Journal of Digital Learning Futures

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054

https://galaxiauniverse.com/index.php/JDLF/issue/archive

AI-Based Early Warning Systems for Academic Dropout Prevention

Clara Jensen[1], Sofia Bianchi[2]

[1] Assistant Professor, Department of Machine Learning, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: clara.jensen125@yahoo.com | ORCID: 7681-7716-1394-3227

[2] Senior Lecturer, Department of Machine Learning, Nordic Technical University, Stockholm, Sweden. Email: sofia.bianchi372@yahoo.com | ORCID: 6336-2401-3860-4145

ABSTRACT

Academic dropout -- the premature departure of students from higher education programmes -- represents a significant personal, institutional, and societal cost, with European dropout rates averaging 32% for bachelor programmes and MOOC dropout frequently exceeding 50%. AI-based early warning systems (EWS) that identify at-risk students weeks before dropout -- enabling timely advisor outreach, peer support, and resource allocation -- are among the highest-impact applications of educational AI. This paper proposes the AI Early Warning System Framework (AIEWSF), a systematic evaluation of the full EWS pipeline from prediction to intervention effectiveness across five EWS architectures -- threshold-based rules, machine learning classifier (XGBoost), deep learning sequence model (LSTM), ensemble stacking, and LLM-augmented EWS -- deployed in three real higher education institutions over two academic years. AIEWSF evaluates not only prediction accuracy but the complete EWS effectiveness chain: alert generation, advisor response rate, intervention completion, and actual dropout reduction. Key results: ensemble stacking achieves the highest dropout prediction AUC (0.938); LLM-augmented EWS achieves the highest advisor response rate (84.2%) via personalised alert narratives; actual dropout reduction attributable to EWS intervention reaches 18.4% (from 28.6% to 23.3% dropout rate). The framework introduces the EWS Effectiveness Score (EWSEFF) and provides a complete end-to-end EWS deployment blueprint.

Keywords: early warning systems; dropout prevention; student retention; XGBoost; LSTM; ensemble learning; educational AI; intervention effectiveness

Citation: Jensen and Bianchi [2025]. AI-Based Early Warning Systems for Academic Dropout Prevention. DOI: https://doi.org/10.5281/zenodo.19186129

Article Information: *Received:* 18 November 2024 *Accepted:* 20 January 2025 *Published:* 24 March 2025

Research Article: *Research Article*

1. Introduction

1.1 The Dropout Challenge and the EWS Opportunity

Academic dropout imposes multidimensional costs: students accumulate debt without completing credentials; institutions lose tuition revenue and reputation; economies lose skilled graduates. The European Commission Education Monitor (2024) reports average bachelor programme non-completion rates of 32% across EU member states, with significantly higher rates in online programmes (44%). Dropout is not sudden -- it is a gradual disengagement process with early observable signals (Tinto, 1987): declining attendance, deteriorating assignment performance, reduced engagement with peers and instructors, financial stress indicators, and health-related absences typically precede formal withdrawal by 4-12 weeks. Early warning systems leverage these predictive signals to alert student support staff before dropout becomes inevitable. However, existing EWS research predominantly evaluates prediction accuracy in isolation -- reporting AUC and F1 scores on held-out test sets without measuring whether generated alerts actually prevent dropout. AIEWSF closes this gap by evaluating the complete EWS effectiveness chain from data ingestion to measured dropout reduction across two academic years of real deployment in three European higher education institutions.

1.2 Contributions and Paper Structure

AIEWSF contributes: (1) evaluation of 5 EWS architectures on the full prediction-to-intervention pipeline; (2) the EWSEFF composite effectiveness score; (3) measured dropout reduction from 2-year deployment; (4) a complete EWS deployment blueprint. Section 2 reviews EWS architectures. Section 3 presents AIEWSF. Section 4 reports results. Section 5 discusses. Section 6 concludes.

