Predicting and Explaining Middle-School Dropout Risk on Imbalanced Data

Mohamed El Jihaoui[1*], *Oum El Kheir* Abra[2], *Khalifa* Mansouri[1]

[1]Modeling and Simulation of Intelligent Industrial Systems Laboratory (M2S2I), ENSET Mohammedia, University Hassan II of Casablanca, Casablanca 20000, Morocco
[2]Regional Center for Education and Training Professions, Rabat 11000, Morocco

Abstract. Early identification of students at risk of dropping out is vital for providing timely support and efficiently allocating educational resources. Using a large, real-world middle-school cohort (N = 810,853; prevalence 4.38%), we develop an explainable TabNet-based model for tabular data. We train with class-weighted loss and early stop on validation PR-AUC to prioritize minority detection, we calibrate probabilities via Platt scaling, and fix a single operating threshold by maximizing Youden's J. The model achieves strong discrimination (ROC-AUC ≈ 0.90; PR-AUC ≈ 0.47 vs. 0.043 baseline) and a recall-centric operating point on test (TPR = 0.793, TNR = 0.833), with balanced metrics confirming robustness (G-Mean ≈ 0.813; MCC ≈ 0.324). Calibration markedly improves probability quality (Brier score 0.1234→0.0317), to better reflect the true likelihood of outcomes. The use of the mask-based TabNet feature provides transparent reasons—where the most recent cumulative grade point average, academic delay, and socioeconomic vulnerability/poverty dominate, supporting targeted and interpretable intervention. The study also reveals that some students are flagged as dropouts despite continuing, necessitating real follow-up costs.

Abbreviation	Full Term
PR-AUC	Precision–Recall Area Under the Curve
ROC-AUC	Receiver Operating Characteristic Area Under the Curve
TPR	True Positive Rate
TNR	True Negative Rate
G-Mean	Geometric Mean
MCC	Matthews Correlation Coefficient
TabNet	Deep Neural Network for Tabular Data

1 Introduction

Education is central to social and economic development, yet student dropout remains a critical challenge in the educational system. In Moroccan middle schools, dropout rates are alarmingly high nationally, with around 10% of students leaving school at this level[1]. Early school leaving undermines efforts to improve literacy and human capital and has far-reaching consequences for individuals and society. Dropping out of school is not only a personal setback but also a societal concern, leading to missed opportunities, reduced lifetime earnings, and broader social costs such as higher poverty and reliance on public assistance [2]. Prior studies have linked school dropout with diminished wages, entry into unskilled labor, and increased risks of criminal behavior and health problems, which in turn burden the

* Corresponding author: mohamed.eljihaoui-etu@etu.univh2c.ma

economy through lost productivity and higher welfare expenditures[2]. These repercussions signal the urgency of addressing dropout to safeguard both individual futures and national development.

Machine learning (ML) offers a promising tool in this context. By leveraging data on students' backgrounds and academic records, ML-based early warning systems can automatically flag students with a high risk of dropping out, allowing educators to target preventive measures[3]. Building such predictive models can help reduce the negative social and economic implications of dropout and guide policymakers in allocating resources for maximum impact. Early warning classification models have been shown to effectively predict at-risk students and thereby support interventions that improve retention[4].

Developing accurate dropout prediction models involves several challenges. Student datasets are typically heterogeneous, combining demographic, socioeconomic, and academic variables in both categorical and numerical formats, which demands models capable of handling diverse tabular data. Another difficulty is the severe class imbalance[5],[6], as dropouts usually represent only a small fraction of the population; this often biases models toward the majority class and reduces their ability to detect at-risk cases. Finally, interpretability[7] is essential for practical adoption—predictions must clearly indicate the underlying risk factors so that educators and policymakers can trust the results and translate them into effective interventions.

In this study, we employ TabNet[8], a deep learning architecture specifically designed for tabular data, to address the challenges of dropout prediction. Unlike traditional multilayer perceptron, TabNet uses a sequential attention mechanism that dynamically selects the most relevant features at each decision step, enabling both high predictive performance and instance-level interpretability. Prior research has shown that TabNet consistently achieves competitive or superior results compared to ensemble tree methods[8], while also offering transparent feature attributions that enhance trust in its predictions[9]. Moreover, it integrates naturally with strategies to handle class imbalance, such as class weighting, and scales effectively to large datasets through end-to-end training.

2 Related Works

A growing body of research has explored machine learning methods to predict student dropouts in secondary education. Traditional statistical models, such as logistic regression, have been widely used as early warning classifiers[10], due to their simplicity and interpretability. In fact, several studies report that logistic regression can serve as a strong baseline for school dropout prediction and sometimes even outperform more complex methods in overall accuracy[3]. Beyond linear models, researchers have applied decision trees, support vector machines, naïve Bayes, and ensemble methods to classify likely dropouts. Decision tree-based classifiers are popular for their interpretability in educational contexts[11], [12]. For instance, Timbal[13]used decision tree models in the Philippines to produce interpretable rule-based patterns—such as links between grades, attendance, and dropout risk—that educators could easily apply in practice. Similarly, Psyridou et al.[2]conducted a study using ensemble algorithms, including Balanced Random Forests and AdaBoost-based models, to anticipate upper-secondary school dropouts from early school data. Their findings demonstrated that ensemble approaches can capture complex patterns of risk and provide valuable insights for early identification.

