



Early Warning Systems for Academic Performance Based on Student Profile Analysis: A Survey and a Conceptual Explainability-Oriented Knowledge-Distillation Framework with a Design Prototype

Cao Minh Thanh¹, Cao Thai Phuong Thanh¹, Tran Son Hai^{2,*,} 

¹Faculty of Information Technology, Saigon University, Ho Chi Minh City, Vietnam

²Faculty of Information Technology, Ho Chi Minh City University of Education, HCMC, Vietnam

Article History

Received: 19 August, 2026

Revised: 01 October, 2026

Accepted: 05 October, 2026

Published: 20 October, 2026

Abstract:

Introduction: The rising number of academic warnings and forced withdrawals in tertiary institutions, particularly in credit-based Vietnamese universities, calls for the implementation of early warning systems (EWS) that predict academic risk based on student profiles obtained periodically. Organizations require accurate and understandable forecasts that enable academic counsellors to translate every notification into practical advice.

Methodology: This paper reviews the existing literature on profile-based EWS for academic performance, and as a conceptual (non-empirical) contribution, it presents a transparent framework of knowledge distillation with a prototype of a client-side design. A three-block structure was then defined using a design science process, and an HTML/JavaScript prototype of the client side was constructed.

Results: This discussion indicates that tree ensembles are the best predictors of tabular student profiles, that class imbalance and realistic temporal validation are significant prerequisites for practice, and that the explainability of prior studies is now compulsory, and SHAP-style attribution and interpretable models are at the forefront. In contrast, limited research based on thorough implementations of EWS with that based solely on models. The proposed system incorporates an instructor group with a simplified, understandable student model and regulation-mindful warning systems. The prototype features personal student risk evaluations, factor-specific explanations, and advisory implications.

Conclusion: There was no model training, real data, or empirical analysis presented in this study; the framework, its exemplifying weights, and the prototype were a design proposal, and a specific evaluation protocol was outlined in future experiments regarding institutional data.

Keywords: Anticipatory alert mechanism, educational data analytics, forecasting student achievement, knowledge compression, interpretable artificial intelligence, tertiary education, quality learning (sustainable development goal 4).

1. INTRODUCTION

Colleges provide a large amount of administrative information on students, including records of their entry and grades in courses on a semester-by-semester basis. However, a high percentage of enrolled students remain incomplete or leave without a degree. Educational Data Mining (EDM) and

Learning Analytics (LA) have rapidly become disciplines that convert such data into useful information [1]. Academic warning is a formal process in Vietnam. Under Circular No. 08/2021/TT-BGDĐT, students will receive a warning at the end of each main semester in case of a failed-credit volume, credit debt accumulated, semester Grade Point Average (GPA), or cumulative GPA lower than the regulated levels. Students who

*Address correspondence to this author at Faculty of Information Technology, Ho Chi Minh City University of Education, HCMC, Vietnam; E-mail: haits@hcmue.edu.vn



© 2026 Copyright by the Authors.

Licensed as an open access article using a [CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/).

are warned multiple times are expelled [2]. Approximately 33% of the 2015–2020 group of Saigon University (SGU) undergraduates received at least one academic warning, and approximately 10% were forced to drop out. Cohorts that completed their training cycle had an attrition rate of approximately 20–27%.

The current system of warning is backward-looking; therefore, the advisers will intervene only after the failure has been realized, and there is little that can be done to rectify the situation because the credit debt has already accumulated. An Early Warning System (EWS) is the opposite of this reasoning: It is designed to forecast, based on the student profile at time t , which students will be in danger of warning states in the following semesters, to enable advising and course-registration support to be given to these students before they fall into the trap of failing. In the last 20 years, cross-country studies have produced an impressive range of prediction models of student performance and dropout [3, 4] and a more limited range of operational EWS with dashboards and intervention processes [5–9]. However, three requirements that determine real-world applicability are unevenly addressed. Initially, models have to deal with the unbalanced, tabular, and time-varying characteristics of institutional data. Second, we require deployment (training on previous cohorts, testing on future cohorts) and not random splits. Third, and most importantly, all alarms must be intelligible: a caution without a reason can accomplish little to educate, and the counsellors will not believe and act on unintelligible prophecies [10, 11]. There are two contributions to this paper. (i) It presents an overview of a study on profile-based academic-performance Early Warning Systems (EWS) research with four axes: data sources, prediction models, learning difficulties, and implementation with trust in relation to the Vietnamese credit-based context. (ii) To bridge the gap between opaque but accurate ensembles and interpretable but weaker simple models, it recommends a framework that learns an explainability-oriented Knowledge Distillation (KD) to train an ensemble of teachers into a student model that is both interpretable and classifies academic risk as well as the source of each alert.

The prototype of the framework that we have implemented is a client-side framework, which includes a rule engine that is sensitive to regulations, a transparent distilled score model, explanations of the scores per factor, and advising recommendations. It is a design proposal survey and does not present new empirical training results, and the scoring weights of the prototype are illustrative and awaiting to be trialled on institutional data.

Accordingly, the paper should be read as a conceptual/position contribution in the design-science tradition: the novelty of explainability-oriented knowledge distillation for academic early warning is proposed and argued on the basis of survey evidence, not empirically demonstrated, and Section 4.3 specifies the evaluation protocol under which the proposal will be tested. The rest of the paper is structured as follows: Section 2 covers the survey procedure and the prototype design technique; Section 3 gives the survey results; Section 4 details the suggested framework and prototype; Section 5. Implications and constraints are discussed, and Section 6 ends.

