

Original Research Article

Performance analysis of CNN-KNN co-learning model for adaptive student learning evaluation in higher education settings

Yuan Kang^{1,2}, M. Kazem Chamran^{3*}

¹City University Malaysia, Kuala Lumpur, 46100, Malaysia

²Sanya City Vocational College, Sanya, Hainan, 572000, China

³Faculty of Information Technology, City University Malaysia, Kuala Lumpur, 46100, Malaysia

Abstract: This study evaluates a convolutional neural network-k-nearest neighbors (CNN-KNN) co-learning model for adaptive student learning evaluation in higher education. A one-dimensional CNN learns a latent representation, while distance-weighted KNN corrects low-confidence cases through a validation-selected gate. Experiments use the public Predict Students' Dropout and Academic Success dataset with 4,424 student records and 36 demographic, socioeconomic, macroeconomic, and academic variables. Three evidence stages are considered: Enrollment, end of Semester 1, and end of Semester 2. Across three stratified splits, the full-stage CNN-KNN achieved 0.740 ± 0.023 accuracy and 0.691 ± 0.017 macro-F1. Accuracy was slightly above standalone CNN (0.737) and clearly above raw KNN (0.691), although Random Forest remained strongest (0.769 accuracy; 0.708 macro-F1). Accuracy increased from 0.577 at enrollment to 0.721 after Semester 1 and 0.740 after Semester 2. The hybrid is therefore better viewed as a selective, stability-oriented decision-support layer than as a universally superior classifier, while the staged design supports better-timed academic advising and student-support allocation.

Keywords: adaptive evaluation; educational data mining; CNN-KNN; learning analytics; higher education management; student success

1. Introduction

Higher education institutions increasingly use educational data mining and learning analytics to support retention, program quality, and evidence-based student services. A recent review found rapid growth in machine learning-based performance prediction but persistent problems with heterogeneous datasets, inconsistent metrics, and limited generalizability ^[1]. Conventional classifiers can still extract useful signals from compact academic records ^[2], while online-course studies use demographic, assessment, and interaction data to predict outcomes such as grades or dropout ^[3]. The Portuguese higher-education dataset of Realinho et al. similarly links demographic, socioeconomic, macroeconomic, and semester-level variables to graduate, enrolled, and dropout outcomes ^[4].

The value of student prediction depends on when evidence becomes available and how predictions inform action. Multimodal learning analytics can reveal behavioral and contextual patterns beyond a single grade snapshot ^[5]. Ouyang et al. combined AI-based prediction with learning-analytics feedback, emphasizing timely teaching decisions ^[6]. Holicza and Kiss showed that model performance varies with the information environment ^[7], while Xue and Niu proposed multi-output hybrid models to characterize learning progress at multiple points ^[8]. These findings motivate adaptive evaluation as the available evidence evolves during the academic process. The prediction target may range from final grades to persistence or dropout risk, making timing and evidence selection central to practical deployment.

A second challenge is model complementarity. Higher-education data are often tabular, moderately sized, and class-imbalanced, so no learner dominates every metric or subgroup. Smadi et al. showed that several machine-learning methods can support competency-based evaluation, with the preferred model depending on the objective ^[9]. Christou et al. reported benefits from feature construction and machine

learning for early performance and dropout prediction, while stressing the distinction between prediction and actionable support ^[10]. Optimized boosting and imbalance-aware processing have also improved early failure detection ^[11]. Recent studies further demonstrate useful signals from longitudinal clickstream data ^[12] and multiple measurable student and learning-context factors ^[13].

Recent literature highlights both opportunity and caution. A 2025 rapid review found substantial potential for university performance prediction but uneven reliability, generalizability, and methodological quality ^[14]. Multi-model fusion can support academic early warning when component learners provide complementary information ^[15]. Accordingly, this study analyzes rather than assumes the superiority of CNN-KNN co-learning. The CNN learns a compact representation of ordered institutional features, and KNN performs local inference in that latent space when CNN confidence is low. Here, "co-learning" refers to this coupled representation-and-local-decision mechanism, not joint gradient updates.

