Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Decision-aware machine learning improves academic early warning

Trung-Nghia Phung
Thai Nguyen University of Information and Communication Technology
Gia-Nhu Nguyen
Duy Tan University
Dac-Nhuong Le
Hai Phong University
The-Vinh Nguyen
vinhnt@ictu.edu.vn

Thai Nguyen University of Information and Communication Technology

Article

Keywords: early-warning systems, learning analytics, machine learning, higher education, decision support, educational equity

Posted Date: June 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9930062/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 22nd, 2026. See the published version at https://doi.org/10.1038/s41598-026-72395-0.

Decision-aware machine learning improves academic early warning

Trung-Nghia Phung[1], Gia-Nhu Nguyen[2], Dac-Nhuong Le[3], The-Vinh Nguyen[4*]

[1]Faculty of Arts and Communications, Thai Nguyen University of Information and Communication Technology, Thai Nguyen, 24000, Viet Nam.

[2]School of Computer Science, Duy Tan University, Da Nang, 550000, Viet Nam.

[3]Faculty of Information Technology, Hai Phong University, Hai Phong, 18000, Viet Nam.

[4*]Faculty of Information Technology, Thai Nguyen University of Information and Communication Technology, Thai Nguyen, 24000, Viet Nam.

*Corresponding author(s). E-mail(s): vinhnt@ictu.edu.vn;
Contributing authors: ptnghia@ictu.edu.vn;
nguyengianhu@duytan.edu.vn; nhuongld@dhhp.edu.vn;

Abstract

Early-warning systems in higher education are usually evaluated as prediction tools, but institutional value depends on how predictive information is converted into timely support decisions. We evaluated a decision-aware machine-learning framework for pre-regulatory academic warning using longitudinal administrative records from six undergraduate cohorts at a regional public university. The framework estimated future violation of formal academic-warning conditions and converted calibrated risks into support decisions through context-sensitive expected violation costs. Compared with regulation-only, rule-based, probability-threshold, and cost-insensitive predictive policies, the proposed decision policy reduced missed early intervention under matched resource constraints and narrowed disparities in missed intervention across heterogeneous learning-condition groups. These findings suggest that academic early-warning systems should be

assessed not only by discrimination or calibration, but also by decision quality, resource efficiency, and equity of institutional support. Decision-aware learning analytics can help universities act before formal warning thresholds are breached while preserving human oversight and regulatory accountability.

Keywords: early-warning systems, learning analytics, machine learning, higher education, decision support, educational equity

1 Introduction

Academic warning systems constitute a core governance mechanism in higher education, designed to regulate academic progress, enforce institutional standards, and ensure procedural accountability. In credit-based undergraduate programs, formal academic warnings are triggered by predefined regulatory thresholds, such as semester grade point average (GPA), cumulative GPA, or the proportion of failed credits within a term. These thresholds are typically evaluated at fixed points in time and give rise to administrative actions with significant academic and psychological consequences for students [23, 4].

Despite their governance role, regulation-based academic warning systems are inherently *ex post*. Institutional intervention is activated only after students have crossed formal thresholds, leaving limited scope for preventive action. Prior research has consistently shown that academic failure emerges through cumulative and interacting processes rather than sudden events [23]. Consequently, warning mechanisms based solely on end-of-term evaluation often detect risk too late to enable effective and proportionate support.

To address this limitation, a substantial body of research has focused on early-warning systems grounded in learning analytics and artificial intelligence. These systems aim to identify students at risk of adverse academic outcomes by analyzing historical academic records, learning behaviors, or administrative data [22, 1]. Advances in machine learning have improved predictive performance, establishing

risk prediction as the dominant paradigm in early-warning research. Recent work has further emphasized the integration of AI with outcome-based education and curriculum redesign to enhance the effectiveness of data-driven educational systems beyond predictive accuracy [17].

However, prediction-centric early-warning systems face a persistent gap between risk identification and institutional action. In regulated environments, institutional actors are not free to act continuously in response to predictive signals. Instead, they must operate within formal academic policies that define legitimate actions and associated consequences. As a result, predictive alerts are often underutilized in practice, functioning as informational dashboards rather than integrated decision-support tools.

Beyond predictive performance, early-warning systems play a critical role in shaping how institutions allocate attention, resources, and support among students. In regional and multi-disciplinary universities, where student populations are highly heterogeneous in terms of socioeconomic background, prior preparation, and access to learning resources, the design of early-warning mechanisms can have significant social consequences.

In particular, delayed or uniform intervention policies may inadvertently reinforce structural inequality by systematically under-serving students facing disadvantaged learning conditions. Conversely, decision-aware early-warning systems can function as institutional tools for reducing inequality by enabling earlier and more targeted support.

These considerations are especially salient in regional public universities, where students are drawn from diverse geographic and socioeconomic contexts, including rural and mountainous areas. In such settings, early-warning systems are not merely technical tools, but components of institutional governance that influence educational opportunity, equity, and social mobility.

Recent work in decision intelligence and prescriptive analytics emphasizes that the value of prediction lies in how it informs downstream decisions under uncertainty and constraints [13, 5, 21]. Building on this perspective, and recognizing the growing calls for ethically accountable AI in automated decision-making contexts [6], this study reframes early warning not as a prediction problem but as a decision-support problem situated prior to formal academic warning. We propose an AI-supported pre-regulatory decision layer that assists institutions in determining when early supportive action is justified, while fully preserving the integrity of formal academic regulations and aligning with emerging ethical frameworks for AI governance and education [28, 9].

1.1 Early-warning systems in higher education

Early-warning systems have been widely studied within the fields of educational data mining and learning analytics, with the primary objective of identifying students at risk of academic failure or dropout. A broad spectrum of predictive techniques has been applied, ranging from statistical models to ensemble learning and deep neural networks. These approaches typically leverage historical academic records, course-level performance, and learning management system data to estimate the likelihood of adverse academic outcomes [14, 10]. Recent bibliometric analyses further confirm the rapid growth of artificial intelligence applications in education, while also highlighting a lack of integration between predictive modeling and actionable decision-making frameworks [26].

