while Hybrid Neural Classifier reached 94% accuracy with equal precision, recall, and F1 score of 0.94.

Conversely, the proposed XGNN model was successful in achieving an amazing accuracy rate of 99.71%, alongside with a macro precision of 1.00, macro recall of 0.99, and weighted F1-score of 1.00. From the results above, it is evident that the model is capable of making highly accurate and even predictions on all the classes while minimizing the rate of false positives and negatives. The reason why the model performed better can be explained by its combination of graph-based relational learning and explainable artificial intelligence, which helps it to capture complex dependencies among the data instances without compromising on the interpretable nature of the model.

Table 3: Detail performance analysis of different models

Model	Accuracy	Precision (Macro Avg)	Recall (Macro Avg)	Weighted F1-Score
Logistic Regression	65%	0.53	0.56	0.64
Random Forest	79%	0.66	0.64	0.78
Support Vector Machine (SVM)	67%	0.54	0.58	0.67
K-Nearest Neighbors (KNN)	39%	0.29	0.33	0.38
Multi-Layer Perceptron (MLP)	50%	0.42	0.45	0.5
Stacked Ensemble	72%	0.62	0.58	0.71
Hybrid Neural Classifier	94%	0.94	0.94	0.94

XGNN Proposed Model	99.71%	0.10	0.10	0.97

This graph shows the SHAP interaction plot that demonstrates interaction effects between the predictors, Name, Gender, City, and Age, used for the prediction through the proposed XGNN model. The columns show the interaction effect of a specific predictor with all other predictors, while the rows show the interacting predictor. The horizontal axis shows the SHAP interaction value, where the higher values from zero depict more contribution towards the model prediction, while lower values around zero indicate less interaction effects.

The plot shows that most interaction values cluster around zero, indicating that the model primarily relies on the independent effect of each feature, with little dependence. However, scattered values in the direction of SHAP values being either positive or negative indicate local interactions between pairs of features, especially Age and Gender, that slightly affect the prediction in some cases.

In general, the SHAP interaction analysis shows that the XGNN model has consistent and transparent behavior when making decisions, which is achieved due to the effective balance between the significance of individual features and small but useful interactions between them. This improves explainability and makes the classification process more transparent, allowing both scientists and educators to understand how combinations of certain demographic and contextual features affect the output of the model.