This study makes three contributions: A staged protocol using only information available at enrollment, after Semester 1, or after Semester 2; a comparison with logistic regression, raw-feature KNN, Random Forest, multilayer perceptron (MLP), and standalone CNN under repeated stratified splits; and a management-oriented interpretation that distinguishes early screening from later evidence-rich evaluation while reporting cases in which conventional ensembles perform better.

2. Materials and methods

2.1. Dataset and adaptive evidence stages

The study uses the public Predict Students' Dropout and Academic Success dataset ^[4] from the UCI Machine Learning Repository ^[16]. It contains 4,424 records, 36 predictors, and three classes: Graduate (2,209), Dropout (1,421), and Enrolled (794), with no reported missing values. To avoid temporal leakage, predictors were grouped by availability: Enrollment uses 24 pre-semester and macroeconomic variables, Semester 1 adds six first-semester curricular variables, and Semester 2 adds six second-semester variables. **Table 1** summarizes the stages. Each stage was standardized using training-set means and standard deviations only.

Table 1. Adaptive evidence stages used for student learning evaluation.

Stage	Feature count	New evidence introduced	Decision use
Enrollment	24	Admission, demographic, socioeconomic, macroeconomic	Initial support screening
Semester 1	30	First-semester curricular activity and grades	Early academic intervention
Semester 2	36	Second-semester curricular activity and grades	Updated status evaluation

2.2. CNN-KNN co-learning model

Figure 1 shows the model flow. At each evidence stage, a one-dimensional CNN processes the standardized feature vector in a fixed conceptual order. Two convolutional layers with 32 and 64 filters (kernel size 3, padding 1) use ReLU activations. The flattened output passes through a 64-unit hidden layer with 0.20 dropout to form a 32-dimensional latent vector. Equation (1) represents the layer- l convolution, where K is the kernel width and ReLU is the rectified linear unit.

$$h_i^{(l)} = \text{ReLU} \left(\sum_{j=0}^{K-1} w_j^{(l)} x_{i+j}^{(l-1)} + b^{(l)} \right) \quad (1)$$

The CNN output estimates a three-class Softmax posterior. A distance-weighted KNN is fitted to the CNN's 32-dimensional training embeddings. For test embedding z , the local posterior for class c is computed from its k nearest training embeddings by Equation (2). Thus, KNN operates on CNN-learned neighborhood structure rather than raw-feature Euclidean distance.

$$P_{KNN}(y = c|z) = \frac{\sum_{n \in \mathcal{N}_k(z)} d(z, z_n)^{-1} \mathbf{1}_{(y_n=c)}}{\sum_{n \in \mathcal{N}_k(z)} d(z, z_n)^{-1}} \quad (2)$$

A confidence gate selects the CNN decision or KNN correction. Let q be the maximum CNN posterior; Equation (3) defines the final rule. The neighborhood size k and threshold τ maximize validation macro-F1.

Across the three splits, selected k values were 41, 11, and 7, with τ values of 0.50, 0.45, and 0.45.

$$\hat{y} = \begin{cases} \operatorname{argmax}_c P_{CNN}(y = c|z), & q \geq \tau \\ \operatorname{argmax}_c P_{KNN}(y = c|z), & q < \tau \end{cases} \quad (3)$$

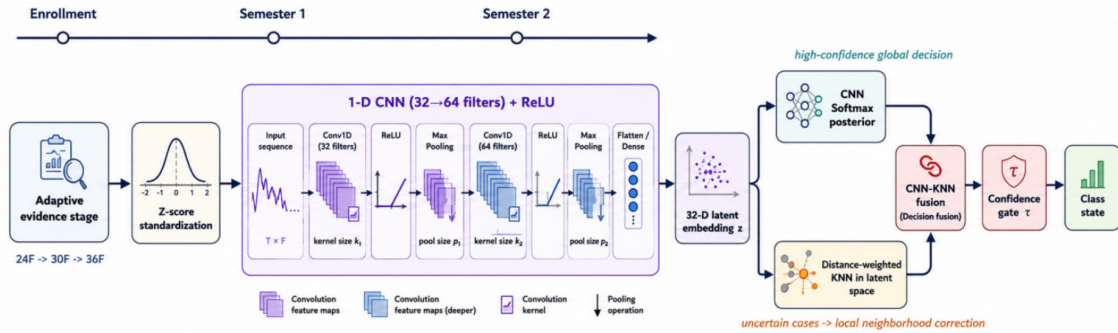


Figure 1. Adaptive CNN-KNN co-learning architecture with confidence-gated local correction.