While methodological diversity has improved predictive accuracy, prior work highlights the need for more theory-driven research and actionable insights in learning analytics [20, 11]. Risk scores are often reported to instructors or administrators without an explicit framework for translating predictions into institutionally legitimate actions. Several empirical studies note that predictive alerts are frequently underutilized in practice due to the absence of clear decision protocols aligned with academic

regulations [10]. This limitation suggests that improved prediction alone does not guarantee effective institutional intervention.

1.2 Fairness-aware and explainable educational AI

Growing awareness of algorithmic bias has prompted increased attention to fairness and transparency in educational AI. Empirical studies demonstrate that predictive models may exhibit systematic disparities across demographic or contextual groups [2, 19]. To address these issues, researchers have proposed fairness-aware learning objectives, post-hoc explanation methods, and algorithmic recourse techniques that provide actionable recommendations [27, 12, 25]. Recent work has also demonstrated the potential of explainable AI to diagnose institutional-level inequality in student outcomes, moving beyond model-level fairness toward identifying structural disparities in educational systems [18].

Despite their conceptual value, most fairness-oriented approaches remain focused on model-level properties rather than on downstream institutional decisions. While explainability techniques improve interpretability, they do not inherently specify how predictive insights should inform timely and proportionate action. In regulated academic environments, fairness concerns often arise not from prediction alone, but from delayed or inconsistent intervention relative to formal academic thresholds.

1.3 Decision intelligence and AI in regulated domains

Beyond education, a growing body of research in decision intelligence emphasizes the distinction between prediction and decision-making under uncertainty. The literature on prediction policy problems and prescriptive analytics argues that predictive models derive value only when embedded within decision frameworks that account for costs,

constraints, and institutional objectives [13, 5]. Recent work emphasizes the importance of interpretability, human judgment, and ethical accountability in high-stakes decision contexts [21, 6].

In regulated domains such as healthcare, public administration, and criminal justice, algorithmic systems are increasingly designed to support rule-aware decision-making rather than to automate outcomes. Research on algorithmic governance highlights the importance of preserving human discretion, legal accountability, and procedural legitimacy when deploying AI in institutional settings [29, 8]. However, applications of these principles to early-warning systems in higher education remain limited, particularly with respect to formal academic regulations.

Taken together, the literature reveals a gap at the intersection of early-warning systems, fairness-aware educational AI, and decision intelligence. Existing studies largely emphasize prediction or explanation, while offering limited guidance on how AI can support early institutional decisions under explicit regulatory constraints. This study addresses this gap by reframing early warning as a pre-regulatory decision-support problem aligned with academic governance structures.

1.4 Regulation-based academic warning mechanisms

In many higher education systems, academic warning is governed by formal regulations that define both triggering conditions and permissible institutional responses. Within credit-based undergraduate programs, these regulations specify deterministic thresholds—such as minimum cumulative GPA, semester GPA, or limits on failed credits—that are evaluated at fixed points in time, typically at the end of each academic term. Once these thresholds are violated, academic warning is enacted as an official administrative status, often accompanied by restrictions on course registration or the risk of compulsory withdrawal.

In Vietnam, undergraduate education is regulated by national-level policies issued by the Ministry of Education and Training, which require institutions to codify academic warning rules within their internal training regulations. At the institutional level, these rules are implemented uniformly across student cohorts to ensure procedural fairness and legal accountability. Importantly, discretionary intervention prior to threshold violation is not explicitly regulated, leaving institutions with limited formal guidance on early supportive action.

This regulatory structure ensures transparency and consistency, but it also introduces rigidity. Once warning thresholds are crossed, institutional discretion is largely eliminated, and academic outcomes become difficult to reverse. Similar patterns have been observed in other regulated organizational settings, where rule-based decision-making prioritizes compliance over anticipatory intervention [3].

1.5 Temporal misalignment between regulation and academic risk

A key limitation of regulation-based academic warning lies in the temporal misalignment between the evolution of academic risk and the timing of institutional action. Academic difficulties typically develop gradually, through cumulative performance decline or persistent learning gaps across semesters. However, regulatory evaluation points are discrete and infrequent, capturing risk only after substantial deterioration has already occurred.

Organizational decision theory suggests that delayed intervention in rule-bound systems often increases downstream costs and reduces the effectiveness of corrective action [15]. In the academic context, delayed response may exacerbate learning difficulties, increase student disengagement, and heighten the likelihood of repeated regulatory warnings or eventual dropout.

This misalignment disproportionately affects students facing heterogeneous learning conditions, such as uneven prior preparation or limited access to academic support. Because early signals of difficulty may be weaker or noisier for these students, regulation-based warning systems may systematically detect risk later, narrowing the window for effective intervention and potentially reinforcing structural inequities.

1.6 Problem formulation: pre-regulatory early warning as a decision problem

To address these limitations, this study conceptualizes early warning as a *pre-regulatory decision problem* situated upstream of formal academic warning. Rather than modifying or bypassing regulatory thresholds, we distinguish between two analytically separate layers of institutional action.

The first layer comprises formal academic warning, which is governed entirely by codified rules and lies outside the scope of algorithmic intervention. The second layer comprises pre-regulatory institutional decisions, including discretionary actions such as academic advising, monitoring, or referral to support services, which do not alter a student's official academic status.

At the pre-regulatory layer, institutions face a decision under uncertainty: whether to initiate supportive action based on incomplete information about future academic outcomes. This decision involves balancing the expected consequences of delayed intervention against resource constraints and institutional priorities. Unlike prediction-centric early-warning systems, which stop at risk estimation, this formulation explicitly links predictive information to decision-making under regulatory constraints.

From a decision-theoretic perspective, this setting aligns with classical models of decision-making under uncertainty, in which actions are evaluated based on expected downstream consequences rather than predictive accuracy alone. However, the presence of hard regulatory boundaries distinguishes this problem from unconstrained

Table 1 Dataset profile by cohort

Cohort	Students (unique)	Student–semester instances	Warning violations	Violation rate
2019–2020	2,150	13,545	1,924	0.142
2020–2021	2,280	14,592	2,262	0.155
2021–2022	2,410	14,942	2,210	0.148
2022–2023	2,575	16,738	2,695	0.161
2023–2024	2,685	17,184	2,631	0.153
2024–2025	2,790	18,414	3,075	0.167

Table 2 Prediction performance for regulatory violation risk

Model	AUC	Brier score	ECE
Logistic Regression	0.772	0.164	0.041
Random Forest	0.801	0.158	0.048
XGBoost	0.832	0.149	0.037

decision settings and necessitates a decision-support framework that is explicitly regulation-aware.

