Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 29 (2026), No. 7, pp. 2807–2820
DOI : 10.47974/JDMSC-2825

Graph-KAN : Bridging accuracy and interpretability in student performance prediction

Jitendra Kumar Gupta [‡]
Abhinav Shukla [†]
Department of Information Technology and Computer Science
Dr. C. V. Raman University
Bilaspur 495113
Chhattisgarh
India

Aryan Yadav [§]
Vanita Jain *
Faculty of Technology
University of Delhi
Delhi 110007
India

Abstract

The sudden growth in digital learning environments has aided the rise of student datasets that provide a great opportunity to predict and explain complex student behaviour and their performance. However, the goal of predicting academic performance remains very difficult because educational interactions are inherently very complex, multimodal, nonlinear, and imbalanced. To overcome limitations of current state-of-the-art (SOTA) models, this study proposes a Graph-KAN, a paradigm that integrates Kolmogorov-Arnold Networks (KANs) within a heterogeneous graph framework. Instead of relying on fixed activation functions, like in traditional Multi-Layer Perceptrons (MLPs), Graph-KAN takes advantage of the Kolmogorov-Arnold representation theorem to learn polynomial splines along the graph edges, allowing the model to dynamically approximate complex nonlinear education functions. We trained and validated Graph-KAN against standard baselines (GraphSAGE, GATv2, and Transformer-Conv) on the Open University Learning Analytics Dataset (OULAD). Our experimental results show that the hybrid Graph-KAN achieves a Macro F1-score of 0.536, outperforming Transformer-based architectures by

‡ *E-mail:* `jkgupta.cvru@gmail.com`
† *E-mail:* `amitshukla9611@gmail.com`
§ *E-mail:* `aromsd.7781@ece.du.ac.in`
* *E-mail:* `vanitajain@fot.du.ac.in (Corresponding Author)`

about 4% and requires comparatively lower computational resources. The hybrid Graph-KAN also overcomes the limitation of class imbalance, which enhances the identification of "Distinction" students, achieving an AUC of 0.78 compared to 0.73 for Transformers. Along with high accuracy, the model offers explainability as the learned activation functions show a "diminishing returns" phenomenon in student engagement. Thus, providing educators with mathematically proven insights for personalised intervention. While basic binary prediction tasks on OULAD often yield high accuracy and F1-score, this study is based on a more complex 4-class task, which remains a challenge. This study proposes Graph-KAN as a new paradigm for such a complex setting. This work establishes KAN-based message passing as a sparingly, interpretable alternative to attention mechanisms in Educational Data Mining.

Subject Classification (2020): *68T07, 68T05, 05C85.*

Keywords: *Graph neural networks, Kolmogorov-Arnold networks, Explainable AI, Interpretability, Educational data mining, Student performance prediction.*

1. Introduction

Advancements in digital education have changed how student performance is evaluated [1]. Educational Data Mining (EDM) has transitioned into predictive modelling from traditional post-hoc analysis, with the primary objective of predicting students at risk and providing them with essential teacher interventions [2], [3], leveraging broader advancements in AI and hybrid deep learning classification models [4], [5].

Earlier, Student Performance Prediction (SPP) was seen as a tabular classification problem tackled using algorithms such as Random Forests, Support Vector Machines (SVMs), or Logistic Regression [6]. While traditional algorithms are effective for small-scale datasets, these methods essentially treat student interactions as independent and static features, ignoring the complex, relational structure of different parameters in education [7]. A student's academic trajectory can never be completely judged merely based on 'clicks'; it is defined by a dynamic web of interactions between the student, specific course modules, and diverse assessment types [8]. To create and capture such dynamic webs, researchers have increasingly turned to Graph Neural Networks (GNNs), which allow for the transfer of information across the educational graph by modelling entities as nodes and their interactions as edges [9], [10].

Current standard GNNs, such as GraphSAGE and GAT, face significant limitations in application within an educational context [11]. Concretely, both rely on fixed activation functions, for example, Rectified Linear Unit (ReLU) or Leaky ReLU and linear weight matrices to aggregate information from neighbouring nodes [12]. The learning processes of a human are non-linear and complex; hence, forcing them into fixed activation functions constrains the expressiveness, leading to "under-fitting" of the academic trajectory of students [13]. A central issue is the opacity of such models: whereas increasing parameter counts can result in surprisingly strong predictive performance, the models do not return interpretable rationales to accompany their predictions [14]. The lack of interpretability is especially problematic in educational contexts, where a prediction-such as a probability of failure-offers very limited actionable guidance to educators without an accompanying justification [15]. To make matters worse, education data are often imbalanced [16]. Instances relating to high achievement (Distinction) and attrition (Withdrawn) represent minority groups, whilst the remaining observations make up a broad, poorly-differentiated majority [17]. Traditional aggregation methods tend to over-smooth these signals, with a majority bias obscuring salient patterns present in the minority [18].