2.3. Experimental protocol and evaluation

Three stratified splits used random seeds 13, 47, and 97, with 70% training, 15% validation, and 15% testing. Standardization was fitted on training data only. CNN training used class-weighted cross-entropy, AdamW (learning rate 0.001; weight decay 0.0001), batch size 64, up to 60 epochs, and early stopping with patience 8 on validation macro-F1. Baselines used the same training partition and fixed settings: Balanced logistic regression; distance-weighted KNN with $k=11$; Random Forest with 300 trees and minimum leaf size 2; and an MLP with 64 and 32 hidden units. Accuracy, macro-precision, macro-recall, and macro-F1 were evaluated on untouched test sets. Equation (4) defines macro-F1, which gives equal weight to all classes.

$$F1_{macro} = \frac{1}{C} \sum_{c=1}^C \frac{2 Precision_c Recall_c}{Precision_c + Recall_c} \quad (4)$$

3. Results

3.1. Comparative model performance

Table 2 gives mean $\pm$ standard deviation over the three test splits. Random Forest achieved the highest accuracy (0.769) and macro-F1 (0.708), so CNN-KNN was not the overall best classifier. CNN-KNN nevertheless raised accuracy from 0.737 for standalone CNN and 0.691 for raw KNN to 0.740. Its macro-F1 standard deviation (0.017) was also about 25% lower than CNN's (0.022). MLP achieved higher accuracy than CNN-KNN but lower macro-F1, reinforcing the need to consider more than accuracy alone.

Table 2. Test performance across three stratified random splits (mean $\pm$ SD).

Model	Accuracy	Macro-precision	Macro-recall	Macro-F1
Logistic regression	0.729 $\pm$ 0.009	0.699 $\pm$ 0.014	0.703 $\pm$ 0.016	0.692 $\pm$ 0.014
KNN	0.691 $\pm$ 0.005	0.637 $\pm$ 0.015	0.574 $\pm$ 0.005	0.581 $\pm$ 0.002
Random forest	0.769 $\pm$ 0.012	0.724 $\pm$ 0.011	0.699 $\pm$ 0.015	0.708 $\pm$ 0.014
MLP	0.752 $\pm$ 0.009	0.684 $\pm$ 0.033	0.656 $\pm$ 0.024	0.660 $\pm$ 0.033
CNN	0.737 $\pm$ 0.025	0.695 $\pm$ 0.020	0.700 $\pm$ 0.021	0.693 $\pm$ 0.022
CNN-KNN co-learning	0.740 $\pm$ 0.023	0.696 $\pm$ 0.018	0.695 $\pm$ 0.013	0.691 $\pm$ 0.017

3.2. Adaptive performance across evidence stages

Figure 2 shows a clear temporal effect. CNN-KNN accuracy/macro-F1 increased from 0.577 $\pm$ 0.001/0.529 $\pm$ 0.004 at enrollment to 0.721 $\pm$ 0.005/0.669 $\pm$ 0.006 after Semester 1 and 0.740 $\pm$ 0.023/0.691 $\pm$ 0.017 after Semester 2. Most improvement therefore occurred after the first semester, with a smaller second-semester gain. Enrollment outputs are better treated as low-confidence screening signals, while stronger intervention decisions can rely on later academic evidence.