2. METHODOLOGY

2.1. Survey Design and Source Selection

Research design: This study is a systematic literature review (scoping-type) of the literature on the prediction of academic performance and early warning using student profile information. It was challenging to conduct a complete PRISMA-compliant systematic review with a meticulous screening process in the present study. Instead, the survey is informed by representative, reputable, and indexed sources and classifies them into an analytical taxonomy.

Publisher platforms indexed in either Web of Science and/or Scopus (Wiley, Elsevier/ScienceDirect, Springer, Taylor and Francis, MDPI, IEEE, ACM) were searched with a combination of the following terms: “early warning system,” “at-risk students,” “student performance prediction,” “dropout prediction,” “educational data mining,” “knowledge distillation,” “explainable AI,” covering 2015–2026, and with additional publications as necessary. National training guidelines and Vietnamese studies were integrated to develop the local context. Moreover, conducting it, bibliographic metadata (authors, venue, year, DOI) of new sources that were added was compared to publisher records. Inclusion and exclusion criteria: The studies that matched the criteria (a) predicted academic performance in higher education (performance classification, at-risk identification, or attrition) with the help of the institutional profile data or (b) reported on the use of Early Warning Systems (EWS).

Studies that only included K-12 were selected; only behavioral MOOC clickstream analysis with no profile features, and gray literature that is not included were excluded. The final list will include the 31 references used in this paper, 25 of which are based on the 2010 doctorate overview report, which is the foundation of this study, and 6 of which have been newly obtained and validated. Unit of analysis and synthesis: There are four dimensions that were categorized per source: (D1) Data sources used, (D2) Model classification, (D3) Learning obstacles addressed (imbalance, drift, validation framework), (D4) Implementation and trust factors (explainability, dashboards, interventions). The claims of efficacy and the rates of occurrence are deliberately separated, as the common approach is not taken as evidence.

Review process: The review was not a PRISMA-compliant systematic review, but the conduct of the review adhered to PRISMA-ScR principles, as appropriate. Searching was done in Scopus, Web of Science, and other publisher sites with the Boolean combinations of (early warning system) OR (at-risk students) AND (student performance prediction) OR (dropout prediction) AND (higher education) and (knowledge distillation) OR (student) OR (education). The initial screening was conducted in two phases: titles and abstracts were sorted against the inclusion criteria outlined below by the first author, and then candidate full texts were sorted, with any questions resolved through discussion among all the authors. The most common exclusion reasons were K-12-only settings, MOOC clickstream-only features, and non-indexed venues. Because 25 of the sources are based on the doctoral overview report, on

which this work is constructed, and the number of searches at the record level is not recorded, we do not present a PRISMA flow diagram but intentionally discuss the result as a structured narrative review; a protocolized systematic review with registered searches and a PRISMA flow chart would be considered in the future.

2.2. Prototype Design Methodology

The presented solution was designed following a design science strategy: problem identification (historical alarms, uncertain forecasting), prescription (a framework to distil information with an emphasis on explainability), invention (a prototype on the client side), and preparation to assess it. In line with the practice of research in design science, the artefact mentioned here is an expository instantiation, the intention of which is to report the proposed design and give an account of it; its summative (empirical) assessment is deliberately pushed off to the protocol in Section 4.3. The prototype was deployed as a single stand-alone HTML/JavaScript app, which does not need a server, such that the interaction model can be tested by any browser by advisors.

It consists of (i) a deterministic rule engine that instantiates the warning criteria of Circular 08/2021 [2], (ii) a linear scoring function that is clearly marked as demonstrative, and (iii) explanatory and advisory views. The overall experiments were assigned to statistical software, in the form of Python 3 with scikit-learn and XGBoost, except the former. No further inferential statistics were needed because the SGU dataset, as

described in Section 3.1, was analyzed descriptively and correlated as shown in this paper. Ethics: This article combines the available research and regulatory literature with a prototype that works. Only user-supplied or artificial data; no personal student data were obtained or used in this research, consequently dispensing with the need to have ethics approval.

3. RESULTS

The taxonomy resulting from the coding of the corpus is summarized in Fig. (1). The four subsections below present the synthesis along its dimensions; Table 1 compares representative primary studies.

3.1. Data Sources: From Single Scores to Longitudinal Profiles (D1)

Although the original studies took the results based on the characteristics of the admission alone [12], the findings suggest that there is a tendency towards longitudinal assessments. Marbouti *et al.* [5] showed that grades based on standards in the course can be used to predict reliably at the beginning of the semester [5]; Yağcı [4] stated that mid-term grades can be used to predict the results of the final exam with 70–75% accuracy [4]; Miguéis *et al.* [6] distinguished engineering students using only the first-year data and achieved an ensemble accuracy of more than 95% [6]. Entrance is gradually being added to program-level studies exam scores, semester GPA patterns, and patterns of credit accumulation [9, 10], and behavioral LMS data where available [11].