A pervasive challenge in student dropout prediction is the class imbalance in datasets. Predicting student dropout reliably is consistently challenged by class imbalance, with dropout cases often constituting a minority. Without corrective measures, classifiers tend to

favor the majority class (continuing students), resulting in poor detection of actual at-risk students—false negatives that may hinder prompt and necessary interventions.

In this context, researchers have developed and applied several strategies to counter this imbalance:

- Resampling techniques, such as oversampling the minority class or undersampling the majority, are widely used. Among these, SMOTE (Synthetic Minority Over-sampling Technique) is particularly prominent due to its ability to synthetically generate new minority instances, improving model sensitivity without simple duplication[14-16]. In educational contexts, Mduma et al.[17] demonstrated that hybrid resampling strategies—like SMOTE combined with Edited Nearest Neighbor—enhanced recall of dropout cases across datasets.
- Cost-sensitive learning, in which the misclassification of a dropout incurs a greater penalty than misclassifying a continuing student, enables models to better prioritize the minority class without altering the dataset. These methods require precise calibration to prevent a surge in false positives [18-20].
- Imbalanced-aware ensembles and hybrid methodologies have also been explored to address both resampling and model enhancement [16], [21].

Overall, prior work confirms that dropout prediction is particularly challenging, especially in highly imbalanced[22] educational datasets, where the majority-class patterns easily obscure rare positive cases. Centering on the core objective of reliably identifying at-risk students, we treat handling imbalances as a means rather than an end. Accordingly, we pair TabNet with imbalance-aware training (class weighting), probability calibration via Platt scaling[23] to align predicted probabilities with empirical rates—and post-training threshold optimization. This configuration enhances minority-class recall while maintaining acceptable precision, and calibrated probabilities support credible risk stratification and policy-relevant decisions.

In doing so, we advance dropout prediction from a technical exercise to actionable decision support—improving the detection of true dropouts, curbing unnecessary false alarms, and explaining why each student is flagged.

3 Materials and Methods

We analyze a large, real-world tabular cohort of middle-school students. Table 1 details feature types, cardinalities, and example values for all 810,853 records, with the target labeled as Dropped/Retained.

Table 1. Dataset structure summary.

Feature	Data Type	Unique Values	Missing Values	Sample Values
SchoolType	object	2	0	Public, Private
Gender	object	2	0	Male, Female
Location	object	2	0	Urban, Rural
Grade	object	3	0	9th Grade, 8th Grade, 7th Grade
HasDisability	int64	2	0	Min: 0, Max: 1
HasPedagSupport	int64	2	0	Min: 0, Max: 1
HasSocialSupport	int64	2	0	Min: 0, Max: 1
Repetition	int64	3	0	Min: 0, Max: 2
Age	int64	8	0	Min: 12, Max: 19
Poverty	float64	105	0	Min: 0.0021, Max: 0.1658
Vulnerability	float64	113	0	Min: 0.3676, Max: 36.5715
Delay	int64	8	0	Min: 0, Max: 7
ClassSize	int64	46	0	Min: 5, Max: 50
ClassSize	int64	46	0	Min: 5, Max: 50

CGPA_Y-1	float64	1957	0	Min: 0.0, Max: 19.98
CGPA_Y-2	float64	1826	0	Min: 0.0, Max: 20.0
Status	object	2	0	Retained, Dropped

Figure 1 summarizes the cohort composition. Panel (a) shows a balanced year-wise distribution: 259,085 students in 2022 (31.95%), 269,157 in 2023 (33.19%), and 282,611 in 2024 (34.85%). Panel (b) reveals a pronounced outcome skew: 775,377 students (95.6%) were retained versus 35,476 (4.38%) who dropped out.

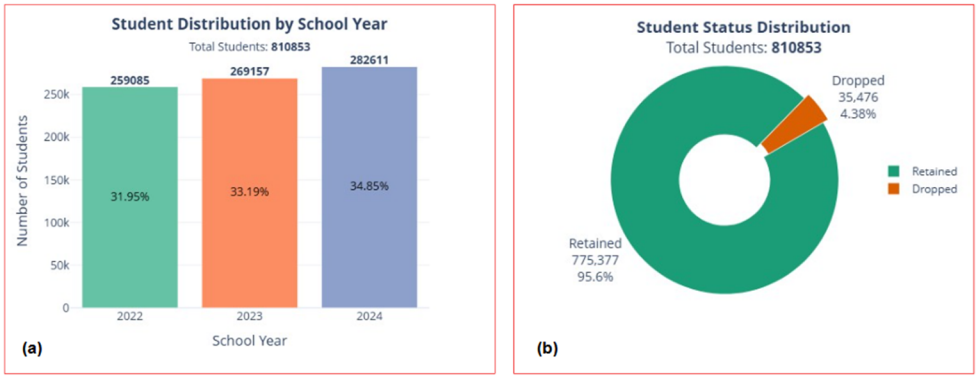


Fig. 1. Student distribution: (a) by school year, (b) by status (Dropped/Retained)

After describing our dataset, our workflow (Figure 2) proceeds in four stages: data ingestion and quality checks, preprocessing of categorical and numerical features, model training, post-hoc calibration and threshold selection, followed by analysis of model performance and explainability.