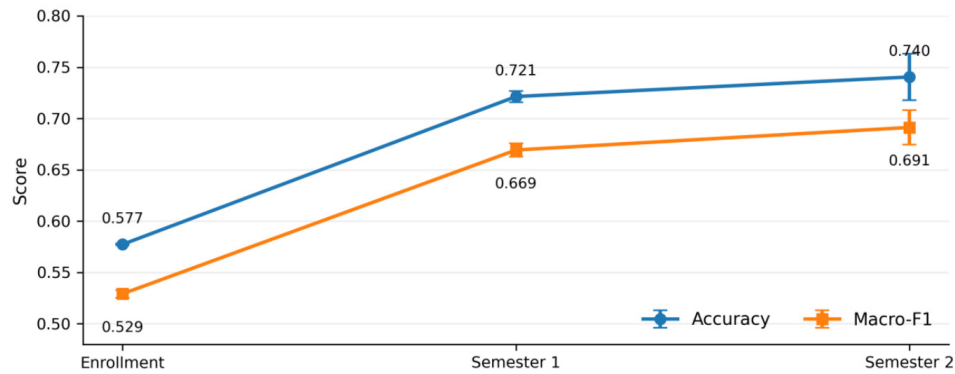


Figure 2. CNN-KNN accuracy and macro-F1 as evidence accumulates across academic stages (error bars: SD across three splits).

4. Discussion

The results suggest modest but useful complementarity between global and local learners. Because CNN-KNN improved standalone CNN accuracy by only 0.3 percentage points and the gate activated infrequently, KNN mainly corrected boundary cases. Lower macro-F1 variability indicates greater stability, not superior mean macro-F1; Random Forest remained stronger on both principal metrics, consistent with the advantages of tree ensembles for heterogeneous tabular data. CNN-KNN is therefore most relevant when local similarity is operationally useful, such as comparing an uncertain student with nearby historical profiles in the learned space.

For higher education management, the stage effect is more important than small classifier differences. Enrollment-only predictions were not accurate enough for high-stakes action, whereas Semester 1 evidence markedly improved accuracy and macro-F1 and can better support targeted advising, tutoring, financial-support review, or workload adjustment. Predictions should prioritize supportive contact rather than automate exclusionary decisions. Human review remains important, particularly for the ambiguous Enrolled class, and institutions should monitor subgroup performance, calibration, concept drift, and intervention outcomes before routine deployment. This is important because prediction errors can affect resource allocation as well as individual advising decisions.

From a business-administration perspective, staged prediction can support different intervention portfolios. Enrollment estimates can inform capacity planning and voluntary orientation; Semester 1 estimates can prioritize advising, tutoring, and scholarship or payment follow-up; and Semester 2 estimates can support program-level retention review. Deployment should define access, refresh frequency, permissible support actions, and adviser override procedures. Predictive scores create managerial value only when they improve the timing or allocation of institutional support.

The study has several limitations. It uses one institutional dataset with structured administrative rather than fine-grained learning-process features, limiting external validity. One-dimensional convolution also imposes locality on a manually ordered feature vector, so other orderings may alter learned interactions. Only three random splits were used for a compact, reproducible analysis rather than exhaustive tuning. The sequential mechanism learns CNN representations before KNN uses them; future work could examine differentiable neighborhood objectives, calibrated gating, cost-sensitive decisions, and external validation across institutions.

5. Conclusion

This study analyzed CNN-KNN co-learning for adaptive student evaluation on a public higher-education dataset. At the full evidence stage, the hybrid achieved 0.740 accuracy and 0.691 macro-F1. It modestly improved CNN accuracy and reduced macro-F1 variability, while Random Forest remained the strongest benchmark. More importantly, prediction quality increased sharply after Semester 1. The practical contribution is therefore a timing-aware decision-support framework: Use early predictions for screening,

update risk as evidence accumulates, and reserve consequential actions for human-reviewed, evidence-rich decisions.

Acknowledgements

This work was supported by the Research Project "Personalized Learning Path Recommendation Integrating Real-Time Feedback and Multidimensional Features" (No. Hnky2026ZD-20).

About the author

*Corresponding author: M. Kazem Chamran.