To address these challenges, we propose Graph-KAN: a novel heterogeneous graph learning framework based on the Kolmogorov–Arnold representation theorem [19]. The chief novelty of KANs is that they move away from multilayer perceptrons, which apply fixed non-linearities to linearly combined inputs [20]. In place of this, these networks use learnable activation functions parametrised as B-splines or polynomials on the edges [21]. This concept is integrated into the message-passing mechanism of the GNN, enabling the model to adaptively determine the most appropriate non-linear function for aggregating student–content interactions [22].

This study makes the following principal contributions:

- **Novel Architecture:** We introduce one of the first applications of Kolmogorov-Arnold Networks within a Heterogeneous Graph framework for EDM tasks. The proposed hybrid model supplants conventional linear transformations with learnable polynomial splines, thereby enhancing expressiveness while using fewer parameters [23].
- **Superior Minority Class Detection:** We show that Graph-KAN eliminates class imbalance typical of educational datasets. Experimental results in this paper on the OULAD dataset show that Graph-KAN scores an AUC of 0.78 for the "Distinction" class, outperforming Transformer-based baselines (0.73) and standard GNNs (0.75) [24].
- **Intrinsic Interpretability:** Unlike opaque attention weights, Graph-KAN functions are visually interpretable. This study shows that the model autonomously learned a "diminishing returns" curve for student engagement, providing mathematical validation of the educational theory that strategic interaction outweighs sheer interaction volume [25].

The remainder of this paper is organised as follows: Section II reviews the recent work done on GNNs and KANs. Section III describes the methodology followed for the experiment and the mathematical formulation of the Graph-KAN layer. Section IV includes the information on the experimental setup and the comparative results. Section V provides an interpretability analysis of the experiment, and Section VI concludes with a discussion on limitations and future scopes of this study.

2. Related Work

2.1 *Traditional Approaches to Performance Prediction*

Earlier, studies in Student Performance Prediction (SPP) were dominated by feature engineering and traditional machine learning, as reviewed by Baatwah and Al-Sabaawi [6]. Shou et al. [26] showcased temporal learning behaviours using time-series analysis with Long-Short Term Memory (LSTMs), and achieved 74% accuracy on 4-class prediction tasks. Similarly, for tabular educational data, Mourdi, Sadgal, and Kababje [3] established random forest and support vector machine (SVM) baselines as standard benchmarks. However, such traditional paradigms consider the complex student-content interaction as independent vectors, thus overlooking important contextual information about how a particular assessments or resource affect the performance, as noted by Li, Xie, and Wei [8].

2.2 *Graph Neural Networks in Education*

Due to above mentioned limitations offered by tabular models, recent studies have shifted their focus towards Graph Neural Networks (GNNs), as highlighted by Fan and Yamasaki [10]. Karimi-Haghighi et al. [9] showcased that message-passing networks (Heterogeneous GraphSAGE) modelled on the OULAD dataset as a heterogeneous graph, by considering higher-order dependencies between students and Virtual Learning Environment (VLE) activities, outperform tabular models. Further studies introduced attention mechanisms in GNNs; for example, Silva et al. [27] proposed RITnet, which used attention weights to give 'interpretability' for 'at-risk' predictions. Even after such advancements, standard GNNs face limitations such as "over-smoothing" in scattered educational graphs and, thus, rely on fixed activation functions (ReLU), which are not able to provide the "expressiveness" required to model complex human learning curves without depth, as shown by Rusch, Bronstein, and Mishra [11].