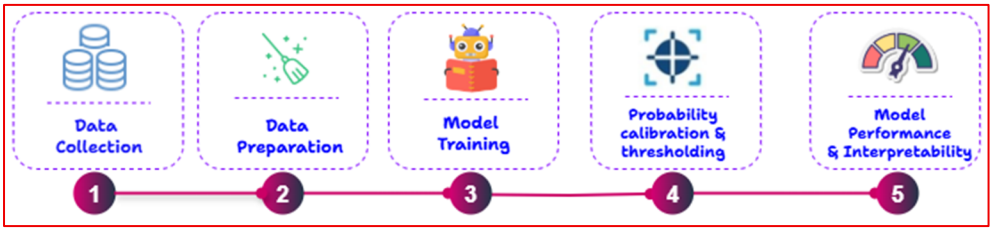


Fig. 2. Proposed Workflow for Student Dropout Prediction.

We train TabNet on a mixed feature space, using its sequential attention to learn sparse, case-specific feature masks that yield interpretable attributions. Because the Dropped class is rare (≈4.38%) and treated as the positive class, we stratify the train/validation/test splits by status (Retained/Dropped) and adopt minority-sensitive evaluation (PR-AUC, class-wise recall/F1). During optimization, we apply cost-sensitive class weights to up-weight minority samples and use PR-AUC as an early stopping technique to favor accurate identification of positive (minority) cases.

After training, we calibrate predicted probabilities on the validation split via Platt scaling to align outputs with real outcomes; using these calibrated scores, we then fix a single operating threshold by minority-aware criteria using Youden's J statistic[24], [25].

4 Results and Discussion

After training and validating our TabNet model, Figure 3 summarizes a well-behaved training process. Although training was scheduled for 250 epochs, we used early stopping on the highest validation PR-AUC, which peaked at epoch 105. Across epochs, log loss for both

training and validation fell sharply and then stabilized with a negligible gap, while PR-AUC rose rapidly and plateaued, with validation closely tracking the training curve. The simultaneous plateau in loss and PR-AUC, together with the absence of train–validation divergence, indicates stable optimization, good probability quality, and strong generalization with no evidence of overfitting.

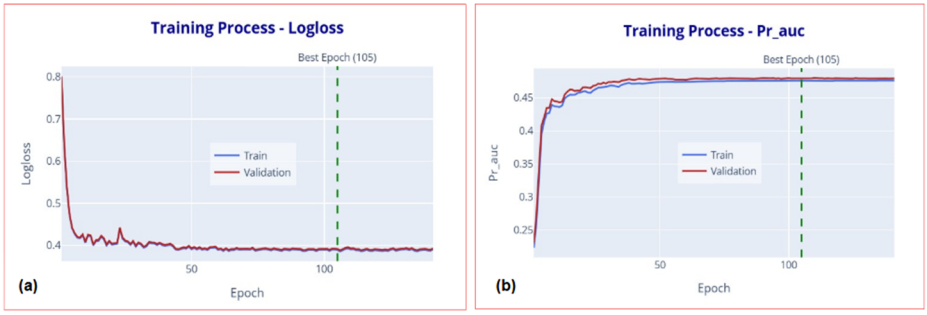


Fig. 3. Training process across epochs: (a) Logloss, (b) PR-AUC

This process is followed by another step: the probability calibration using Platt scaling (Figure 4). Deep models—especially with class weighting and optimization geared toward PR-AUC—often produce well-ranked but miscalibrated probabilities[26], [27]. Model calibration adjusts a classifier's probability outputs so that the predicted probabilities align with the observed frequencies of outcomes, yielding probabilities that reflect true event likelihoods[27].