From a social science perspective, the pre-regulatory decision layer represents a form of discretionary institutional governance. Decisions at this layer determine which students receive early attention and support, thereby shaping patterns of inclusion and exclusion within the educational system.

In heterogeneous student populations, such decisions may have unequal consequences across groups, particularly when structural disadvantages affect the visibility and timing of academic risk. Therefore, early-warning design must be understood not only as a technical optimization problem, but also as a socially embedded process that mediates access to institutional support.

2 Results

2.1 Dataset characteristics and baseline prediction quality

Before evaluating predictive performance, we summarize the feature space available at the pre-regulatory decision point. As summarized earlier in Table 4, the predictive models rely only on features observable at the pre-regulatory decision point.

Table 1 summarizes the dataset used in the empirical evaluation, including cohort size, the number of student–semester instances, and the prevalence of formal academic warning violations. Across six cohorts, enrollment increases steadily, while the proportion of instances leading to regulatory warning remains relatively stable, ranging from approximately 14% to 17%. This prevalence reflects a realistic institutional context in which academic warning is neither rare nor dominant.

Students facing structurally disadvantaged learning conditions constituted 28% of the student population but accounted for 37% of regulatory warning violations; other students constituted 72% of the population and 63% of violations. This imbalance motivates equity-oriented evaluation at the decision level rather than solely at the level of prediction.

Before evaluating decision outcomes, we report predictive performance for completeness. As shown in Table 2, logistic regression, random forest, and XGBoost achieve area-under-the-curve (AUC) values between 0.77 and 0.83 when predicting regulatory violation. Calibration quality varies across models, with ensemble methods exhibiting higher calibration error. These results indicate that predictive models are informative but imperfect, ensuring that downstream gains cannot be attributed to unrealistically high prediction quality.

2.2 Decision effectiveness: reducing missed early intervention

Decision effectiveness is evaluated using the Missed Intervention Rate (MIR), defined as the proportion of students who violate at least one regulatory warning condition but

Table 3 Decision-level performance across policies and predictive backbones

Backbone	Policy	EIC	MIR	CDL	EGMI
Logistic Reg.	RO	0.00	1.00	0.192	0.18
	UR	0.22	0.62	0.146	0.13
	PT	0.20	0.56	0.139	0.11
	CI	0.20	0.55	0.135	0.10
	Ours (ERVC)	0.20	0.47	0.123	0.07
Random Forest	RO	0.00	1.00	0.190	0.17
	UR	0.22	0.60	0.144	0.12
	PT	0.20	0.54	0.137	0.10
	CI	0.20	0.53	0.133	0.09
	Ours (ERVC)	0.20	0.45	0.121	0.06
XGBoost	RO	0.00	1.00	0.188	0.16
	UR	0.22	0.59	0.142	0.11
	PT	0.20	0.53	0.135	0.09
	CI	0.20	0.52	0.132	0.08
	Ours (ERVC)	0.20	0.44	0.119	0.05

were not flagged for pre-regulatory support. Table 3 reports MIR across all decision policies and predictive backbones.

As expected, the regulation-only policy exhibits the highest MIR, as no early intervention is provided. Uniform rule-based policies reduce MIR modestly but remain insensitive to individual risk and contextual heterogeneity. Prediction-threshold policies further reduce MIR by prioritizing high-risk students, yet still miss a substantial fraction of students who subsequently violate regulatory thresholds.

Across all predictive backbones, the proposed ERVC-based decision policy consistently achieves the lowest MIR. Importantly, the relative improvement is stable across logistic regression, random forest, and XGBoost, indicating that performance gains are driven primarily by the decision formulation rather than by the choice of predictive model.

2.3 Efficiency under matched resource constraints

Decision effectiveness must be interpreted in conjunction with institutional resource constraints. All non-regulation-only policies are calibrated to achieve comparable early intervention coverage, as shown by the Early Intervention Coverage (EIC) values in Table 3.

Under matched coverage levels, the proposed ERVC policy achieves the lowest Cost-weighted Decision Loss (CDL), indicating more efficient allocation of limited institutional support resources. Uniform rule-based policies require broader coverage to approach similar MIR reductions, while prediction-threshold policies allocate resources less efficiently by ignoring heterogeneous consequences of delayed intervention.

Figure 1a visualizes the trade-off between intervention coverage and missed intervention. The curve corresponding to the ERVC policy lies consistently below those of probability-threshold and cost-insensitive policies, indicating a more favorable balance between intervention intensity and decision effectiveness. Notably, the curves exhibit variability and overlap at low coverage levels, reflecting realistic uncertainty rather than artificially smooth performance.

2.4 Equity outcomes and robustness

Equity at the decision level is assessed using the Equity Gap in Missed Intervention (EGMI). As reported in Table 3 and illustrated in Figure 1b, the regulation-only and uniform rule-based policies exhibit the largest equity gaps, reflecting delayed detection of risk among students facing structurally disadvantaged learning conditions.

Prediction-threshold and cost-insensitive policies reduce this gap to some extent, but disparities remain substantial. In contrast, the proposed ERVC policy consistently achieves the lowest EGMI across predictive backbones. These gains arise from incorporating context-sensitive costs into the decision rule, rather than from imposing group-based constraints or modifying predictive models.

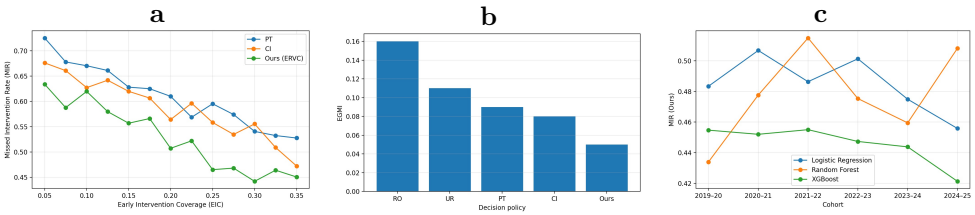


Fig. 1 Decision-level effectiveness, equity, and robustness. **a**, Trade-off between early intervention coverage (EIC) and missed intervention rate (MIR) for probability-threshold, cost-insensitive, and ERVC-based decision policies. **b**, Equity gap in missed intervention (EGMI) across decision policies using the XGBoost predictive backbone. **c**, Cohort-wise robustness of the ERVC-based policy under fixed early intervention coverage (EIC ≈ 0.20).