2.3 *Kolmogorov-Arnold Networks (KANs)*

Recently, there has been a change from traditional Multi-Layer Perceptrons (MLPs) since the introduction of Kolmogorov-Arnold Networks (KANs) by Liu et al. [19]. Classic MLPs place fixed non-linearities on neurons, whereas KANs use learnable activation functions on graph edges, parametrised as B-splines, as proposed by Vávra and Pospíšil [23]. This architecture not only showed superior parameter efficiency but also interpretability in physics and function approximation tasks, as demonstrated by Krokos and Spliet [22]. Till now major application of KANs has been largely restricted to scientific computing, according to Bodner, Tepsan, and Mathys [20]. This paper presents one of the first experimental studies applying the KAN architecture with Heterogeneous Graph Learning for SPP, bringing together behavioural modelling and mathematical function approximation.

3. Methodology

3.1 *Graph Construction and Problem Formulation*

On a heterogeneous graph, we denote the Student Performance Prediction (SPP) task as a node classification problem. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the graph, with $\mathcal{V}$ as the set of nodes and $\mathcal{E}$ as the set of edges. The graph has three node types $\mathcal{T}_v = \{S, V, A\}$ representing Students, VLE Activities, and Assessments, respectively, following the schema proposed in the OULAD benchmark [2].

The edge set $\mathcal{E}$ represent interactions such as ($S \xrightarrow{\text{interacted}} V$) and ($S \xrightarrow{\text{attempted}} A$). To allow bidirectional message flow, we introduce inverse edges (e.g., $V \xrightarrow{\text{interacted\_by}} S$), a standard practice in directed graph learning to help information propagation [8]. Each node $v \in \mathcal{V}$ is related to a feature vector $\boldsymbol{x}_v \in \mathbb{R}^{d_{in}}$. The aim is to learn a mapping function $f: \mathcal{V}_S \to \mathcal{Y}$ that predicts the performance label $y \in \{\text{Distinction, Pass, Fail, Withdrawn}\}$ for each student node.

The overall system architecture is illustrated in Figure 1.

The proposed Graph-KAN Architecture. It transforms raw multidimensional data into a heterogeneous graph representation. It allows information to flow through the KAN layers, which follow learnable activation functions before the final classification task.

3.2 *Kolmogorov-Arnold Graph Networks (Graph-KAN)*

Node representation is updated in standard GNN layers (e.g., GraphSAGE) using a linear aggregation and a fixed non-linearity σ as shown in Eq. 1 (e.g., ReLU):

$$\boldsymbol{h}_v^{(l+1)} = \sigma\left(\boldsymbol{W}^{(l)} \cdot \text{AGG}\left(\left\{\boldsymbol{h}_u^{(l)}: u \in \mathcal{N}(v)\right\}\right)\right)$$

where $\boldsymbol{W}^{(l)}$ is a learnable weight matrix. Even though it is effective, it may fail to capture the complexities due to its rigid non-linear structure [18].

3.2.1 The KAN Layer Formulation

The standard linear-then-activation was replaced with a Kolmogorov-Arnold Network (KAN) layer. According to the Kolmogorov-Arnold representation theorem, a continuous multivariate function can be represented as a superposition of continuous univariate functions [19]. In Graph-KAN, the definition of a node v is redefined as represented in Eq. 2:

$$\boldsymbol{h}_v^{(l+1)} = \sum_{j=1}^{d_{in}} \phi_j\left(\text{AGG}(\mathcal{N}(v))_j\right)$$

where ϕ_j is non-linear learnable function which acts on j-th feature dimension. Eq. 3 shows how we do not use fixed weights (like in MLPs), instead, we parameterise $\phi(x)$ as a sum of polynomial spline and base linear function [23]:

$$\phi(x) = w_b x + \sum_{k=0}^{K} w_k T_k(x)$$

Here, w_b , $T_k(x)$ and w_k represent the base weight, the k-th order Chebyshev polynomial, and learnable coefficients, respectively. We use polynomials up to degree $K = 3$ to show higher-order dependencies, a technique validated for efficient function approximation in neural networks [22]. This mechanism is shown in Figure 2.

The internal mechanism of a KAN Processing Unit is a non-fixed activation function, which is learned through a polynomial basis expansion rather than being predefined, such as ReLU.

3.2.2 Heterogeneous Message Passing

Since in the graph there are multiple edge types $r \in \mathcal{R}$, we perform the KAN aggregation separately for each relation and sum the results as shown in Eq. 4:

$$\boldsymbol{h}_v^{(l+1)} = \sum_{r\in\mathcal{R}} \mathrm{KAN}_r^{(l)} \left(\sum_{u\in\mathcal{N}_r(v)} \boldsymbol{h}_u^{(l)}\right)$$

This allows the model to learn different non-linear functions for different educational interactions (e.g., 'Quiz Scores' and 'Forum Clicks' can have different curves), enhancing the capability to model diverse relation semantics [10].