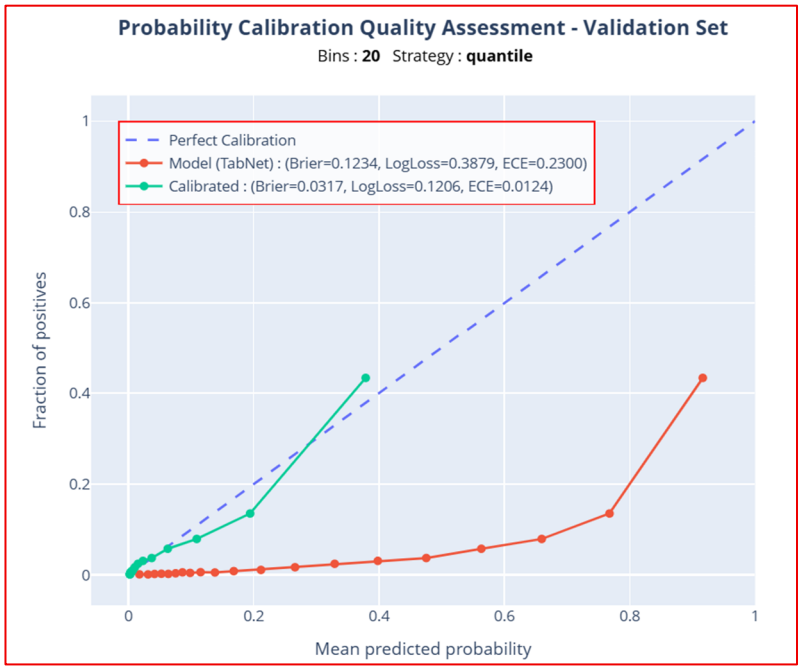


Fig. 4. Probability Calibration Quality Assessment

Calibration markedly improved probability quality: the Brier score dropped from 0.1234 to 0.0317 (~74% reduction), log loss from 0.3879 to 0.1206 (~69% reduction), and ECE from 0.2300 to 0.0124 (~95% reduction). The reliability curve moved toward the diagonal, indicating probabilities that more closely reflect true outcome frequencies. We then fixed the

decision threshold on these calibrated scores. Using the calibrated validation scores, we selected the operating point by maximizing Youden's J statistic (sensitivity + specificity − 1) as in Figure 5. This criterion balances true-positive and true-negative performance in a single number and avoids accuracy-driven bias under class imbalance. The chosen threshold is then fixed for the test dataset for generalization.

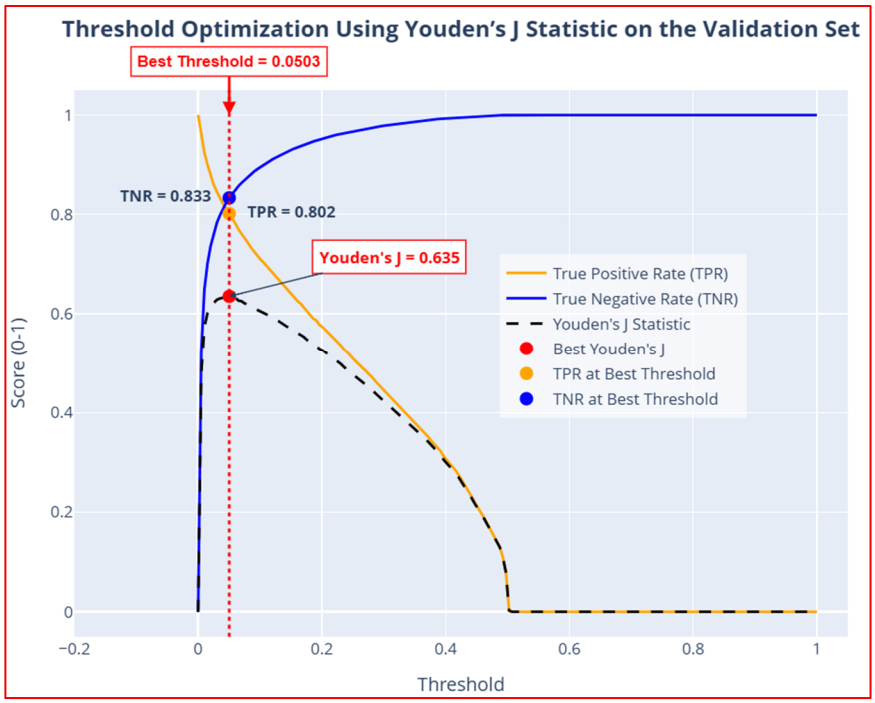


Fig. 5. Threshold optimization using Youden's J statistic

As shown in Fig. 5, the validation-based threshold that maximizes Youden's J is 0.0503, yielding TPR = 0.802, TNR = 0.833, and J = 0.635. This operating point recovers ~80% of true dropouts while maintaining strong specificity—an appropriate trade-off for early-warning systems that prioritize recall without excessive false alarms.

With the validation-selected threshold fixed at thr = 0.0503, Table 2 reports class-wise and aggregate metrics on the held-out test set. The model achieves TPR = 0.79 and TNR = 0.83 (G-Mean ≈ 0.81; Index of BA = 0.66). As expected under low prevalence and a recall-oriented operating point, precision for the Dropped class is modest (0.18; F1 = 0.29), while Retained precision remains high (0.99, F1 = 0.90).

Table 2. Classification Report (Imbalanced) - **Test Set (thr=0.0503)**
Unique labels: **0**: 116307, **1**: 5321} | Prevalence (Dropped=1) =4.375%

Classes	Precision	Recall	Specificity	F1-score	G-Mean	Index of BA*	Support
Dropped	0.18	0.79	0.83	0.29	0.81	0.66	5321
Retained	0.99	0.83	0.79	0.90	0.81	0.66	116307
avg /total	0.95	0.83	0.79	0.88	0.81	0.66	121628