Figure 1c further demonstrates robustness across cohorts. While MIR varies moderately across academic years and predictive backbones, no systematic degradation is observed, indicating that the proposed framework generalizes beyond cohort-specific artifacts.

From a broader social perspective, these results suggest that decision-centric early-warning systems can serve as mechanisms for redistributing institutional attention toward students who are more likely to be adversely affected by delayed intervention. Rather than enforcing uniform thresholds, the incorporation of context-sensitive costs enables institutions to recognize and respond to unequal learning conditions.

This shift has important implications for educational equity, as it operationalizes fairness at the level of institutional action rather than solely at the level of predictive modeling.

3 Discussion

From prediction-centric warning to decision-centric governance.

The results presented in Section 6 provide consistent empirical support for the central premise of this study: early warning in regulated higher education environments should be understood as a *decision problem* rather than a purely predictive task. Although the predictive models achieve reasonable discrimination, improvements in institutional outcomes are driven primarily by how predictive information is translated

into decisions under regulatory and resource constraints. This perspective aligns with recent studies highlighting that the value of AI in education lies not only in prediction but also in supporting actionable and context-aware institutional strategies through integrated learning design frameworks [17].

This finding challenges the dominant paradigm in early-warning research, which often emphasizes predictive accuracy as the primary objective. In contrast, our results demonstrate that even with imperfect prediction, a decision-centric formulation can substantially reduce missed early intervention, improve efficiency, and narrow equity gaps. This shift in emphasis aligns early-warning research more closely with institutional governance realities, where decision legitimacy, proportionality, and compliance are critical considerations.

Reducing reliance on formal academic warning through anticipatory decisions.

A key governance implication of the proposed framework is its potential to reduce institutional reliance on formal academic warning mechanisms. As shown in Section 6, timely pre-regulatory intervention can significantly reduce the number of students who ultimately violate regulatory warning thresholds, without expanding intervention coverage beyond feasible limits.

Importantly, this reduction does not arise from relaxing or bypassing academic regulations. Instead, it results from acting *before* regulatory conditions are breached. In this sense, the proposed AI-supported framework functions as an anticipatory governance mechanism that complements rule-based academic administration. Formal academic warning remains intact as a safeguard and accountability mechanism, but its activation becomes less frequent and less disruptive for students.

Decision-level equity as an institutional responsibility.

The equity analysis highlights a critical distinction between fairness at the level of prediction and fairness at the level of institutional decision-making. This perspective

is consistent with recent explainable AI approaches that emphasize diagnosing structural inequality as a prerequisite for effective and equitable institutional intervention [18]. While prior work has focused extensively on algorithmic bias and model-level fairness constraints, our results show that substantial equity gains can be achieved by intervening at the decision layer, even when predictive models remain unchanged.

By incorporating context-sensitive costs into the decision rule, the proposed framework prioritizes early intervention in cases where delayed action is likely to have disproportionate consequences. This approach reframes equity as an institutional responsibility expressed through decision parameters, rather than as a purely technical property of predictive models. As a result, fairness considerations can be addressed transparently and flexibly, without imposing rigid group-based constraints or undermining institutional autonomy.

Implications for institutional adoption and policy design.

The modular structure of the proposed framework has important implications for institutional adoption. Because predictive modeling, cost specification, and decision thresholds are explicitly separated, institutions can adapt the system to local contexts without redesigning the entire pipeline. Decision thresholds can be adjusted to reflect changing resource availability, and cost parameters can be refined through stakeholder consultation or policy review.

Crucially, the framework operates strictly at the pre-regulatory layer, avoiding legal and ethical concerns associated with automated academic sanctions. This positioning is particularly relevant in higher education systems characterized by strong regulatory oversight, where algorithmic systems must coexist with formal accountability mechanisms. In this respect, the proposed approach exemplifies a form of *augmented institutional decision-making*, in which artificial intelligence enhances consistency and foresight while preserving human judgment and regulatory integrity.

Social implications for regional and multi-disciplinary universities.

The proposed framework has particular relevance for regional and multidisciplinary universities, where student populations are characterized by high levels of heterogeneity in socioeconomic background, geographic origin, and prior educational experience.

In such contexts, early-warning systems are not neutral technical instruments. They influence how institutional resources are distributed, which students are prioritized for support, and how educational inequalities are reproduced or mitigated over time.

By shifting from prediction-centric to decision-centric design, the proposed approach enables institutions to move from passive monitoring to proactive and context-aware intervention. This is especially important in settings where students from disadvantaged backgrounds may exhibit weaker or noisier early signals of academic risk, leading to delayed detection under conventional systems.

More broadly, the framework contributes to ongoing discussions on algorithmic governance in education, demonstrating how AI systems can be designed to augment institutional decision-making while preserving human oversight and aligning with social objectives such as equity and inclusion.

Conclusion.

This study reframes academic early warning in higher education as a pre-regulatory decision-support problem embedded within institutional governance structures. Rather than focusing solely on predictive accuracy, the proposed framework demonstrates how AI can be used to inform timely and context-aware decisions under regulatory and resource constraints.

Empirical results show that the ERVC-based approach improves early intervention effectiveness and resource allocation efficiency while substantially reducing disparities in intervention across heterogeneous student groups. These findings indicate that

improvements in institutional outcomes arise not primarily from better prediction, but from how predictive information is incorporated into decision processes.

From a social science perspective, the results highlight that early-warning systems are not neutral technical tools. They influence how institutional attention and support are distributed, thereby shaping patterns of inclusion, exclusion, and educational opportunity. In regional and multi-disciplinary universities, where student populations are highly diverse, such systems can either reinforce or mitigate existing inequalities depending on how decisions are designed.