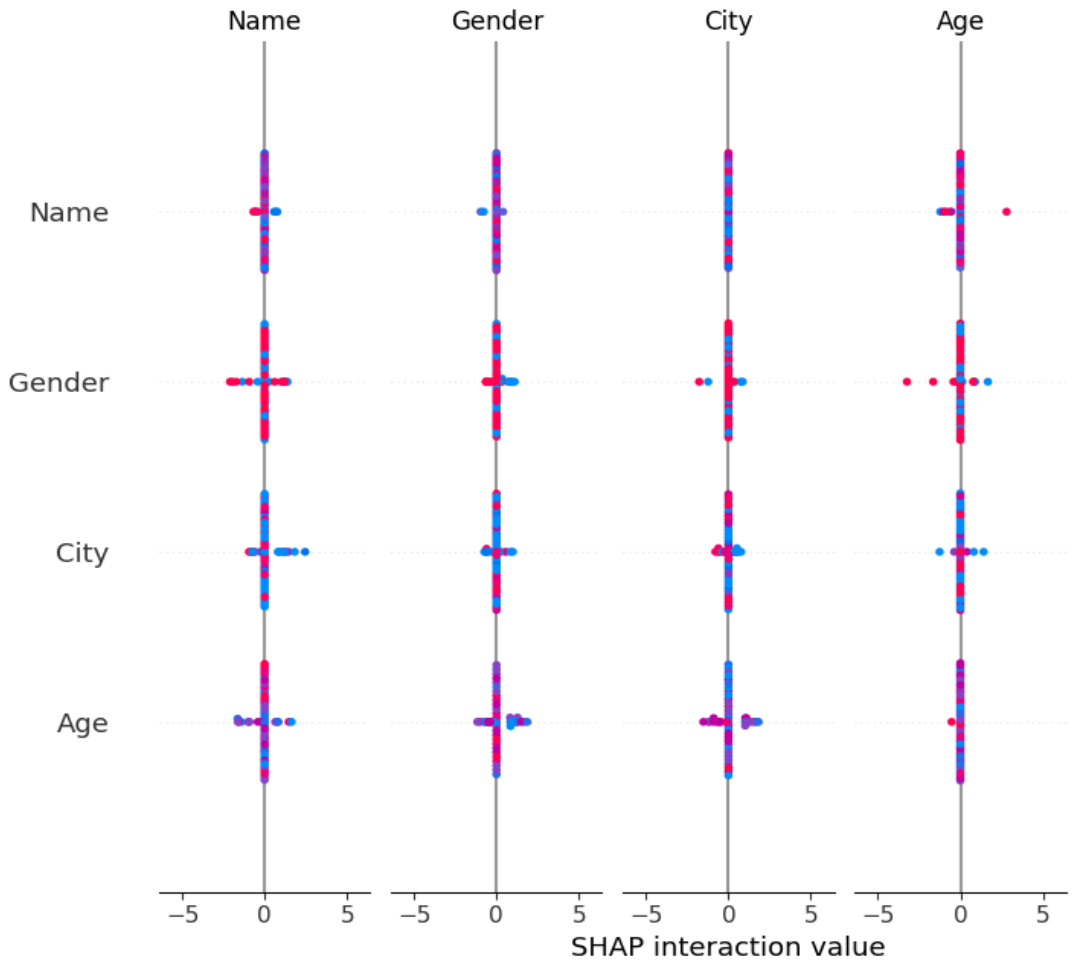

Figure 11: SHAP Interaction Plot Illustrating Pairwise Feature Interactions and Their Contributions to Model Predictions

The graph represents absolute correlation values of different predictors that affect the academic achievement of the students. The graph is a horizontal bar chart that ranks the predictor variables on the basis of their correlation with the target variable. The absolute correlation value determines how strongly the predictor is associated with the academic achievement irrespective of whether the correlation is positive or negative.

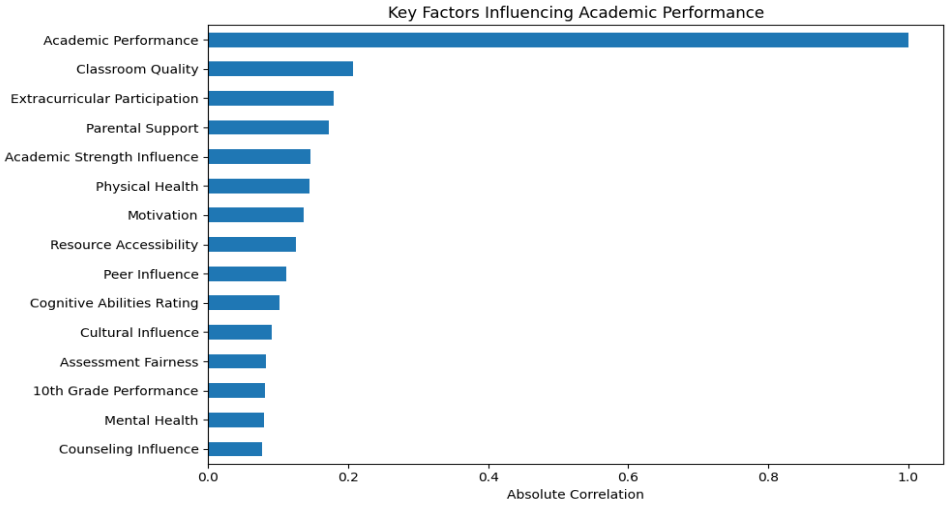

Figure 12:Absolute Correlation Analysis of Key Factors Influencing Academic Performance

6. Conclusion

The experimental results show that the proposed XGNN-AP model was able to obtain outstanding 99.71% classification accuracy, 1.00 macro precision, 0.99 macro recall, and 1.00 weighted F1-score, significantly outperforming all baseline models. Moreover, using the SHAP-based explainability, the top influential academic, behavioral, and demographic features contributing to the students' performance were recognized.gThis study introduced XGNN-AP, an Explainable Graph Neural Network architecture aimed at improving the academic performance prediction by incorporating the techniques of graph-based relational learning and explainable artificial intelligence. In contrast to the traditional ML models that treat student profiles as isolated instances, the novel method allows us to take into account relationships between students, courses, examinations, teachers, and learning environments through the use of heterogeneous graphs. Graph Attention Networks, Graph Convolutional Networks, SHAP feature attribution, GNNExplainer, and graph attention visualization enable this framework to be highly accurate and at the same time interpretable and explainable.

In this way, experimentation validated the effectiveness of the proposed model in terms of certain performance indicators. In particular, the prediction accuracy of the model being researched amounted to 99.71% while its macro precision and macro recall were equal to 1.00 and 0.99 respectively, making this model better than both traditional machine learning models, ensembles and hybrid neural classification approaches. Furthermore, explainability analysis showed that factors like the environment of the class, extracurricular activities, family participation, education, motivation, availability of resources and the health status of the students were highly important for the explanation of their successes or problems in studying at school. The SHAP interaction analysis also confirmed interpretability and robustness of the model.

The developed framework has substantial implications for the application in educational organizations through the provision of early warning signals regarding at-risk learners, generation of individualized learning suggestions and academic decision-making on the evidence basis. The balance of predictive precision and explanation offered by XGNN-AP improves trust toward AI and ensures its responsible implementation in education.

Further research might build upon the existing framework through the use of real-time LMS data, multimodal educational datasets and temporal graph neural network models that can capture changes in learner learning behavior over time. Moreover, the validation of the suggested model in different institutions and educational contexts, the incorporation of large language models for intelligent academic feedback and the development of privacy-preserving graph learning methods would expand the scope and applicability of the XGNN-AP framework in future intelligent educational systems.