*: Balanced accuracy

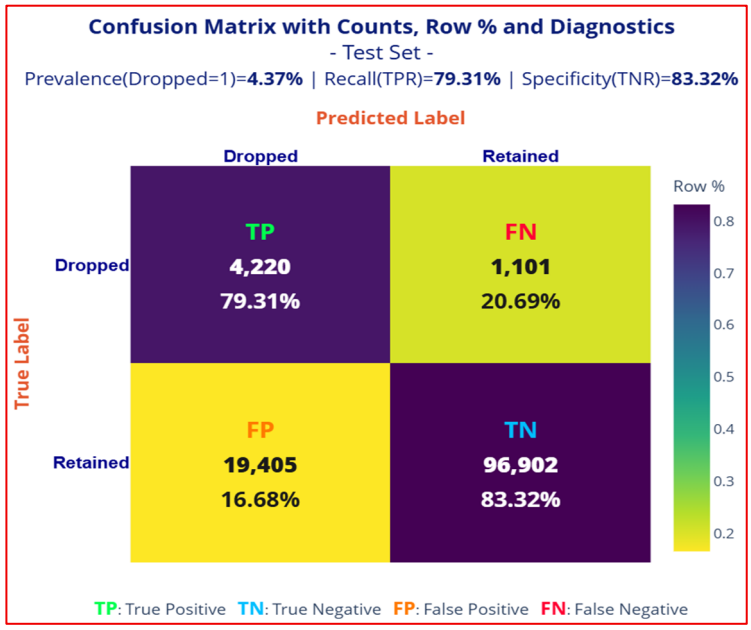


Fig. 6. Confusion matrix on the test set showing counts, row percentages, and diagnostics.

On the held-out test set (prevalence ≈ 4.37%) and at the fixed operating point, the model attains TPR = 0.793 and TNR = 0.833 (G-Mean ≈ 0.81; IBA ≈ 0.66). Figure 6 shows the corresponding counts: 4,220 TP / 1,101 FN among dropouts and 96,902 TN / 19,405 FP among retained students—row rates 79.31/20.69% and 83.32/16.68%, respectively. These results confirm that the system retrieves most true dropouts while keeping false alarms controlled, matching the recall-oriented goal of an early-warning tool.

Figure 7 summarizes TabNet's performance across training, validation, and test at the calibrated, fixed threshold. We report ranking metrics (ROC-AUC, PR-AUC for the minority class) alongside class-balanced measures (Balanced Accuracy, G-Mean, MCC) and class-wise precision/recall/F1 for Dropped=1.

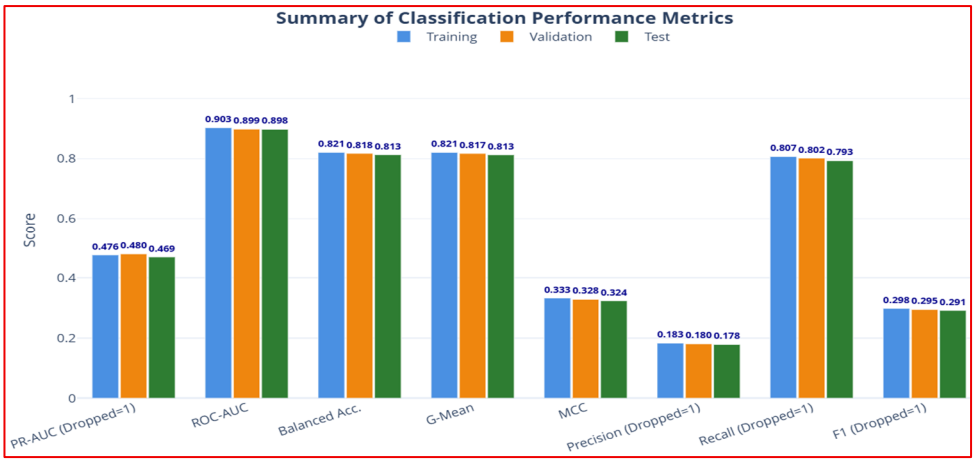


Fig. 7. Model Performance across Training, validation, and test sets**.**

This figure shows stable generalization across splits: ROC-AUC remains ≈0.90, and PR-AUC (~0.47) surpasses the 4.38% baseline, confirming effective minority ranking. Balanced

Accuracy and G-Mean (~0.81–0.82) indicate a solid sensitivity–specificity trade-off, with MCC (~0.33) capturing useful signal under imbalance. At the chosen threshold, dropout recall (~0.80) meets early-warning goals, while precision (~0.18) and F1 (~0.29) reflect the expected false-positive cost. Overall, the model consistently identifies most at-risk students, supporting deployment for early intervention.

To ensure the model's outputs are actionable, we complement discrimination metrics with explanations derived from TabNet. These yield global importances (which features the model relies on most across the cohort) and instance-level rationales[28] (why a given student was flagged). We report the global view below and use instance explanations in practice to align alerts with targeted support. Figure 8 summarizes the importance of features.