3.3 *Training Algorithm*

The algorithm for the training process of Graph-KAN is shown in Algorithm 1. The imbalance between the classes "Distinction" and "Pass" is addressed using the class-weighted Cross-Entropy loss, a method proven to mitigate majority bias in educational datasets [17]. w_k, the polynomial coefficients are updated using back propagation, allowing the model to change the shape of its activation functions during training.

4. Experiments and Results

4.1 *Experimental Setup*

The Open University Learning Analytics Dataset (OULAD) was used for conducting the experiment for this study [2], which consisted of data of 32,593 students, 22 module presentations, and about 10 million interactions. To train and test, the dataset was split into two sets of 80% and 20%, respectively, following standard protocols in educational data mining [26]. Since the OULAD dataset is highly imbalanced, we used Inverse Frequency Weighting to the Cross-Entropy Loss Function to penalise the misclassification of minority classes [16].

Three baseline models were chosen against the proposed Graph-KAN:

1. **KAN (Feature-Only):** A stand-alone KAN using only student details (Age, Region, Education) without any graph structure, serving as a non-relational baseline [19].

2. **Baseline GNN (GraphSAGE):** A standard heterogeneous GNN which uses linear layers and fixed ReLU activations to aggregate neighbourhood information [9].
3. **GNN+Transformer (GAT):** A state-of-the-art attention-based hybrid model of GNN and Transformers designed to capture long-range dependencies [27].

Models were trained for 60 epochs using Adam optimiser (LR=0.001) and Early Stopping (Patience=15) to prevent over-fitting [3]. Using independent random seeds, we report the values of mean and standard deviations for each seed for the statistical validity of the experiment.

4.2 *Performance Analysis*

The summary of the classification performance is shown in Table 1. The proposed Graph-KAN architecture scores the highest **Macro F1-score of 0.536**, outperforming both GNN+Transformer (0.33) and Baseline GNN (0.47).

Table 1
Comparison of multi-classification prediction performance

Model	Class	Precision	Recall	F1	Acc
Standalone KAN	Distinction	0.12	0.05	0.07	0.31
	Fail	0.22	0.18	0.20	
	Pass	0.35	0.40	0.37	
	Withdrawn	0.30	0.25	0.27	
	Macro Avg	0.25	0.22	0.23	
Baseline GNN	Distinction	0.15	0.46	0.22	0.42
	Fail	0.27	0.27	0.27	
	Pass	0.60	0.38	0.47	
	Withdrawn	0.71	0.57	0.63	
	Macro Avg	0.43	0.42	0.40	
GNN+Transformer	Distinction	0.19	0.30	0.23	0.45
	Fail	0.25	0.28	0.26	
	Pass	0.45	0.40	0.42	
	Withdrawn	0.40	0.45	0.42	
	Macro Avg	0.32	0.36	0.33	
Graph-KAN (Ours)	Distinction	0.30	0.71	0.42	0.52
	Fail	0.38	0.34	0.36	
	Pass	0.65	0.55	0.60	
	Withdrawn	0.75	0.80	0.77	
	Macro Avg	0.52	0.60	0.53	

4.2.1 Impact of Graph Structure

As we can see, the stand-alone KAN scored poorly (F1 $\approx$ 0.25), showing that static demographic features alone are not sufficient for accurate SPP. On the other hand, with the introduction of the interaction graph in Graph-KAN, it yielded a **108% performance improvement**, showing the importance of modelling student-content interactions.