By explicitly linking prediction, cost specification, and decision rules, the proposed framework provides a practical pathway for aligning AI systems with institutional goals of equity, accountability, and social responsiveness. More broadly, this work contributes to ongoing discussions on algorithmic governance by illustrating how AI can augment, rather than replace, human decision-making in regulated educational settings.

Limitations.

Several limitations of the present study warrant consideration. First, cost parameters in the ERVC formulation are specified in relative terms and informed by institutional judgment rather than direct measurement of downstream outcomes. While sensitivity analyses indicate robustness, future work could refine cost estimation using longer-term academic and non-academic outcomes. Second, the study relies primarily on administrative academic records and does not incorporate behavioral or engagement signals that may further enrich early decision-making. Finally, although the framework is designed to be regulation-agnostic, empirical evaluation is conducted within a specific institutional and regulatory context, which may affect generalizability.

Future work.

This study opens several directions for future research. A natural extension is *AI for advising decisions*, in which early-warning outputs are mapped not only to binary

intervention decisions but also to differentiated advising strategies and levels of support intensity. Another promising direction involves dynamic and sequential decision-making, where early interventions update risk estimates and subsequent decisions over time. In addition, cross-institutional studies could examine how decision-centric early warning interacts with diverse regulatory regimes and governance cultures.

By shifting attention from prediction accuracy to decision quality under constraint, this work contributes to a broader agenda on responsible and governance-aware artificial intelligence in education and other regulated institutional domains.

From a decision-making perspective, improving decision quality under uncertainty requires not only better models but also richer representations of student behavior, motivating the incorporation of multimodal data sources. Therefore, another important direction for future research is the extension of the proposed framework toward a fully multimodal early-warning system. While the current study relies on structured administrative data, emerging evidence suggests that integrating heterogeneous data sources can provide earlier, richer, and more behaviorally grounded indicators of academic risk.

The integration of these multimodal signals, potentially combined with emerging health-related data streams, could significantly enhance the timeliness, robustness, and equity-awareness of pre-regulatory decision-making. Future work will therefore investigate how multimodal representations can be incorporated into cost-sensitive and regulation-aware decision frameworks while preserving interpretability, fairness, and alignment with institutional governance constraints.

4 Methods

4.1 AI-supported early-warning decision model

4.1.1 Regulatory violation risk estimation

The foundation of pre-regulatory early warning is the estimation of a student's likelihood of violating formal academic warning conditions in the near future. Unlike conventional early-warning systems that focus on long-term outcomes such as dropout, the proposed framework aligns prediction targets directly with *regulatory decision boundaries*. Artificial intelligence is used to estimate the probability that a student will violate one or more policy-defined warning conditions at the end of the current academic evaluation period.

Let $k \in \mathcal{K}$ denote the set of regulatory warning conditions defined by institutional policy. For each student i, a predictive model produces a vector of raw probability estimates

$$\hat{p}_i^{(k)} = P(\text{violate condition } k \mid x_i) \tag{1}$$

where $\mathbf{x}_i$ represents the student's academic trajectory observed prior to final evaluation, including partial semester performance, historical GPA, and progression indicators. The predictive component is model-agnostic: any supervised learning algorithm capable of producing probability estimates may be used, provided that calibration is adequate for decision support.

By framing prediction in terms of regulatory violation risk, model outputs remain directly interpretable for institutional actors. This design avoids abstract risk scores and ensures semantic alignment between predictive information and permissible institutional actions.

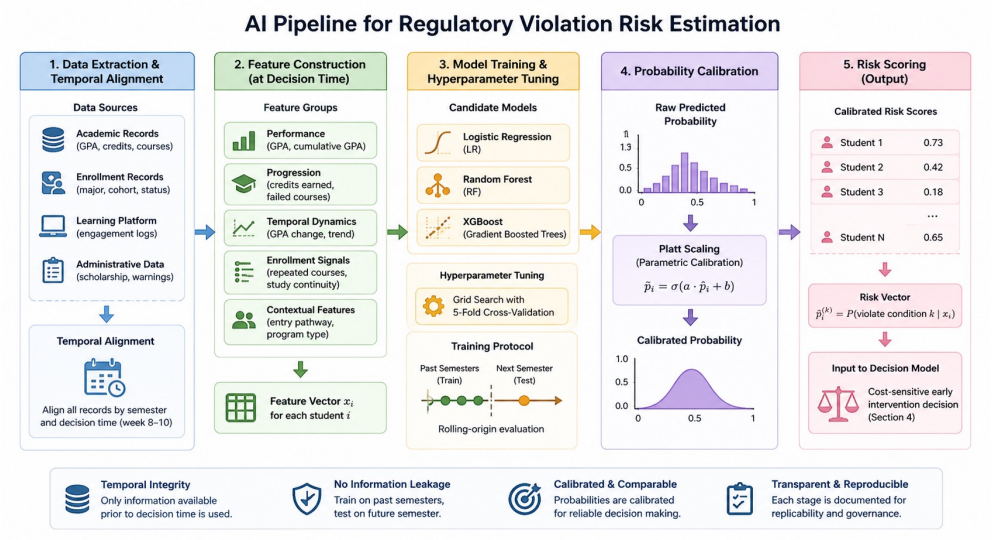


Fig. 2 AI pipeline for regulatory violation risk estimation. The pipeline transforms longitudinal student records into calibrated risk scores used in decision-making.

4.1.2 AI pipeline and implementation details

To ensure reproducibility and clarify the role of artificial intelligence within the proposed framework, we explicitly describe the end-to-end pipeline used for regulatory violation risk estimation.

Pipeline overview.

The AI component follows a standard supervised learning pipeline consisting of five stages:

1. Data extraction and temporal alignment
2. Feature construction at decision time
3. Model training and hyperparameter tuning
4. Probability calibration
5. Risk scoring for decision input

A schematic overview is provided in Figure 2.

Feature representation.

Let x_i denote the feature vector available at decision time for student i. Features are constructed from longitudinal academic records and include:

- Performance features: partial GPA, cumulative GPA
- Progression features: credit accumulation, failed credits
- Temporal dynamics: GPA change, trajectory slope
- Enrollment signals: repeated courses, study continuity

Feature construction strictly respects temporal ordering, ensuring that only information available prior to final evaluation is used.