References

- [1] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, 11(9), 552. <https://doi.org/10.3390/educsci11090552>.
- [2] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9, Article 11. <https://doi.org/10.1186/s40561-022-00192-z>.
- [3] Alhothali, A., Albsisi, M., Assalahi, H., & Aldosemani, T. (2022). Predicting Student Outcomes in Online Courses Using Machine Learning Techniques: A Review. *Sustainability*, 14(10), 6199. <https://doi.org/10.3390/su14106199>.
- [4] Realinho, V., Machado, J., Baptista, L., & Martins, M. V. (2022). Predicting Student Dropout and Academic Success. *Data*, 7(11), 146. <https://doi.org/10.3390/data7110146>.
- [5] Liao, C.-H., & Wu, J.-Y. (2022). Deploying multimodal learning analytics models to explore the impact of digital distraction and peer learning on student performance. *Computers & Education*, 190, 104599. <https://doi.org/10.1016/j.compedu.2022.104599>.
- [6] Ouyang, F., Wu, M., Zheng, L., Zhang, L., & Jiao, P. (2023). Integration of artificial intelligence performance prediction and learning analytics to improve student learning in online engineering course. *International Journal of Educational Technology in Higher Education*, 20, Article 4. <https://doi.org/10.1186/s41239-022-00372-4>.
- [7] Holicza, B., & Kiss, A. (2023). Predicting and Comparing Students' Online and Offline Academic Performance Using Machine Learning Algorithms. *Behavioral Sciences*, 13(4), 289. <https://doi.org/10.3390/bs13040289>.
- [8] Xue, H., & Niu, Y. (2023). Multi-Output Based Hybrid Integrated Models for Student Performance Prediction. *Applied Sciences*, 13(9), 5384. <https://doi.org/10.3390/app13095384>.
- [9] Smadi, A., Al-Qerem, A., Nabot, A., Jebreen, I., Aldweesh, A., Alauthman, M., Abaker, A. M., Al Zuobi, O. R., & Alzghoul, M. B. (2023). Unlocking the Potential of Competency Exam Data with Machine Learning: Improving Higher Education Evaluation. *Sustainability*, 15(6), 5267. <https://doi.org/10.3390/su15065267>.
- [10] Christou, V., Tsoulos, I. G., Loupas, V., Tzallas, A. T., Gogos, C., Karvelis, P. S., Antoniadis, N., Glavas, E., & Giannakeas, N. (2023). Performance and early drop prediction for higher education students using machine learning. *Expert Systems with Applications*, 225, 120079. <https://doi.org/10.1016/j.eswa.2023.120079>.
- [11] López-García, A., Blasco-Blasco, O., Liern-García, M., & Parada-Rico, S. E. (2023). Early detection of students' failure using Machine Learning techniques. *Operations Research Perspectives*, 11, 100292. <https://doi.org/10.1016/j.orp.2023.100292>.
- [12] Borna, M.-R., Saadat, H., Hojjati, A. T., & Akbari, E. (2024). Analyzing click data with AI: implications for student performance prediction and learning assessment. *Frontiers in Education*, 9, 1421479. <https://doi.org/10.3389/feduc.2024.1421479>.
- [13] Lu, J., Liu, Y., Liu, S., Yan, Z., Zhao, X., Zhang, Y., Yang, C., Zhang, H., Su, W., & Zhao, P. (2024). Machine learning analysis of factors affecting college students' academic performance. *Frontiers in Psychology*, 15, 1447825. <https://doi.org/10.3389/fpsyg.2024.1447825>.
- [14] Turkmenbayev, A., Abdykerimova, E., Nurgozhayev, S., Karabassova, G., & Baigozhanova, D. (2025). The application of machine learning in predicting student performance in university engineering programs: a rapid review. *Frontiers in Education*, 10, 1562586. <https://doi.org/10.3389/feduc.2025.1562586>.
- [15] Zou, W., Zhong, W., Du, J., & Yuan, L. (2025). Prediction of Student Academic Performance Utilizing a Multi-Model Fusion Approach in the Realm of Machine Learning. *Applied Sciences*, 15(7), 3550. <https://doi.org/10.3390/app15073550>.
- [16] Realinho, V., Vieira Martins, M., Machado, J., & Baptista, L. (2021). Predict Students' Dropout and Academic Success [Dataset]. *UCI Machine Learning Repository*. <https://doi.org/10.24432/C5MC89>.