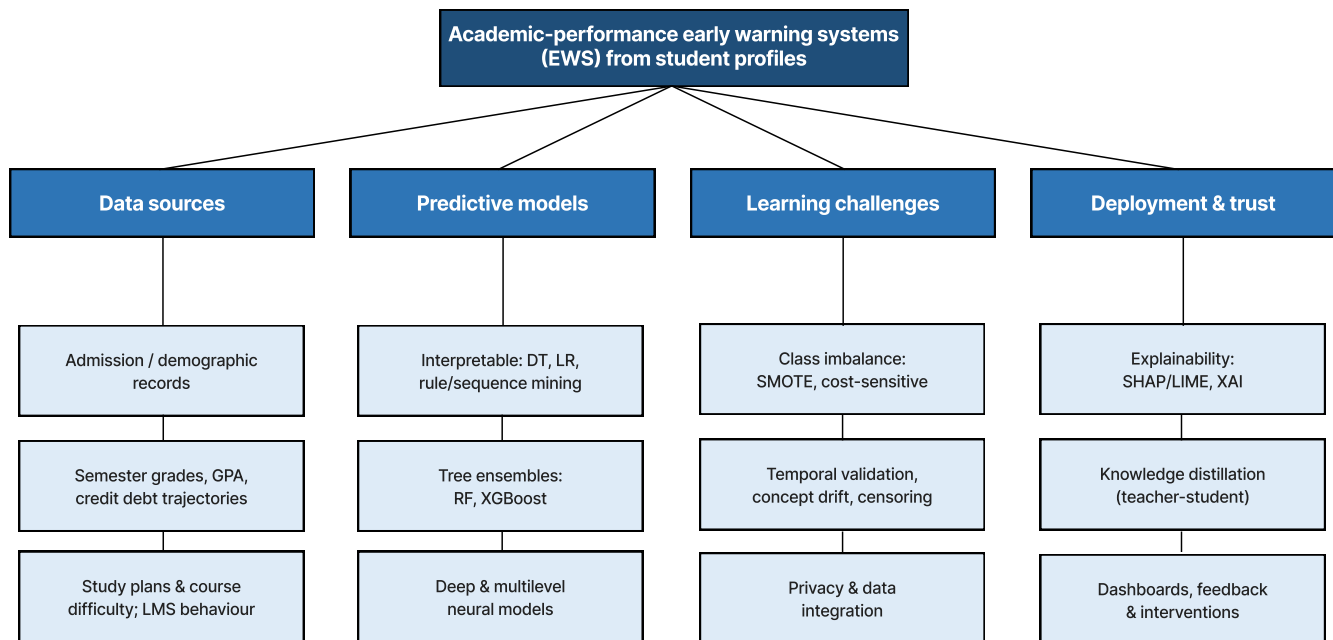


Fig. (1). Taxonomy of profile-based academic-performance early warning research synthesized in this survey. EWS: early warning system; DT: decision tree; LR: logistic regression; RF: random forest; LMS: learning management system; SMOTE: synthetic minority over-sampling technique; SHAP: Shapley additive explanations; LIME: local interpretable model-agnostic explanations; XAI: explainable artificial intelligence.

Table 1. Representative main papers on profile-based academic risk prediction and early warning synthesized in this study (higher education settings; publication years 2016–2025).

Study	Data	Models	Validation and Imbalance	Explainability
Marbouti <i>et al.</i> [5]	Standards-based grades, USA	LR, DT, NB	In-course early horizons	Feature-level discussion; no system
Miguéis <i>et al.</i> [6]	More than 2,400 engineering students, data, Portugal	RF	Holdout; accuracy more than 95%	Segmentation for advising
Akçapınar <i>et al.</i> [7]	LMS interactions, online course, Türkiye	Multiple ML classifiers	Week-by-week horizons	Early-warning framing; alerts
Bañeres <i>et al.</i> [8]	Institutional data mart, Spain	NB, DT, KNN, SVM	Semester-wise testing; quality thresholds	Full EWS: dashboards, traffic lights.
Çırak <i>et al.</i> [9]	Records of 11,698 freshmen, Türkiye	ML classifiers	1-year horizon	Policy factor analysis
Cannistrà <i>et al.</i> [10]	Administrative data, Italian university	Multilevel statistics and ML	Cohort-based early prediction	EWS design
Waheed <i>et al.</i> [11]	VLE big data (OULAD-scale)	Deep networks vs. LR, SVM	Standard splits	Accuracy-focused
Ghorbani and Ghousi [18]	Student benchmark data	ML + resampling variants	Systematic imbalance comparison	-
Nagy and Molontay [19]	University dropout, Hungary	Gradient boosting and XAI	Institutional data	SHAP-based personalized advice; user study
Mustafa <i>et al.</i> [20]	Records of 4,000 student	Hybrid LR–NN	Accuracy ~96% reported	SHAP and LIME risk-factor analysis
Nguyen <i>et al.</i> [27]	Credit records, Vietnam	Sequential rules	Historical records	Interpretable rules for advising

Abbreviation: LR: logistic regression; DT: decision tree; NB: naïve Bayes; RF: random forest; KNN: k-nearest neighbors; SVM: support vector machine; NN: neural network; ML: machine learning; VLE: virtual learning environment; OULAD: Open University Learning Analytics Dataset; XAI: explainable artificial intelligence; SHAP: Shapley additive explanations; LIME: local interpretable model-agnostic explanations; EWS: early warning system.

The first look at the SGU data applicable to the doctoral study demonstrates a regular tendency: entrance-score attrition rates can be related to quartiles (we find 20.4% attrition in the lowest quartile and 9.3% attrition in the highest; point-biserial $r = -0.128$, $p < 0.001$), but the effect size is small, indicating that the data of admissions do not provide strong predictive power for an Early Warning System. The multi-faceted profile of admission characteristics is informed by this observation. Semester patterns and study-plan items, such as historical failure rates of courses being taken, are discussed in Section 4.