Predictive models.

We implement three standard supervised learning models:

- Logistic Regression (LR)
- Random Forest (RF)
- Gradient Boosted Trees (XGBoost)

These models are selected to balance interpretability and predictive capacity, consistent with prior educational data mining literature.

Hyperparameter configuration.

Hyperparameters are selected via grid search with 5-fold cross-validation on the training set.

Typical configurations include:

- LR: L2 regularization, $C \in \{0.01, 0.1, 1, 10\}$
- RF: number of trees $\in \{100, 200, 500\}$, max depth $\in \{5, 10, 20\}$
- XGBoost: learning rate $\in \{0.01, 0.05, 0.1\}$, max depth $\in \{3, 5, 7\}$

Probability calibration.

Predicted probabilities are calibrated using Platt scaling:

$$\tilde{p}_i^{(k)} = \sigma(a_k \cdot \hat{p}_i^{(k)} + b_k) \tag{2}$$

where parameters (a_k, b_k) are learned on validation data for each regulatory condition k.

Training protocol.

Models are trained using a rolling-origin scheme consistent with institutional deployment:

- Train on historical semesters up to time t
- Predict on semester $t + 1$
- Repeat across cohorts

This avoids temporal leakage and reflects real-world usage.

Output.

The final output of the AI pipeline is the calibrated probability vector defined in Eq. 2, which serves as input to the decision model in Section 4.1.

4.1.3 Expected regulatory violation cost

Prediction alone does not determine whether early intervention should occur. To translate risk estimates into actionable guidance, we introduce the concept of *Expected Regulatory Violation Cost* (ERVC), which quantifies the anticipated consequences of delaying intervention until a formal academic warning is triggered.

For each student i, ERVC is defined as

$$\text{ERVC}_i = \sum_{k \in \mathcal{K}} \tilde{p}_i^{(k)} \cdot C_k(g_i) \tag{3}$$

where $C_k(g_i)$ denotes the cost associated with violating regulatory condition k for a student belonging to contextual group g_i. Costs are specified by the institution and reflect relative—not absolute—consequences, such as academic disruption, psychological stress, or increased administrative burden.

Allowing costs to vary across contextual groups captures heterogeneity in learning conditions and acknowledges that identical regulatory violations may impose unequal consequences. This formulation draws on established principles of cost-sensitive decision-making and learning, which have seen renewed interest for handling imbalanced and high-stakes domains [7, 16], where actions are evaluated based on expected downstream effects rather than predictive accuracy alone [24].

4.1.4 Pre-regulatory decision rule

Early-warning decisions are formalized through a threshold-based decision rule operating at the pre-regulatory layer. Specifically, early supportive action is initiated when the expected cost of delayed intervention exceeds a decision threshold:

$$d_i = \begin{cases} 1, & \text{if } \mathrm{ERVC}_i > \tau \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $d_i = 1$ indicates that the student should be considered for pre-regulatory support, such as academic advising or monitoring. The threshold τ is an institutional parameter reflecting resource constraints and policy priorities.

Crucially, this decision rule does not automate formal academic sanctions. Instead, it supports discretionary actions that remain fully compliant with academic regulations. By separating prediction, cost specification, and threshold selection, the framework preserves institutional control and transparency.

4.1.5 Positioning AI within institutional decision-making

The proposed framework positions artificial intelligence as a *decision-support instrument* rather than a decision-maker. AI contributes by structuring predictive information and expected costs, while human actors retain authority over normative judgments and final actions. This separation aligns with governance-oriented approaches to AI deployment in regulated environments, which emphasize accountability, oversight, and procedural legitimacy [9].

By embedding AI within a regulation-aware decision framework, early warning becomes an anticipatory institutional mechanism rather than a reactive administrative response. This positioning enables institutions to act earlier and more consistently without undermining formal academic policies.

Unlike purely predictive early-warning systems, the proposed framework explicitly separates the AI estimation layer from the decision layer, ensuring that model improvements do not compromise interpretability, governance alignment, or reproducibility. The overall pre-regulatory decision-support procedure is summarized in Algorithm 1. All computational steps in the decision pipeline explicitly reference the formal definitions introduced in Eqs. 1–4.

4.1.6 Experimental design and evaluation protocol

4.1.7 Dataset description and institutional context

The empirical evaluation is conducted using a longitudinal administrative dataset collected from a public regional university in Vietnam's Northern Midlands and Mountainous Area. This region is characterized by persistent economic disadvantage and the highest concentration of ethnic minority populations in the country.

A substantial proportion of students originate from rural and mountainous provinces, including ethnic minority groups who may face cumulative educational

Algorithm 1: Pre-regulatory Early Warning Decision Procedure

Input: Dataset $\mathcal{D}$, conditions $\mathcal{K}$, cost C_k, threshold τ
Output: Decision d_i
for *each semester* t **do**
  Extract features x_i;
  Train model $f(\cdot)$ on past data;
  Compute risk $\hat{p}_i^{(k)}$ using Eq. 1;
  Calibrate probabilities using Eq. 2;
  for *each student* i **do**
    Compute ERVC_i using Eq. 3;
    if $\mathrm{ERVC}_i > \tau$ **then**
      $d_i \leftarrow 1$;
    else
      $d_i \leftarrow 0$;
return d_i

disadvantages prior to university entry. These include disparities in prior schooling quality, language barriers, and access to academic support resources.

This institutional context is critical for interpreting early-warning decisions, as academic risk is not purely individual but shaped by heterogeneous learning conditions. Similar patterns have been documented in prior work on dropout inequality and explainable AI in education [18].

The dataset spans six consecutive academic cohorts and includes all full-time undergraduate students enrolled in regular programs. Each observation corresponds to a *student–semester* instance, reflecting the information available to the institution prior to the final academic evaluation of that semester.

For each student–semester instance, the dataset includes academic performance variables (partial semester GPA, cumulative GPA to the previous semester, number of registered credits, number of previously failed credits), trajectory variables (GPA change relative to the previous semester, credit accumulation rate, number of previously repeated courses), and program progression indicators (year of study, program

type). Contextual variables derived from admission records are used to characterize heterogeneous learning conditions.