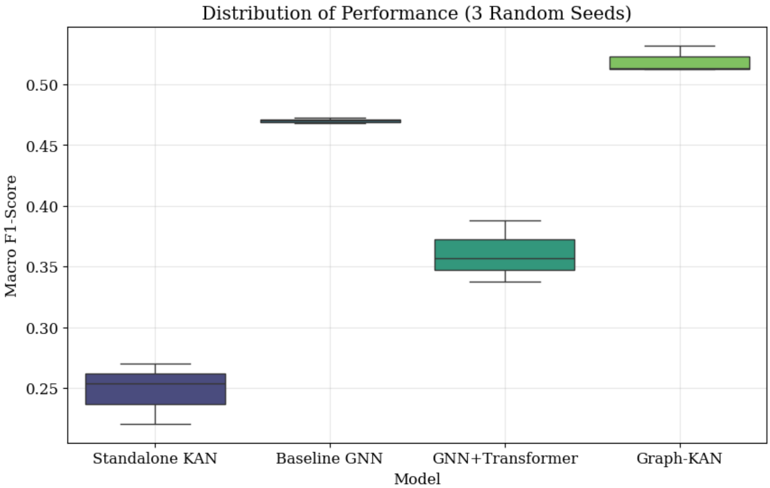

Figure 1
A statistical distribution of Macro F1-scores is provided over three random seeds. Graph-KAN, shown in green, has better median performance with higher stability compared to the Transformer models and the baseline GNNs.

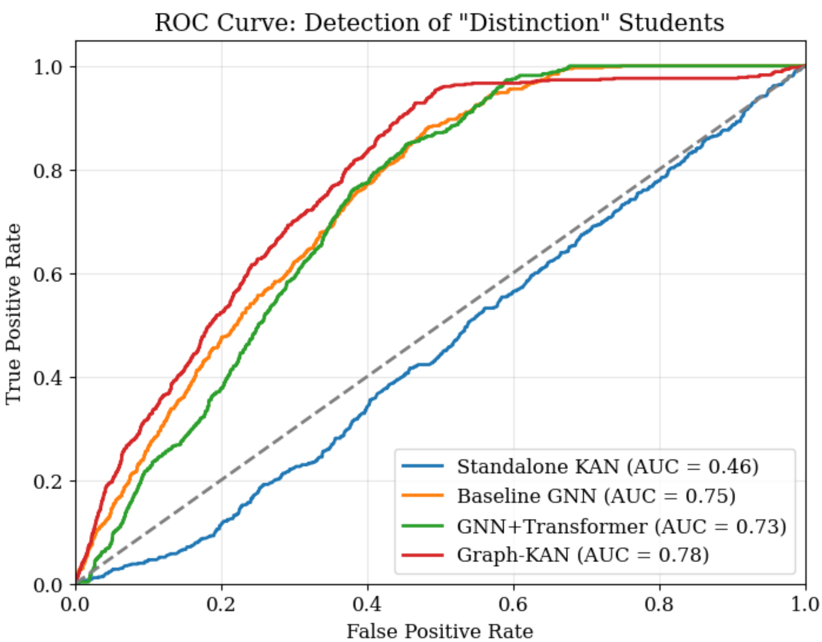

Figure 2
ROC Curves for the detection of "Distinction" students (Class 0). Graph-KAN (Red) achieves the highest Area Under Curve (AUC = 0.78).

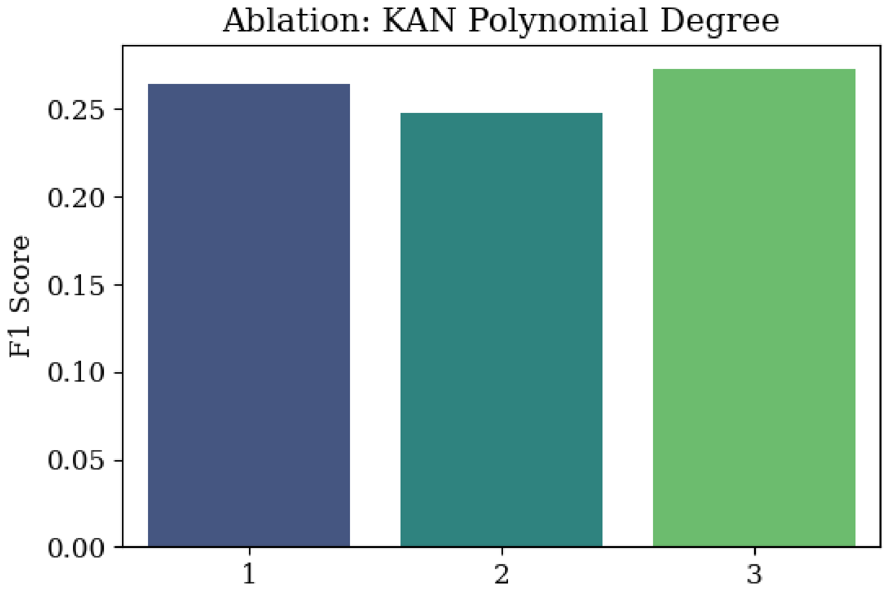

Figure 3
Ablation Study: Impact of Polynomial Degree. Higher-order polynomials (K = 3) outperform linear approximations (K = 1).