3.2. Predictive Models: Tree Ensembles as the Tabular Benchmark (D2)

Despite the fact that ensemble techniques prevail in terms of accuracy, interpretable classical models (*e.g.*, decision trees), logistic regression, sequential rules) can still be appealing to the corpus because of their advisory nature. Comparative studies on university dropout always place random forest and gradient boosts at the top or at the forefront [13, 14]. Waheed *et al.* [11]

demonstrated the superiority of deep models over logistic regression and SVM when applied to large virtual-learning-environment datasets [11]. Two systematic studies in the education domain compare institutional data, where tree-based ensembles consistently outperform specialised deep architectures with less tuning effort on traditional tabular datasets [15, 16], with XGBoost being the prevailing benchmark [17]. The EWS of Cannistrà.

3.3. Learning Challenges: Imbalance, Drift, and Honest Validation (D3)

Addressing imbalance is considered vital and not superficial in the literature. The specific category that an Early Warning System must identify is the minority group, pupils at risk of warning or withdrawal. Ghorbani and Ghousi have systematically studied resampling methods (SMOTE, Borderline-SMOTE, SVM-SMOTE) and reported significant impacts on the recall of the minority class [18]. Other methods include cost-sensitive learning and decision threshold adjustment [3].

The second ongoing challenge is temporal validity. As programs, regulations, and student demographics change over time, models validated through random partitioning overestimate real-world performance. Deployment-appropriate designs train on completed cohorts and evaluate later cohorts. When labels are based on final outcomes, right-censoring must be addressed for ongoing cohorts. Third, privacy and data aggregation pose concrete challenges: profiles are present in different formats across admission, training management, and evaluation systems, and anonymization is required for research use [1, 3]. Çırak *et al.* [9] used the freshman-year-specific frameworks on 11,698 datasets of freshmen, suggesting that a careful feature extraction within a well-defined early period may partially bypass drift, while still being useful at the institutional level.

3.4. Deployment and Trust: Explainability and Complete Systems (D4)

EWS was implemented in only a small minority of the works in the corpus. Bañeres *et al.* [8] describe a full framework in a virtual university, which includes risk models for every activity, quality thresholds that regulate the categories of messages, dashboards for both students and instructors, and feedback systems [8]. Traffic-light visuals inspired by Course-Signals became the dominant metaphor for interaction [8, 9]. Akçapınar *et al.* [7] showed the feasibility of alerts from the third week of an online course [7]. Recent research highlights the need for explainability to build confidence. Nagy and Molontay [19] introduce SHAP-based local explanations, which are translated into targeted intervention suggestions for dropout prediction, using user research with students and decision-makers [19]; Mustofa *et al.* [20] integrate SHAP and LIME evaluations with a hybrid dropout model to unveil risk determinants [20]; and the comprehensive 72-study PRISMA review by Boujmiraz *et al.* [21] is a history of the parallel evolution of machine learning, deep learning, and explainable artificial intelligence in performance forecasting, with the quality of explanation now viewed as a separate review standard [21].

Another method of ensuring reliable predictions would be to ensure that the model used is interpretable yet has a precise ensemble. Knowledge distillation, first proposed by Hinton *et al.* [22] and further refined by Gou *et al.* [23], facilitates the transfer of predictive prowess from a robust teacher (or a collective of teachers) to a more compact student through temperature-adjusted output distributions; it is part of the larger model-efficiency toolkit that includes pruning and quantization [24-26].

Response-based distillation is essentially education-based, not requiring architecture on the instructor side, and allows a trade-off between XGBoost, random forest, and neural models to learn a shallow decision tree or a sparse linear learner. However, no study in the literature uses the concept of knowledge distillation with an explainable emphasis on the problem of issuing academic warnings based on institutional profiles; the literature on distillation of knowledge emphasizes compression, and the literature on the early warning system attaches explainable artificial intelligence to opaque models

post-hoc. Section 4 discusses these discrepancies. This is the discrepancy discussed in Section 4.

3.5. The Vietnamese Context

National research is at a rudimentary level and reflects the worldwide trend. Nguyen *et al.* [27] were the first to build a sequential-rule prediction of academic outcomes using the precondition structure of credit-based programs, producing understandable advice for counselling.

According to Cao *et al.*, [28] universities in Ho Chi Minh City have a large amount of training data, but academic management is still semi-manual, and automated warnings are a significant area of digital transformation [28]. Triet *et al.* [29] show that technological intervention has a significant impact on learning motivation, which is one of the factors that the warning system's recommendations can affect [29]. The regional momentum is met by the basic EDM/LA framework [1, 30] and preliminary intervention data [31]. There are three typical gaps: (1) most domestic works end up with model building with no real warning system; (2) modern model-optimization methods, e.g., knowledge distillation, are absent; (3) the combination of multi-source profiles, including admission scores, semester progressions, and academic plans, is not well studied in the Vietnamese credit system.

4. PROPOSED CONCEPTUAL FRAMEWORK AND ILLUSTRATIVE DESIGN PROTOTYPE

4.1. Explainability-Oriented Distillation Framework

In Fig. (2), the proposed framework, which operationalizes the need mentioned in Section 3.4 of the Vietnamese credit-based setting, is provided. Block 1 (profile and features). At time t , there are three sets of features for every student: admission features (normalized entrance score by admission method, priority attributes); trajectory features (cumulative GPA, sequence of semester GPAs, number of credits earned vs. number of credits owed in a semester, trend indicators); and study-plan features (registered credit load, historical failure rates of courses registered in). SGU data show that failure is concentrated in identifiable math and foundational courses, with course-level failure rates ranging between 19% and 31%.