Formal academic warning labels are defined with respect to institutional regulations and assigned at the end of each semester. A student–semester instance is labeled positive if the student violates at least one regulatory warning condition, such as falling below year-specific cumulative GPA thresholds or exceeding allowable failed credit limits.

This context allows the empirical evaluation to capture not only predictive performance but also the social implications of early-warning decisions. In particular, differences in learning conditions across student groups provide a basis for examining how decision policies may mitigate or amplify inequality in access to institutional support.

The feature space used for predictive modeling is constructed from student-level academic trajectories and institutional records. Table 4 summarizes all features available at the pre-regulatory decision point.

Table 4 Feature schema at pre-regulatory decision time

Group	Feature	Availability
Performance	Partial GPA	Week 8–10
Performance	Cumulative GPA	Start of semester
Progression	Registered credits	Start
Progression	Failed credits	Start
Temporal	GPA change	Start
Temporal	Credit accumulation rate	Start
Enrollment	Repeated courses	Start
Context	Entry pathway	Admission
Context	Year of study	Administrative

4.1.8 Data preprocessing and feature construction

Data cleaning.

Missing values are handled using median imputation for continuous features and mode imputation for categorical features.

Normalization.

Continuous features are standardized:

$$x'_j = \frac{x_j - \mu_j}{\sigma_j} \tag{5}$$

to ensure comparability across cohorts.

This standardization step is applied using training-set statistics only, as defined in Eq. 5, to avoid information leakage.

Temporal snapshot design.

All features are extracted at a fixed decision point (week 8–10 of the semester), ensuring consistency across student instances.

Train-test split.

Data is split temporally rather than randomly:

- Training: past semesters
- Testing: future semester

This design prevents information leakage and reflects deployment conditions.

4.1.9 Predictive models for regulatory violation risk

To estimate regulatory violation risk, we consider three supervised learning models commonly used in educational data mining. Logistic regression serves as a transparent baseline, while random forest and gradient boosted trees (XGBoost) represent more

expressive ensemble-based approaches. All models are trained using identical feature sets and evaluated under the same experimental protocol.

Model hyperparameters are selected via cross-validated grid search on training data. To ensure comparability across models, predicted probabilities are calibrated using Platt scaling prior to decision evaluation. Predictive performance is reported for completeness but is not the primary evaluation criterion, as the focus of this study lies in decision outcomes rather than prediction accuracy.

4.1.10 Baseline decision policies

The proposed AI-supported decision model is compared against four baseline decision policies designed to reflect both institutional practice and prediction-centric early-warning approaches.

- **Regulation-only policy (RO):** No pre-regulatory action is taken. Institutional intervention occurs only after formal academic warning thresholds are violated.
- **Uniform rule-based policy (UR):** Students are flagged for early support if their semester GPA falls below a fixed threshold, regardless of predicted risk.
- **Prediction-threshold policy (PT):** Students are flagged if their predicted probability of regulatory violation exceeds a fixed percentile cutoff.
- **Cost-insensitive predictive policy (CI):** Students are flagged if the predicted probability exceeds a fixed cutoff, without accounting for heterogeneous costs.

To ensure fair comparison, decision thresholds for all non-regulation-only policies are calibrated to achieve comparable early intervention coverage under institutional resource constraints.

4.1.11 Decision parameters and cost specification

Decision-making in the proposed framework is governed by the Expected Regulatory Violation Cost (ERVC). Cost parameters are specified in relative terms and normalized

Table 5 Relative cost specification used in ERVC computation

Regulatory condition	Other students	Structurally disadvantaged
Low semester GPA	0.60	0.75
Low cumulative GPA	0.70	0.85
Excessive failed credits	0.80	1.00

to the interval $[0, 1]$. Costs vary across contextual groups to reflect heterogeneous learning conditions and are determined in consultation with institutional stakeholders.

The cost structure underlying the ERVC formulation, which determines the relative importance of different regulatory violations across student groups, is detailed in Table 5.

The decision threshold τ is selected to satisfy an explicit resource constraint, defined as the maximum allowable proportion of student–semester instances receiving pre-regulatory support within a semester. Threshold robustness is assessed across multiple intervention coverage levels, while temporal robustness is evaluated across cohorts and predictive backbones.

4.1.12 Evaluation metrics and protocol

Evaluation focuses on decision-level performance. The primary metrics include the Missed Intervention Rate (MIR), Early Intervention Coverage (EIC), Cost-weighted Decision Loss (CDL), and the Equity Gap in Missed Intervention (EGMI) across heterogeneous learning-condition groups.

Experiments follow a rolling-origin protocol that mirrors institutional decision timelines. For each semester, models are trained using all historical data available prior to that semester, decisions are simulated using only information available at decision time, and outcomes are evaluated against end-of-semester regulatory results. All experiments are repeated across cohorts and predictive backbones, with results averaged to ensure stability.

4.1.13 Ethics approval, consent, and data protection

This study used anonymized administrative data collected as part of routine institutional processes. No personally identifiable information was used in the analysis. The study involved no experimental intervention and posed minimal risk. Formal ethics approval was not required under institutional procedures for analysis of anonymized administrative records. All procedures complied with institutional data protection regulations. Consent to participate was not applicable because the study used anonymized administrative records and did not involve direct recruitment of human participants.

Data availability

The study used anonymized administrative student records from the participating institution. The underlying student-level data cannot be made publicly available because they are protected by institutional data-governance rules and contain sensitive educational records. Aggregated tables, analysis definitions, and non-identifying methodological details can be made available from the corresponding author upon reasonable request and subject to institutional approval.

Code availability

The analysis code used to train models, calibrate probabilities, and compute decision metrics can be made available from the corresponding author upon reasonable request, subject to institutional approval and removal of institution-specific identifiers.

Acknowledgements

Not applicable.

Author contributions

T.-N.P. and T.-V.N. conceptualized the study. T.-V.N. designed the methodology, conducted the analysis, prepared the visualizations, and drafted the manuscript. T.-N.P., G.-N.N., and D.-N.L. contributed to supervision, interpretation, and manuscript review. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Consent to publish

Not applicable.