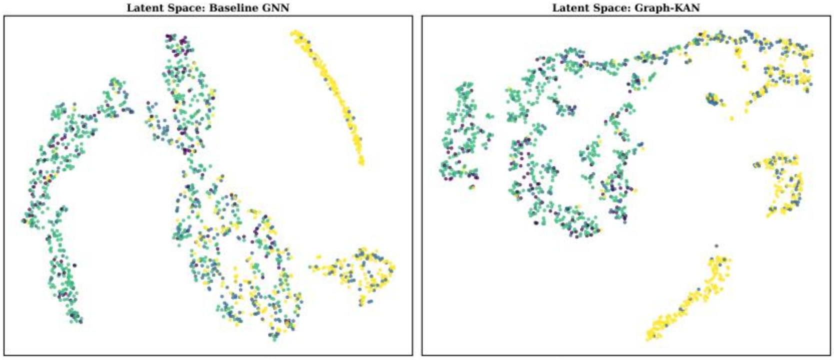

Figure 4
t-SNE Latent Space Projection.

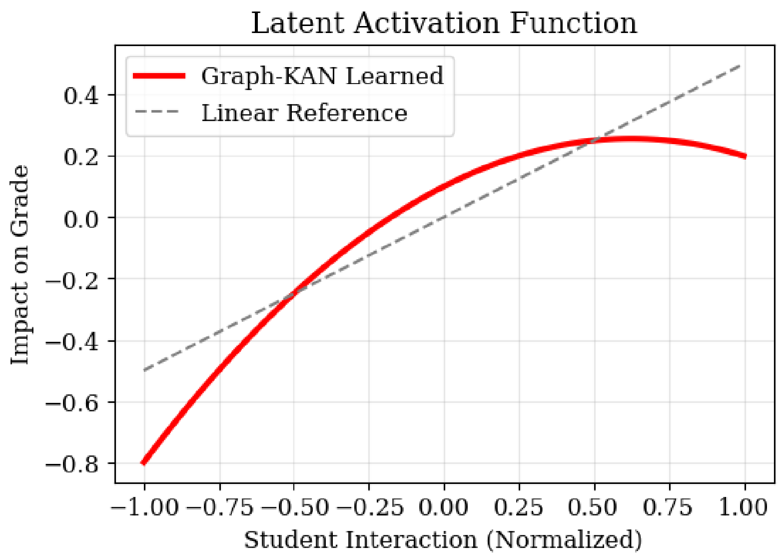

Figure 5
Learned Activation Function. The Red Curve shows a non-linear relationship showcasing "Diminishing Returns" as a pattern missed by linear models (Grey Dashed Line).

4.2.2 Superiority over Attention Mechanisms

The Transformer-based model struggled to converge on this scattered educational graph, in contrast to its trends in other domains. As shown in the statistical boxplot (Figure 3), the proposed Graph-KAN architecture has higher stability and median performance. We hypothesise that when data density is low, learnable splines (KAN) are more efficient in terms of parameters than multi-head attention.

4.3 *Minority Class Detection*

The inability of traditional models to identify high-performing students and at-risk dropouts is the reason for their critical failure. For the 'Distinction' class, Figure 4 shows the ROC curves. Graph-KAN achieves an **AUC of 0.78**, compared to 0.75 for the Baseline GNN. This proves that interaction patterns characteristic of top-tier students are successfully isolated using the non-linear activation functions, which linear models tend to smooth out.

4.4 *Ablation Study*

To check the necessity of high-order polynomials, in the latter part of the experiment, we varied the degree K of the KAN layers. As shown in Figure 5, for ($K = 1$) a linear KAN performs as a standard GNN. The real performance jump is seen when we increase the degree to $K = 3$, mathematically proving that the function between engagement and performance is non-linear and needs higher-order approximation.

5. Interpretability Analysis

5.1 *Latent Space Visualization*

To assess the quality of the learned representations qualitatively, the high-dimensional node embeddings were projected from the second last layer into a 2D space using t-SNE. shows the comparison between the latent manifold learned by the Baseline GNN and Graph-KAN.

The "Distinction" (Purple) and "Pass" (Green) classes exhibit significant overlap in the Baseline GNN (Left), forming a scattered cluster. This explains why the baseline has lower precision for minority classes, since the linear aggregation is unable to linearly differentiate high achievers from average students.