First, Block 2 (explainability-oriented KD) chooses the teacher empirically from tree ensembles (XGBoost, random forest) and deep networks, or an ensemble of them, according to the tabular-data evidence [15-17]; teacher predictions are softened with temperature T and a student model with an interpretable structure, a depth-limited decision tree, or a sparse linear model, is trained on the combined objective $L = \alpha \cdot \text{CE}(\text{student}, \text{hard labels}) + (1 - \alpha) \cdot T^2 \cdot \text{KL}(\text{student} \parallel \text{teacher soft labels})$, transferring response-based knowledge [22, 23]. The aim, to be experimentally validated, is to obtain a student model that is as accurate as the teacher, while every alarm can be tracked back to explicit feature conditions or coefficients. Block 3 (warning services) combines the model output with the deterministic warning rules of Circular 08/2021 [2]. It provides per-student risk levels, factor-level explanations, and recommendations (retake planning, credit-load adjustment)

exposed to advisors and managers through dashboards in the spirit of deployed EWS [8, 19].

4.2. Prototype: EduAlert (HTML/JavaScript)

In an effort to instantiate the interaction model in a tangible way prior to complete experimentation, the framework's Block 3 is instantiated as EduAlert, a single-page HTML/JavaScript design mock-up (interaction prototype) presented with this paper; it contains no trained model and makes no predictive-performance claims (file `edualert_prototype.html`). It is all client-side, with no dependencies. Its components are aligned to the framework as follows:

(1) Input profile The advisor accesses the admission profile (admission score on the scale of 30, admission method), per-semester table (credits registered, credits failed, GPA per semester and cumulative GPA on the scale of 4, accumulated credit debt) and the next-semester study plan (credits registered, for each planned course its historical failure-rate band). Click to load a demo profile.

(2) Rule engine for regulations. The prototype deterministically evaluates the alert conditions of Circular 08/2021 [2]: credits failed more than 50% of registered credits in the semester, credit debt more than 24 credits, semester GPA less than 0.8 (first semester) or 1.0 (subsequent semesters), and cumulative GPA less than 1.2/1.4/1.6/1.8 by the study year. Each of the triggered conditions is given in its pure form so that the retrospective institutional status would always appear in the prediction.

(3) Scoring model for distillation (example). The student model, which is distilled, is a transparent linear-logistic scoring model: visible signed weights are added to the standardized features (cumulative GPA gap, GPA trend, credit-debt ratio, entrance-score percentile, planned-load pressure, planned-course difficulty index) to form a risk probability and a three-level risk class (normal, at risk of warning, high risk). All the weights are displayed in the interface and indicated as placeholders to show that they will be altered with the coefficients, depending on the institutional data. The prototype shows an explanation contract, as opposed to a trained model.

(4) Recommendation explanation and opinion. In both tests, the prototype produces (i) a risk gauge, (ii) a signed risk score (iii) a factor-contribution bar chart (the per-feature terms of the linear score), (iv) a plain language explanation sentence of each dominant factor, (v) recommendations based on rules (*e.g.*, prioritized retake of failed foundational courses, cut down of scheduled credit load at a level of difficulty that is high, referral to advising at a negative GPA trend). The cohort tab shows a view of the manager dashboard, with a demo table of synthetic students with risk badges that can be sorted. This design maintains the eventual substitution surgical: the replacement of the illustrative coefficient vector (or shallow tree) does not change any other component; and the rule engine ensures that no student in a controlled warning state is ever overlooked by the system, no matter how the model behaves. The model brings in the aspect of anticipation; the regulations ensure conformity.