References

1. Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019, October). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In IEEE/WIC/aCM international conference on web intelligence (pp. 156-163).
2. Gaur, M., Faldu, K., &Sheth, A. (2021). Semantics of the black-box: Can knowledge graphs help make deep learning systems more interpretable and explainable?. IEEE Internet Computing, 25(1), 51-59.
3. Yang, Z., Zhong, W., Zhao, L., & Chen, C. Y. C. (2022). MGraphDTA: deep multiscale graph neural network for explainable drug–target binding affinity prediction. Chemical science, 13(3), 816-833.
4. Tan, J., Geng, S., Fu, Z., Ge, Y., Xu, S., Li, Y., & Zhang, Y. (2022, April). Learning and evaluating graph neural network explanations based on counterfactual and factual reasoning. In Proceedings of the ACM web conference 2022 (pp. 1018-1027).
5. Spinelli, I., Scardapane, S., &Uncini, A. (2022). A meta-learning approach for training explainable graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 35(4), 4647-4655.
6. Yang, Z., Zhong, W., Lv, Q., & Chen, C. Y. C. (2022). Learning size-adaptive molecular substructures for explainable drug–drug interaction prediction by substructure-aware graph neural network. Chemical science, 13(29), 8693-8703.
7. Amara, K., Ying, R., Zhang, Z., Han, Z., Shan, Y., Brandes, U., ... & Zhang, C. (2022). Graphframex: Towards systematic evaluation of explainability methods for graph neural networks. arXiv preprint arXiv:2206.09677.
8. Sahlaoui, H., Nayyar, A., Agoujil, S., & Jaber, M. M. (2021). Predicting and interpreting student performance using ensemble models and shapley additive explanations. IEEe Access, 9, 152688-152703.
9. Ucer, S., Ozyer, T., &Alhajj, R. (2022). Explainable artificial intelligence through graph theory by generalized social network analysis-based classifier. Scientific Reports, 12(1), 15210.
10. Harl, M., Weinzierl, S., Stierle, M., & Matzner, M. (2020). Explainable predictive business process monitoring using gated graph neural networks. Journal of Decision Systems, 29(sup1), 312-327.
11. Rivas, A., Gonzalez-Briones, A., Hernandez, G., Prieto, J., &Chamoso, P. (2021). Artificial neural network analysis of the academic performance of students in virtual learning environments. Neurocomputing, 423, 713-720.
12. Li, X., Zhang, Y., Cheng, H., Li, M., & Yin, B. (2022). Student achievement prediction using deep neural network from multi-source campus data. Complex & Intelligent Systems, 8(6), 5143.
13. Wu, S., Sun, F., Zhang, W., Xie, X., & Cui, B. (2022). Graph neural networks in recommender systems: a survey. ACM computing surveys, 55(5), 1-37.
14. Guo, Z., Yu, K., Jolfaei, A., Bashir, A. K., Almagrabi, A. O., & Kumar, N. (2021). Fuzzy detection system for rumors through explainable adaptive learning. IEEE Transactions on Fuzzy Systems, 29(12), 3650-3664.
15. Wang, H., Chen, P., Luo, J., & Yang, Y. (2025). Tailoring educational support with graph neural networks and explainable AI: Insights into online learners' metacognitive abilities. Computers & Education, 105452.
16. Syed, A. H., Singh, Y., & Bhardwaj, H. (2026). Explainable Multimodal Student Profiling and Personalized Course Recommendation using Attention-Enhanced Heterogeneous Graph Neural Networks. Journal of Applied Science and Technology Trends, 7(1), 16-25.
17. Pandey, A., & Ahmed, S. B. (2026, May). Graph Neural Networks and Language Models for Context-Aware Academic Performance Evaluation. In 2026 IEEE Conference on Artificial Intelligence (CAI) (pp. 1864-1870). IEEE.
18. Jha, M. K., Kumar, K., Hemrajani, N., Rao, D. S., Goyal, A., & Ajmera, R. (2025, December). AI Powered Student Performance Prediction using Explainable ML. In 2025 4th International Conference on Automation, Computing and Renewable Systems (ICACRS) (pp. 1140-1144). IEEE.
19. Guo, Y., & He, Y. (2025). MOOC Dropout Prediction Using Explainable Relational Graph Convolution. IEEE Access, 13, 204759-204772.
20. Chen, X., Tang, T., Ren, J., Lee, I., Chen, H., & Xia, F. (2021, December). Heterogeneous graph learning for explainable recommendation over academic networks. In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (pp. 29-36).
21. Niu, K., Cao, X., & Yu, Y. (2021). Explainable student performance prediction with personalized attention for explaining why a student fails. arXiv preprint arXiv:2110.08268.
22. Li, Y., Liu, L., Wang, G., Du, Y., & Chen, P. (2022). EGNN: Constructing explainable graph neural networks via knowledge distillation. Knowledge-Based Systems, 241, 108345.
23. Qiang, M., Liu, Z., & Zhang, R. (2026). Explainable AI in education: integrating educational domain knowledge into the deep learning model for improved student performance prediction. Scientific Reports.