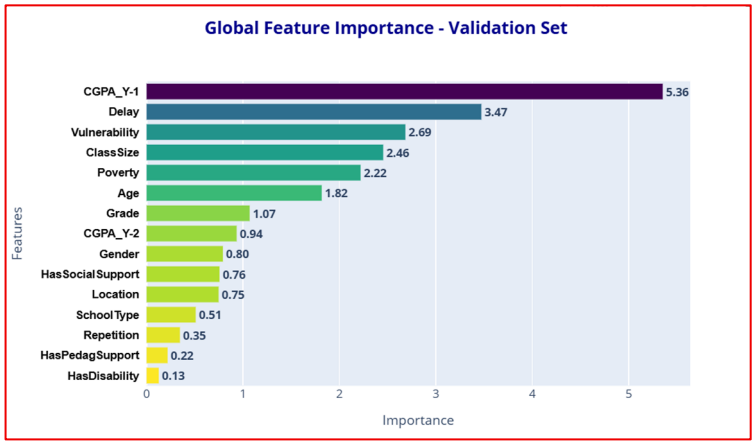
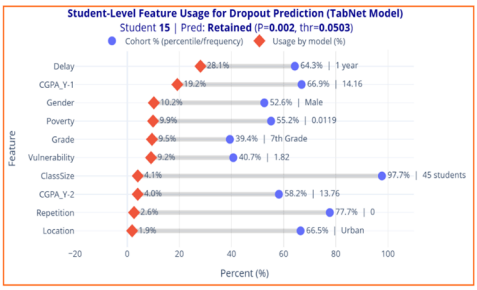


Fig. 8. Global feature importance

In Figure 8, risk estimation is driven chiefly by recent academic performance (CGPA_Y-1), followed by academic delay, socio-economic vulnerability/poverty, and class size. These patterns highlight prior achievement and progression disruptions as the dominant predictors, with socio-economic and classroom factors providing complementary signals. Age, grade, and earlier CGPA (CGPA_Y-2) contribute modestly; gender, location, school type, and support variables have smaller effects, and repetition adds little once delay and GPA are considered—suggesting overlap. While these importances are associational and cohort-specific, they point to actionable levers: academic remediation for low CGPA, catch-up support for delayed students, and targeted socio-economic assistance.

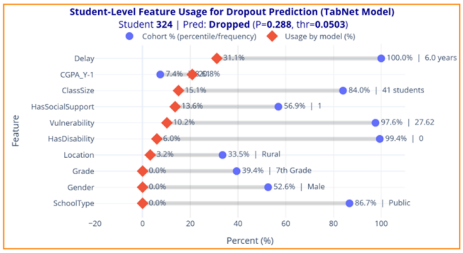
To complement the global analysis, we examine local (per-student) explanations that show how the model arrives at individual predictions. For each student, we plot two quantities per feature: the model's usage and the student's cohort position.



	Feature	Importance	Usage %	Value	Cohort context
0	Delay	5.789	28.11	1 year	64.3th percentile
1	CGPA_Y-1	3.964	19.24	14.16	66.9th percentile
2	Gender	2.110	10.24	Male	52.6% of cohort has 'Male'
3	Poverty	2.045	9.93	0.0119	55.2th percentile
4	Grade	1.962	9.52	7th Grade	39.4% of cohort has '7th Grade'
5	Vulnerability	1.890	9.17	1.82	40.7th percentile
6	ClassSize	0.839	4.07	45 Students	97.7th percentile
7	CGPA_Y-2	0.833	4.04	13.76	58.2th percentile
8	Repetition	0.544	2.64	0	77.7% of cohort has '0'
9	Location	0.392	1.91	Urban	66.5% of cohort has 'Urban'

Fig. 9. Student-Level Feature Usage for Dropout Prediction (Student Id: **15 / Retained**)

For the student in Figure 9, the model assigns a very low calibrated dropout probability (0.002 < 0.0503). Although the student shows a one-year delay (64th percentile), this is outweighed by strong recent achievement (CGPA Y-1: 67th percentile; CGPA Y-2: 58th). Other attributes are near cohort norms, and while class size is large (~98th percentile), it received minimal model usage. On balance, the evidence supports a retained outcome.



	Feature	Importance	Usage %	Value	Cohort context
0	Delay	6.651	31.12	6.0 years	100.0th percentile
1	CGPA_Y-1	4.447	20.81	8.61	7.4th percentile
2	ClassSize	3.217	15.05	41 Students	84.0th percentile
3	HasSocialSupport	2.916	13.64	1	56.9% of cohort has '1'
4	Vulnerability	2.178	10.19	27.62	97.6th percentile
5	HasDisability	1.280	5.99	0	99.4% of cohort has '0'
6	Location	0.681	3.19	Rural	33.5% of cohort has 'Rural'
7	SchoolType	0.000	0.00	Public	86.7% of cohort has 'Public'
8	Gender	0.000	0.00	Male	52.6% of cohort has 'Male'
9	Grade	0.000	0.00	7th Grade	39.4% of cohort has '7th Grade'

Fig. 10. Student-Level Feature Usage for Dropout Prediction (Student Id: **324 / Dropped**)

Figure 10 shows attention concentrated on Delay, CGPA_Y-1, ClassSize, and HasSocialSupport/Vulnerability. The student's profile—extreme delay (100th percentile), very low recent CGPA (7th), high vulnerability (97.6th), and a large class size (84th)—yields a Dropped prediction.

These local attributions align with the global pattern: performance and academic delay are primary drivers, with socio-economic and classroom factors modulating risk. The per-instance view makes decisions auditable and guides intervention (academic remediation, catch-up support, targeted socio-economic assistance).