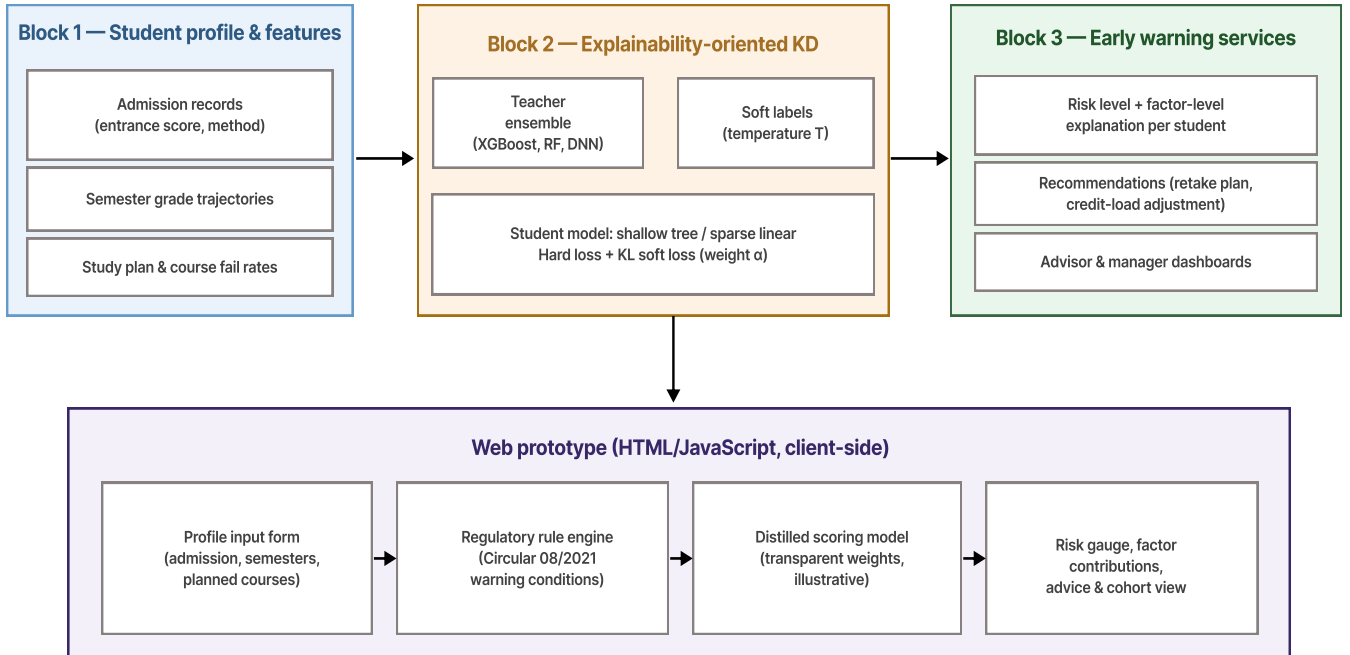


Fig. (2). Proposed three-block framework (top) and the corresponding client-side web prototype pipeline (bottom). KD = knowledge distillation; RF = random forest; DNN = deep neural network; KL = Kullback–Leibler divergence; α = loss-mixing weight; T = distillation temperature.

4.3. Planned Evaluation Protocol

To render it falsifiable, we pre-specify the protocol that we will use to test the conceptual proposal. Dataset: de-identified SGU student data of the cohort of 2015–2023 (admission, semester trajectories, study plans) with institutional data governance permission. Validation design: time splits that are deployment-realistic, training on cohorts that have been completed (*e.g.*, 2015–2020) and testing on cohorts that begin later (2021 and later) and right-censoring of cohorts (when they are still ongoing). Baselines: (a) the teacher (XGBoost/random forest, tuned); (b) the teacher (depth-limited tree; sparse logistic model) trained on hard labels, but not distilled; (c) the teacher using the post-hoc SHAP explanations.

Hypotheses: H1, the distilled student performs better than the directly trained student of the same ability; H2, the distilled student performs in a narrow band of performance as compared to the post-hoc attributions on the teacher; H3, it has more precise explanations of the deployed decision performance as compared to the post-hoc attributions on the teacher. Metrics; per-class recall (with focus on the minority warned/withdrawn classes), macro-F1, AUC, and calibration, and imbalance-sensitive training (resampling, or cost-sensitive losses). The usefulness of explanations: a research study of academic advisors that compared distilled factor-level explanations with SHAP outputs on the same cases, of think-alouds and a standardization of the explanation-satisfaction scale, in the user-study format of Nagy and Molontay [19]. Such experiments will only be followed by the making of predictive or usability claims of the framework.

5. DISCUSSION

The survey proposes that bright picture, image-based academic risk prediction has already attained the technological maturity of a model scale, but imbalance-conscious training, simulation confirmation of the deployment environment, and interpretability are the measures of reliable implementation. The conceptual form of the main hypothesis of the present paper, which will be tested in Section 4.3, is that these conditions can be satisfied simultaneously through explainability-oriented distillation, rather than one of the following: an accurate black box (with post-hoc XAI) or a weaker transparent model. This position is different from the two adjacent bodies of literature.

Compared to the XAI-on-black-box framework [19–21], a distilled interpretable student provides a decision function with auditable properties without the post-hoc interpretation issues that can be encountered otherwise. Compared to the compression-oriented KD literature [22–26], attention is paid to the accuracy-interpretability trade-off in the advisory problem, which is negligible in the case of tens of thousands of tabular records. Tabular-data evidence [15–17] also warns against the fallacy of imagining the instructor to be a deep network; in institutional profiles, the teacher will be a tree ensemble, and the summary by responses will naturally encourage this. The Vietnamese credit system has three tangible implications.

First, prediction tethered to the advised warning situations endows alerts with an institutional connotation and makes them

readily combinable with training-management software, which is embraced as a national push for digital transformation. Second, the features of study plans are an under-extracted, locally available indicator: since failure often occurs in certain courses, the organization of further attendance can be used to predict it, which can be immediately addressed with the help of registration counselling. Third, the quality of explanations should be measured with advisers, and not necessarily with the help of metrics as is the case with user-study patterns as developed in [19]. The prototype is a design of the interaction and explanation, but without a trained model; the coefficients are illustrative and all the desired statements of predictive performance of the framework (student accuracy close to that of the teacher; distillation better than simple training of the same student architecture) are just a conjecture to be checked in further experiments with the SGU data in deployment-realistic, cohort-based experiments with imbalance-sensitive measures (per-class recall, macro-F1, confusion matrices). Section 4.3 pre-specifies these experiments making sure that the present proposal can be verified or disproved and not only rephrased. The difference in retention in the SGU data has motivated the use of equality in the assessment of gender and academic departments before an institution can use it.

CONCLUSION

We have performed a review of the literature on early warning systems for academic achievement based on student profiles. We have classified 31 indexed and regulatory references related to data, models, learning barriers, and implementation with trust. We have proposed a knowledge-distillation framework focused on explainability, illustrated by a client-side design prototype. It is concluded that (i) three ensembles are the state of the art concerning institutional tabular data accuracy, (ii) the real requirement in an EWS is minority classes recall in deployment-relevant validation, and (iii) explainability is both a requirement of deployment and no longer an option. By satisfying both requirements simultaneously, consolidating the ensemble knowledge into a student model that can be comprehended satisfies both of these needs and has not been previously used in the context of establishing academic warnings in Vietnam.