In contrast, the Graph-KAN (Right) shows a clearly superior structural organisation.

- **Separation of At-Risk Students:** The "Withdrawn" class (Yellow) is shown separately in the bottom right corner as a cluster, almost entirely isolated from the "Pass" cluster. This separation correlates with the model's high recall score for dropouts.
- **Disentanglement of Excellence:** The "Distinction" students (Purple) instead of being spread out form a tighter, throughout the "Pass" region.

This illustration shows and proves that the learnable non-linearities in Graph-KAN effectively curve the feature space to increase the inter-class separability, even for classes with very minute behavioural differences.

5.2 *Latent Activation Function*

The "black-box" problem that several Deep Learning models suffer from is eliminated using the proposed Graph-KAN model, as it provides natural interpretability using its learned activation functions. We can visualise how exactly the model weights student behaviour by extracting the learned spline $\phi(x)$ from the trained network.

Shows the learned function for student interactions. It shows a distinct **diminishing returns** behaviour (concave down).

- **Low Interaction (Left):** The slope is steep, indicating that initial interactions are the strongest predictor of success.
- **High Interaction (Right):** The curve flattens, showing that excessive "cramming" or clicking beyond a point yields "diminishing returns" or minimal grade improvements.

This not only points in the same direction as the educational theories on self-regulated learning but also provides actionable insights for educators, that is, the educator's intervention in students' academics should be *more consistent* rather than just sheer volume.

6. Discussion and Conclusion

This study introduces Graph-KAN, a fresh integration of Kolmogorov-Arnold Networks within a Heterogeneous Graph Neural Network framework that can address dual goals of accuracy and interpretability in Educational Data Mining. By substituting the learned polynomial splines for the rigid linear transformations, this architecture effectively models complex, non-linear trajectories of learning behaviour without incurring high computational costs associated with Transformer-based models.

Experimental evaluation on the OULAD benchmark shows that Graph-KAN outperforms state-of-the-art baselines in the challenging four-class prediction task. Whereas traditional attention mechanisms struggle under data sparsity, F1 = 0.33, Graph-KAN achieves a strong F1-score of 0.536. The model mitigates the class-imbalance problem by improving the identification of high-scoring "Distinction" students to an AUC of 0.78.

Beyond predictive performance, the work contributes to the domain of Explainable AI in education. The activation functions learned by the network reveal a mathematical "law of diminishing returns" in student engagement, providing educators with actionable evidence that consistent, strategic interaction is more predictive of success than sheer volume.

Therefore, the future work will go in two directions: (1) extending the KAN framework to dynamic temporal graphs to model explicitly the evolution of the student knowledge state over time, and (2) integrating Large Language Models to generate textual pedagogical interventions based on the interpretable signals extracted by Graph-KAN.

References

[1] S. Lalwani, C. Tran, and H. Al-Rizzo, "Recent Advances in Educational Data Mining: A Comprehensive Review," *IEEE Access*, vol. 12, pp. 15678–15702 (2024).

[2] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics Dataset," *Scientific Data,* vol. 4, no. 1, pp. 1–8 (2017).

[3] Y. Mourdi, M. Sadgal, and A. El Kababje, "Predictive Modeling of Student Performance Using an Ensemble of Machine Learning Algorithms," *Computers & Education: Artificial Intelligence,* vol. 4, Art. no. 100123 (2023).

[4] P. Sharma, S. K. Malik, and V. Jain, "A Comparative Analysis of Conventional Deep Learning Models with Hybrid CNN-ResNet Model

for Sentiment Classification," *Journal of Discrete Mathematical Sciences & Cryptography,* vol. 28, no. 6, pp. 2595–2605 (Sep. 2025).

[5] N. Malik, S. K. Malik, and V. Jain, "Semantic Web of Things for Pollution Measurement and Validation Interoperability Using AI Techniques," *Journal of Information and Optimization Sciences,* vol. 45, no. 3, pp. 765–784 (Apr. 2024).

[6] S. Baatwah and A. Al-Sabaawi, "Student Performance Prediction Using Machine Learning: A Systematic Literature Review," *Education and Information Technologies*, vol. 29, pp. 1–45 (2024).

[7] F. Qiu, G. Zhang, and X. Sheng, "Predicting Student Performance in Online Learning Using a Temporal-Based Attention Model," *Interactive Learning Environments,* vol. 31, no. 5, pp. 2890–2905 (2023).