5 Conclusion

In this work, we developed a calibration-aware TabNet pipeline that turns strong ranking performance into reliable risk scores for early identification of students at risk of dropout. Because the data are highly imbalanced (Dropped ≈ 4.38%), we trained with class weights and used early stopping on PR-AUC to keep focus on the minority class. After training, we applied Platt scaling so that predicted probabilities match observed rates, which is important for creating risk tiers, planning resources, and setting clear decision rules.

We fixed the operating threshold using Youden's J on the validation split (threshold = 0.0503). On the test set, the model achieved high recall for dropouts (TPR = 0.793) together with strong specificity (TNR = 0.833). Class-balanced metrics (G-Mean ≈ 0.813; MCC ≈ 0.324), along with high ROC-AUC (~0.90) and PR-AUC (~0.47, well above the 4.38% baseline), show dependable discrimination in a rare-event setting. In practice, this means the system can identify most at-risk students while keeping false alarms at a manageable level.

The approach is also interpretable. TabNet's sequential attention provides global and student-level explanations that consistently highlight low recent CGPA, academic delay, and socio-economic vulnerability as key drivers. These explanations help educators connect alerts to concrete actions—such as academic remediation, catch-up support, and targeted social assistance. Calibration further improves decision quality (lower Brier score, log loss, and ECE), and a single operating threshold ensures consistent, reproducible triage.

This study focuses on one region and feature set. Future work will test generalization in other contexts, add temporal signals (e.g., attendance), explore resource-aware cost-sensitive training, and conduct subgroup fairness audits. A prospective pilot embedded in school workflows will be important to measure real-world impact on intervention timing and student retention. Overall, our results show that calibrated and interpretable deep tabular models can meaningfully strengthen data-informed efforts to reduce dropouts.

References

[1] A. Ibourk and S. Raoui, 'Territorial obstacles causing early school dropout in Morocco: Multivariate spatial analysis', *Heliyon*, vol. 11, no. 1, Jan. 2025, doi: 10.1016/j.heliyon.2024.e41386.

[2] M. Psyridou, F. Prezja, M. Torppa, M. K. Lerkkanen, A. M. Poikkeus, and K. Vasalampi, 'Machine learning predicts upper secondary education dropout as early as the end of primary school', *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-63629-0.

[3] K. S. Selim and S. S. Rezk, 'On predicting school dropouts in Egypt: A machine learning approach', *Educ Inf Technol (Dordr)*, vol. 28, no. 7, pp. 9235–9266, Jul. 2023, doi: 10.1007/s10639-022-11571-x.

[4] D. Bañeres, M. E. Rodríguez, A. E. Guerrero-Roldán, and A. Karadeniz, 'An early warning system to detect at-risk students in online higher education', *Applied Sciences (Switzerland)*, vol. 10, no. 13, Jul. 2020, doi: 10.3390/app10134427.

[5] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, 'Imbalanced Data Problem in Machine Learning: A Review', *IEEE Access*, vol. 13, pp. 13686–13699, 2025, doi: 10.1109/ACCESS.2025.3531662.

[6] E. Rendón, R. Alejo, C. Castorena, F. J. Isidro-Ortega, and E. E. Granda-Gutiérrez, 'Data sampling methods to dealwith the big data multi-class imbalance problem', *Applied Sciences (Switzerland)*, vol. 10, no. 4, Feb. 2020, doi: 10.3390/app10041276.

[7] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, 'Machine learning interpretability: A survey on methods and metrics', Aug. 01, 2019, *MDPI AG*. doi: 10.3390/electronics8080832.

[8] S. Arık and T. Pfister, 'TabNet: Attentive Interpretable Tabular Learning', *35th AAAI Conference on Artificial Intelligence, AAAI 2021*, vol. 8A, pp. 6679–6687, Aug. 2019, doi: 10.1609/aaai.v35i8.16826.

[9] K. McDonnell, F. Murphy, B. Sheehan, L. Masello, and G. Castignani, 'Deep learning in insurance: Accuracy and model interpretability using TabNet', *Expert Syst Appl*, vol. 217, p. 119543, May 2023, doi: 10.1016/J.ESWA.2023.119543.

[10] S. Koon and Y. Petscher, 'Comparing Methodologies for Developing an Early Warning System: Classification and Regression Tree Model versus Logistic Regression. REL 2015-077.', *Regional Educational Laboratory Southeast*, Feb. 2015, Accessed: Sep. 08, 2025. [Online]. Available: http://ies.ed.gov/

[11] E. M. Queiroga *et al.*, 'A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course', *Applied Sciences 2020, Vol. 10, Page 3998*, vol. 10, no. 11, p. 3998, Jun. 2020, doi: 10.3390/APP10113998.

[12] A. Villar and C. R. V. de Andrade, 'Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study', *Discover Artificial Intelligence*, vol. 4, no. 1, pp. 1–24, Dec. 2024, doi: 10.1007/S44163-023-00079-Z/FIGURES/9.