The regulation-based rules, definite risk assessment, articulate description of factors, and advisory recommendations are combined into one single tool, which can be utilized by advisors in the prototype, which enables the next substitution of the model component with the optimized classifier identified during experiments without the need to redesign the tool.

LIMITATIONS AND FUTURE DIRECTION

The review is not a PRISMA-conformant systematic review, but a narrative synthesis and is conceptual and has not been empirically tested on actual institutional data; the prototype is not a working trained system, but a demonstration of what a system of this kind should be like; and the exploratory nature of explainability-based distillation in academic early warning is thus suggested as opposed to evidence-based. The next step will be the training and refining of models trained on the Saigon

University data with cohort-centric measurements, advisor-centric research on the usefulness of explanations, fairness in the demographic and faculty population, and the implementation of the system into the institutional training-management software that will transform the framework into the practical improvement of educational quality and SDG 4.

LIST OF ABBREVIATIONS

DT	= Decision Tree
EDM	= Educational Data Mining
EWS	= Early Warning Systems
GPA	= Grade Point Average
HTML	= HyperText Markup Language
KD	= Knowledge Distillation
KNN	= K-Nearest Neighbors
LA	= Learning Analysis
LR	= Logistic Regression
LMS	= Learning Management System
LIME	= Local Interpretable Model Agnostic Explanations
ML	= Machine Learning
MOOC	= Massive Open Online Course
NB	= Naïve Byes
NN	= Neural Network
OULAD	= Open University Learning Analysis Dataset
RF	= Random Forest
SVM	= Support Vector Machine
SHAP	= Shapley Additive Explanations

AUTHORS' CONTRIBUTIONS

C.M.T. Conceptualization, Methodology, Investigation, Writing original draft. C.T.P.T.: Supervision, Validation, Writing review & editing. T.S.H.: Supervision, Methodology, Writing review and editing. All authors have read and approved the final manuscript.

ETHICAL APPROVAL AND INFORMED CONSENT

Not applicable. This study involved a literature search and design recommendations based on published materials and regulatory articles. The prototype used only user-provided or synthetic data. No human participants were involved, and no personal student data were collected.

AVAILABILITY OF DATA AND MATERIALS

No datasets were generated or analysed for this survey. The prototype source code http://eduaitch.site/PoCs/edualert_prototype.html is available from the corresponding author upon reasonable request.

FUNDING

This research is supported by Sai Gon University under grant number **CSB.2026.030**.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interests.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Saigon University for supporting the data for this research.

DECLARATION OF AI

Authors only used an AI assistant (Claude, Anthropic) as an aid to organizing the literature, drafting, and writing of the prototype code. The authors thoroughly guided and revised this whole procedure. Authors completely accept culpability of the accuracy, originality, and integrity of the manuscript and whatever it comprises.

REFERENCES

- [1] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowl. Discov.*, vol. 10, no. 3, Art. no. e1355, 2020, <https://doi.org/10.1002/widm.1355>
- [2] Ministry of Education and Training of Vietnam, "Circular No. 08/2021/TT-BGDĐT promulgating the Regulation on undergraduate training," Hanoi, Vietnam, 2021. Available from: <https://congbao.chinhphu.vn/van-ban/thong-tu-so-08-2021-tt-bgdtdt-33673.htm>
- [3] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Appl. Sci.*, vol. 11, no. 1, Art. No. 237, 2021, <https://doi.org/10.3390/app11010237>
- [4] M. Yağcı, "Educational data mining: Prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, Art. No. 11, 2022, <https://doi.org/10.1186/s40561-022-00192-z>
- [5] F. Marbouti, H. A. Diefes-Dux, and K. Madhavan, "Models for early prediction of at-risk students in a course using standards-based grading," *Comput. Educ.*, vol. 103, pp. 1–15, 2016, <https://doi.org/10.1016/j.compedu.2016.09.005>
- [6] V. L. Miguéis, A. Freitas, P. J. V. Garcia, and A. Silva, "Early segmentation of students according to their academic performance: A predictive modelling approach," *Decis. Support Syst.*, vol. 115, pp. 36–51, 2018, <https://doi.org/10.1016/j.dss.2018.09.001>
- [7] G. Akçapınar, A. Altun, and P. Aşkar, "Using learning analytics to develop early-warning system for at-risk students," *Int. J. Educ. Technol. High. Educ.*, vol. 16, no. 1, Art. No. 40, 2019, <https://doi.org/10.1186/s41239-019-0172-z>