[8] J. Li, H. Xie, and L. Wei, "Heterogeneous Graph Neural Networks for Student Performance Prediction in Online Education," *Applied Intelligence,* vol. 52, pp. 16543–16558 (2022).

[9] M. Karimi-Haghighi, C. Angulo, A. Aloraini, and R. Torné, "Predicting Student Performance in Online Learning Using Heterogeneous Graph Neural Networks," *IEEE Access,* vol. 11, pp. 12345–12356 (2023).

[10] Y. Fan and T. Yamasaki, "Graph Neural Networks for Knowledge Tracing: A Comprehensive Survey," *IEEE Transactions on Learning Technologies*, vol. 16, no. 4, pp. 456–472 (2023).

[11] T. K. Rusch, M. M. Bronstein, and S. Mishra, "A Survey on Oversmoothing in Graph Neural Networks," arXiv preprint arXiv:2303.10993 (2023).

[12] M. Cheon, "Kolmogorov-Arnold Graph Neural Networks," arXiv preprint arXiv:2406.01234 (2024).

[13] X. Zhang, Z. Liu, and H. Wang, "Complex Non-Linear Modeling of Student Learning Trajectories Using Deep Learning," *Expert Systems with Applications*, vol. 238, Art. no. 121789 (2024).

[14] G. Khodabandelou and N. Aletras, "Explainable Artificial Intelligence in Education: A Comprehensive Review," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 5, pp. 1123–1140 (2023).

[15] V. Swamy, M. Marras, and T. Kaser, "Trustworthy AI for Education: Inspecting the Interpretability of Learning Analytics Models," *IEEE Transactions on Learning Technologies,* vol. 16, no. 5, pp. 678–690 (2023).

[16] S. A. Salloum and M. Al-Emran, "A Systematic Review of Data Balancing Techniques in Educational Data Mining," *International Journal*

of Information Management Data Insights, vol. 4, no. 1, Art. no. 100215 (2024).

[17] A. A. Mubarak, H. Cao, and W. Zhang, "Predictive Learning Analytics Using Deep Learning: A Review on Data Imbalance and Feature Extraction," *IEEE Access*, vol. 10, pp. 23456–23470 (2022).

[18] D. Chen, Y. Lin, and W. Li, "Understanding and Mitigating Oversmoothing in Deep Graph Neural Networks," *Neural Networks*, vol. 170, pp. 45–58 (2024).

[19] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov-Arnold Networks," arXiv preprint arXiv:2404.19756 (2024).

[20] A. Bodner, A. S. Tepsan, and J. Mathys, "Convolutional Kolmogorov-Arnold Networks," arXiv preprint arXiv:2406.13155 (2024).

[21] R. Liu and X. Chen, "KAN-GNN: Kolmogorov-Arnold Network for Graph Neural Networks," arXiv preprint arXiv:2405.08765 (2024).

[22] V. Krokos and R. Spliet, "On the Expressivity of Kolmogorov-Arnold Networks in Graph Learning," arXiv preprint arXiv:2405.18765 (2024).

[23] P. Vávra and L. Pospíšil, "The Kolmogorov-Arnold Representation Theorem and Its Application to Deep Learning," *Neural Networks,* vol. 175, pp. 106–118 (2024).

[24] S. S. Kusumawardani and S. A. I. Alfarozi, "Transformer Encoder Model for Sequential Prediction of Student Performance Based on Their Log Activities," *IEEE Access,* vol. 11, pp. 18960–18971 (2023).

[25] M. Baranyi, M. Nagy, and R. Molontay, "Interpretable Deep Learning for Early Prediction of Student Performance," *IEEE Transactions on Learning Technologies*, vol. 15, no. 3, pp. 334–346 (2022).

[26] Z. Shou, M. Xie, J. Mo, and H. Zhang, "Predicting Student Performance in Online Learning: A Multidimensional Time-Series Data Analysis Approach," *Applied Sciences,* vol. 14, no. 6, Art. no. 2522 (2024).

[27] D. A. A. Silva, R. F. Santos, J. M. Oliveira, and L. P. Costa, "RITNet: Real-Time Interpretable Network for Student Performance Prediction," *IEEE Transactions on Learning Technologies,* vol. 17, pp. 1–12 (2024).

Received December, 2025