[13] M. A. Timbal, 'Analysis of Student-at-Risk of Dropping out (SARDO) Using decision tree: An Intelligent predictive model for reduction', *Int J Mach Learn Comput*, vol. 9, no. 3, pp. 273–278, Jun. 2019, doi: 10.18178/IJMLC.2019.9.3.798.

[14] I. K. R. Arthana, I. M. D. Maysanjaya, G. A. Pradnyana, and G. R. Dantes, 'Optimizing Dropout Prediction in University Using Oversampling Techniques for Imbalanced Datasets', *International Journal of Information and Education Technology*, vol. 14, no. 8, pp. 1052–1060, 2024, doi: 10.18178/ijiet.2024.14.8.2133.

[15] R. Suguna, J. Suriya Prakash, H. Aditya Pai, T. R. Mahesh, V. Vinoth Kumar, and T. E. Yimer, 'Mitigating class imbalance in churn prediction with ensemble methods and SMOTE', *Sci Rep*, vol. 15, no. 1, pp. 1–20, Dec. 2025, doi: 10.1038/S41598-025-01031-0;SUBJMETA=1041,1042,639,705;KWRD=APPLIED+MATHEMATICS,COMPUTATIONAL+SCIENCE.

[16] A. Jain, A. K. Dubey, S. Khan, A. Panwar, M. Alkhatib, and A. M. Alshahrani, 'A PSO weighted ensemble framework with SMOTE balancing for student dropout prediction in smart education systems', *Sci Rep*, vol. 15, no. 1, pp. 1–28, Dec. 2025, doi: 10.1038/S41598-025-97506-1;SUBJMETA=1042,117,639,705;KWRD=COMPUTATIONAL+SCIENCE,COMPUTER+SCIENCE.

[17] S. Cabezuelo, R. González, D. Campo, R. Barbero, and N. Mduma, 'Data Balancing Techniques for Predicting Student Dropout Using Machine Learning', *Data 2023, Vol. 8, Page 49*, vol. 8, no. 3, p. 49, Feb. 2023, doi: 10.3390/DATA8030049.

[18] A. Khan, F. Khan, S. Khan, I. A. Khan, and M. Saeed, 'Cost Sensitive Learning and SMOTE Methods for Imbalanced Data', *Journal of Applied and Emerging Sciences*, vol. 8, no. 1, p. 32, Jan. 2018, doi: 10.36785/JAES.81240.

[19] E. Kalita *et al.*, 'Educational data mining: a 10-year review', *Discover Computing*, vol. 28, no. 1, pp. 1–25, Dec. 2025, doi: 10.1007/S10791-025-09589-Z/TABLES/4.

[20] S. Lee and J. Y. Chung, 'The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction', *Applied Sciences 2019, Vol. 9, Page 3093*, vol. 9, no. 15, p. 3093, Jul. 2019, doi: 10.3390/APP9153093.

[21] A. A. Khan, O. Chaudhari, and R. Chandra, 'A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation', *Expert Syst Appl*, vol. 244, p. 122778, Jun. 2024, doi: 10.1016/J.ESWA.2023.122778.

[22] V. Kumar *et al.*, 'Addressing Binary Classification over Class Imbalanced Clinical Datasets Using Computationally Intelligent Techniques', *Healthcare 2022, Vol. 10, Page 1293*, vol. 10, no. 7, p. 1293, Jul. 2022, doi: 10.3390/HEALTHCARE10071293.

[23] A. Niculescu-Mizil and R. Caruana, 'Predicting good probabilities with supervised learning', *ICML 2005 - Proceedings of the 22nd International Conference on Machine Learning*, pp. 625–632, 2005, doi: 10.1145/1102351.1102430.

[24] M. Hassanzad and K. Hajian-Tilaki, 'Methods of determining optimal cut-point of diagnostic biomarkers with application of clinical data in ROC analysis: an update review', *BMC Med Res Methodol*, vol. 24, no. 1, pp. 1–13, Dec. 2024, doi: 10.1186/S12874-024-02198-2/TABLES/5.

[25] R. Fluss, D. Faraggi, and B. Reiser, 'Estimation of the Youden Index and its associated cutoff point', *Biometrical Journal*, vol. 47, no. 4, pp. 458–472, Aug. 2005, doi: 10.1002/BIMJ.200410135;WGROUP:STRING:PUBLICATION.

[26] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, 'On Calibration of Modern Neural Networks', *34th International Conference on Machine Learning, ICML 2017*, vol. 3, pp. 2130–2143, Jun. 2017, Accessed: Sep. 05, 2025. [Online]. Available: https://arxiv.org/pdf/1706.04599

[27] S. Rajaraman, P. Ganesan, and S. Antani, 'Deep learning model calibration for improving performance in class-imbalanced medical image classification tasks', *PLoS One*, vol. 17, no. 1, p. e0262838, Jan. 2022, doi: 10.1371/JOURNAL.PONE.0262838.

[28] B. Zhang *et al.*, 'TabNet: Locally Interpretable Estimation and Prediction for Advanced Proton Exchange Membrane Fuel Cell Health Management', *Electronics 2024, Vol. 13, Page 1358*, vol. 13, no. 7, p. 1358, Apr. 2024, doi: 10.3390/ELECTRONICS13071358.