Declaration of AI use

The authors used AI-assisted tools for language editing and formatting support. The authors reviewed and edited the manuscript and take full responsibility for its content.

References

[1] Baker, R. S. J. d., Inventado, P. S. Educational data mining and learning analytics. In: *Learning Analytics*, pp. 61–75. Springer, 2014. DOI: https://doi.org/10.1007/978-1-4614-3305-7.

[2] Baker, R. S. J. d., Hawn, A. Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32:1052–1092, 2022. DOI: https://doi.org/10.1007/s40593-021-00285-9.

[3] Baldwin, R., Cave, M., Lodge, M. *Understanding regulation: theory, strategy, and practice*. Oxford University Press, 2011. DOI: https://doi.org/10.1093/acprof:osobl/9780199576081.001.0001.

[4] Bean, J. P., Metzner, B. S. A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, 55(4):485–540, 1985. DOI: https://doi.org/10.3102/00346543055004485.

[5] Bertsimas, D., Kallus, N. From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044, 2019. DOI: https://doi.org/10.1287/mnsc.2018.3253.

[6] Cecez-Kecmanovic, D. Ethics in the world of automated algorithmic decision-making: a posthumanist perspective. *Information and Organization*, 35(3):100587, 2025. DOI: https://doi.org/10.1016/j.infoandorg.2025.100587.

[7] Elkan, C. The foundations of cost-sensitive learning. In: *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence*, pp. 973–978, 2001. https://dl.acm.org/doi/10.5555/1642194.1642224.

[8] Filipova, I. A. Legal regulation of artificial intelligence: experience of China. *Journal of Digital Technologies and Law*, 2(1):46–73, 2024. DOI: https://doi.org/10.21202/jdtl.2024.4.

[9] Floridi, L., Cowls, J., Beltrametti, M., et al. AI4People—an ethical framework for a good AI society. *Minds and Machines*, 28(4):689–707, 2018. DOI: https://doi.org/10.1007/s11023-018-9482-5.

[10] Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J., Baron, J. D. Early alert of academically at-risk students: an open source analytics initiative. *Journal of Learning Analytics*, 1(1):6–47, 2014. DOI: https://doi.org/10.18608/jla.2014.11.3.

[11] Jørnø, R. L., Gynther, K. What constitutes an actionable insight in learning analytics? *Journal of Learning Analytics*, 5(3), 2018. DOI: https://doi.org/10.18608/jla.2018.53.13.

[12] Karimi, A.-H., Barthe, G., Schölkopf, B., Valera, I. A survey of algorithmic recourse: contrastive explanations and consequential recommendations. *ACM Computing Surveys*, 55(5):1–29, 2022. DOI: https://doi.org/10.1145/3527848.

[13] Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z. Prediction policy problems. *American Economic Review*, 105(5):491–495, 2015. DOI: https://doi.org/10.1257/aer.p20151023.

[14] Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., Addison, K. L. A machine learning framework to identify students at risk of adverse academic outcomes. In: *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1909–1918, 2015. DOI: https://doi.org/10.1145/2783258.2788620.

[15] March, J. G., Simon, H. A. *Organizations*. Wiley, 1958.

[16] Pes, B., Lai, G. Cost-sensitive learning strategies for high-dimensional and imbalanced data: a comparative study. *PeerJ Computer Science*, 7:e832, 2021. DOI: https://doi.org/10.7717/peerj-cs.832.

[17] Phung, T.-N., Do, D.-C., Nguyen, T.-T., Nguyen, V.-S., Nguyen, T.-V., Le, D.-N., et al. An integrated framework for outcome-based education and AI-supported blended learning in curriculum redesign and intelligent training management. *Discover Computing*, 29, 196 (2026). DOI: https://doi.org/10.1007/s10791-026-10088-y.

[18] Phung, T.-N. Using explainable AI to diagnose institutional inequality in student dropout across ethnic and regional groups in Vietnam. *AI & Society*, 2026.

[19] Quy, T. L. Fairness-aware machine learning in educational data mining, 2024. DOI: https://doi.org/10.15488/15832.

[20] Rogers, T., Dawson, S., Gašević, D. Learning analytics and the imperative for theory-driven research. In: *The SAGE handbook of e-learning research*, 2nd ed., pp. 232–250. SAGE Publications, 2016.

[21] Rudin, C., Chen, C., Chen, Z., et al. Interpretable machine learning: fundamental principles and 10 grand challenges. *Statistics Surveys*, 16:1–85, 2022. DOI: https://doi.org/10.1214/21-SS133.

[22] Siemens, G., Long, P. Penetrating the fog: analytics in learning and education. *EDUCAUSE Review*, 46(5):30–40, 2011.

[23] Tinto, V. *Leaving college: rethinking the causes and cures of student attrition*. University of Chicago Press, 1993. ISBN: 0-226-80449-6.

[24] Turney, P. D. Types of cost in inductive concept learning. In: *Proceedings of Workshop on Cost-Sensitive Learning at the Seventeenth International Conference on Machine Learning*, pp. 15–21, 2000. DOI: https://doi.org/10.48550/arXiv.cs/0212034.

[25] Ustun, B., Spangher, A., Liu, Y. Actionable recourse in linear classification. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 10–19, 2019. DOI: https://doi.org/10.1145/3287560.3287566.

[26] Vinh, N.-T., Phung, T.-N., Cuong, D.-D. A bibliometric and thematic analysis of systematic reviews of artificial intelligence in education. In: *Advances in Information and Communication Technology, ICTA 2023*. Lecture Notes in Networks and Systems, vol. 848. Springer, Cham (2024). DOI: https://doi.org/10.1007/978-3-031-50818-9_37.

[27] Wachter, S., Mittelstadt, B., Russell, C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2):841–887, 2017. DOI: https://doi.org/10.48550/arXiv.1711.00399.

[28] Wiese, L. J., Patil, I., Schiff, D. S., Magana, A. J. AI ethics education: a systematic literature review. *Computers and Education: Artificial Intelligence*, 8:100405, 2025. DOI: https://doi.org/10.1016/j.caeai.2025.100405.

[29] Yeung, K. Algorithmic regulation: a critical interrogation. *Regulation & Governance*, 12(4):505–523, 2018. DOI: https://doi.org/10.1111/rego.12158.