- [8] 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," *Appl. Sci.*, vol. 10, no. 13, Art. no. 4427, 2020, <https://doi.org/10.3390/app10134427>
- [9] C. R. Çırak, H. Akıllı, and Y. Ekinçi, "Development of an early warning system for higher education institutions by predicting first-year student academic performance," *High. Educ. Q.*, vol. 78, no. 4, Art. no. e12539, 2024, <https://doi.org/10.1111/hequ.12539>
- [10] M. Cannistrà, C. Masci, F. Ieva, T. Agasisti, and A. M. Paganoni, "Early-predicting dropout of university students: An application of innovative multilevel machine learning and statistical techniques," *Stud. High. Educ.*, vol. 47, no. 9, pp. 1935–1956, 2022, <https://doi.org/10.1080/03075079.2021.2018415>
- [11] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Comput. Human Behav.*, vol. 104, Art. no. 106189, 2020, <https://doi.org/10.1016/j.chb.2019.106189>
- [12] D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, "Application of machine learning in predicting performance for computer engineering students: A case study," *Sustainability*, vol. 11, no. 10, Art. no. 2883, 2019, <https://doi.org/10.3390/su11102833>
- [13] J. S. Gil, A. J. P. Delima, and R. N. Vilchez, "Predicting students' dropout indicators in public school using data mining approaches," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 1, pp. 774–778, 2020, <https://doi.org/10.30534/ijatcse/2020/110912020>
- [14] J. Kabadjov and M. Drik, "Towards predicting students' dropout in university courses using different machine learning techniques," *Appl. Sci.*, vol. 11, no. 7, Art. No. 3130, 2021, <https://doi.org/10.3390/app11073130>
- [15] L. Grinsztajn, E. Oyallon, and G. Varoquaux, "Why do tree-based models still outperform deep learning on typical tabular data?" in *Adv. Neural Inf. Process. Syst.* 35 (NeurIPS 2022), Datasets and Benchmarks Track, New Orleans, LA, USA, 2022, Available from: https://openreview.net/forum?id=Fp7_phQszn
- [16] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Inf. Fusion*, vol. 81, pp. 84–90, 2022, <https://doi.org/10.1016/j.inffus.2021.11.011>
- [17] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794, 2016, <https://doi.org/10.1145/2939672.2939785>
- [18] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020, <https://doi.org/10.1109/ACCESS.2020.2986809>
- [19] M. Nagy and R. Molontay, "Interpretable dropout prediction: Towards XAI-based personalized intervention," *Int. J. Artif. Intell. Educ.*, vol. 34, no. 2, pp. 274–300, 2024, <https://doi.org/10.1007/s40593-023-00331-8>
- [20] S. Mustofa, Y. R. Emon, S. B. Mamun, S. A. Akhy, and M. T. Ahad, "A novel AI-driven model for student dropout risk analysis with explainable AI insights," *Comput. Educ.: Artif. Intell.*, vol. 8, Art. no. 100352, 2025, <https://doi.org/10.1016/j.caeai.2024.100352>
- [21] S. Boujmiraz, H. Darhmaoui, and A. Drissi El Maliani, "Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches," *Comput. Educ.: Artif. Intell.*, vol. 10, Art. No. 100548, 2026, <https://doi.org/10.1016/j.caeai.2026.100548>
- [22] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015, <https://doi.org/10.48550/arXiv.1503.02531>
- [23] J. Gou, B. Yu, S. J. Maybank, and D. Tao, "Knowledge distillation: A survey," *Int. J. Comput. Vis.*, vol. 129, no. 6, pp. 1789–1819, 2021, <https://doi.org/10.1007/s11263-021-01453-z>
- [24] M. Zhu and S. Gupta, "To prune, or not to prune: Exploring the efficacy of pruning for model compression," *arXiv preprint arXiv:1710.01878*, 2017, <https://doi.org/10.48550/arXiv.1710.01878>
- [25] H. Wu, P. Judd, X. Zhang, M. Isaev, and P. Micikevicius, "Integer quantization for deep learning inference: Principles and empirical evaluation," *arXiv preprint arXiv:2004.09602*, 2020, <https://doi.org/10.48550/arXiv.2004.09602>
- [26] G. Menghani, "Efficient deep learning: A survey on making deep learning models smaller, faster, and better," *ACM Comput. Surv.*, vol. 55, no. 12, Art. no. 259, pp. 1–37, 2023, <https://doi.org/10.1145/3578938>
- [27] H. Q. Nguyen, T. T. Pham, V. B. Vo, and T. T. Quan, "The predictive modeling for learning student results based on sequential rules," *Int. J. Innov. Comput. Inf. Control*, vol. 14, no. 6, pp. 2129–2140, 2018, <https://doi.org/10.24507/ijicic.14.06.2129>
- [28] T. Cao, T. M. Cao, H. Huynh, T. Nguyen, and H. Tran, "Building digital transformation solutions for higher education institutions in Ho Chi Minh City," *DS J. Digit. Sci. Technol.*, vol. 3, no. 1, pp. 11–19, 2024, <https://doi.org/10.59232/DST-V3I1P102>
- [29] L. M. Triet, L. T. H. Hien, and T. S. Hai, "Developing software technology application solutions in teaching web application development to enhance students' learning interest," *Ho Chi Minh City Univ. Educ. J. Sci.*, vol. 20, no. 11, Art. No. 1967, 2023, <https://doi.org/10.54607/hcmue.js.20.11.3682%282023%29>
- [30] R. Baker and P. Inventado, "Educational data mining and learning analytics: Potentials and possibilities for online education," in *Emergence and Innovation in Digital Learning: Foundations and Applications*, G. Veletsianos, Ed. Athabasca, Canada: Athabasca Univ. Press, pp. 81–98, 2016, Available from: https://www.degruyterbrill.com/document/doi/10.15215/aupress/9781771991490.008/html?srsId=AU7gw4UpFBKpLqstaknpTYtuWqpr_skwphE-Fs33hBTWTVbWqlsdlrz
- [31] V. Tinto, "Learning better together," in *Transitioning Students into Higher Education*. London, U.K.: Routledge, pp. 13–24, 2019, <https://doi.org/10.4324/9780429279355-3>