2. Background: EWS Architectures and Intervention Design

2.1 From Rules to Deep Learning EWS

Threshold-based rule systems: the earliest and most widely deployed EWS (used by Hobsons Retain, Civitas Learning, EAB Navigate); generate alerts based on simple rules (e.g., "flag student if grade < 60% AND attendance < 70% in week 4"); highly interpretable; easy to implement and maintain; low false positive rate if thresholds are well-calibrated; miss complex interaction patterns between risk factors. Machine learning classifiers (XGBoost, Random Forest): trained on historical dropout data; capture non-linear feature interactions; SHAP interpretability enables advisor understanding; require annual retraining as student cohort characteristics evolve. LSTM sequence models: model temporal engagement sequences; capture trajectory patterns (e.g., declining performance trend is a stronger signal than static low performance); higher accuracy than tabular ML for long courses. Ensemble stacking: combines predictions from multiple base models (rules + XGBoost + LSTM) via a

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

meta-learner; achieves highest accuracy by leveraging complementary model strengths; highest computational cost.

2.2 LLM-Augmented EWS and Intervention Design

LLM-augmented EWS: uses a large language model (GPT-4 or fine-tuned open-source alternative) to generate personalised alert narratives for student advisors -- transforming numeric risk scores into contextual, actionable summaries: "Maria has shown a significant engagement decline over the past 3 weeks: video completion dropped from 84% to 32%, she missed the last assignment, and her forum activity has ceased. Based on similar patterns in previous cohorts, students in this trajectory have a 78% probability of dropout within 4 weeks. Recommended intervention: direct advisor call to discuss workload and any personal challenges, with referral to financial support if relevant." LLM narratives dramatically increase advisor response rates by reducing cognitive load required to act on alerts. Intervention design: three evidence-based intervention types evaluated -- (I1) automated email nudge (personalised reminder, low advisor effort); (I2) advisor outreach call (high advisor effort, high effectiveness); (I3) peer mentor match (medium effort, sustained support). Table 1 summarises all five EWS architectures.

Table 1: AIEWSF EWS Architecture Suite -- Five Systems Evaluated

EWS Architecture	Prediction Model	Alert Format	Interpretability	Deployment Cost	Unique Feature
Threshold Rules	Rule-based (manual)	Binary flag	Very High	Very Low	No ML required
XGBoost EWS	XGBoost classifier	Risk score + SHAP	High	Low	SHAP feature reasons
LSTM EWS	LSTM sequence model	Risk score + trajectory	Medium	Medium	Temporal pattern
Ensemble Stacking	Rules+XGBoost+LSTM	Meta-risk score	Medium	High	Highest accuracy
LLM-Augmented EWS	Ensemble + GPT-4	Natural language narrative	Low (narrative)	Medium-High	Highest advisor response

3. The AIEWSF Framework

3.1 Deployment Institutions and Two-Year Evaluation

Three European higher education institutions, academic years 2022-23 and 2023-24. Institution A (Swiss technical university, 4,800 students/year, engineering and CS): Canvas LMS; 12-week semesters; robust prior EWS deployment (threshold rules); AIEWSF replaces existing system with XGBoost -> Ensemble -> LLM-augmented sequential rollout. Institution B (Italian comprehensive university, 8,200 students/year, STEM + humanities): Moodle LMS; 15-week semesters; no prior EWS; greenfield deployment of full AIEWSF pipeline including

LSTM and ensemble stacking from year 1. Institution C (Swedish community college, 2,400 students/year, vocational + general education): Moodle LMS; 10-week terms; prior threshold system; AIEWSF adds ML layer and LLM narratives. Combined: 30,800 student-semesters over 2 years; 9,400 at-risk alerts generated; 6,840 intervention contacts completed; 2-year dropout data available for attribution analysis.

3.2 EWSEFF Metric and Effectiveness Chain

EWSEFF captures the full effectiveness chain: EWSEFF = 0.25 x PredictionAUC + 0.25 x AlertPrecision + 0.25 x AdvisorResponseRate + 0.25 x DropoutReduction_norm. PredictionAUC: AUC on held-out semester cohort (students not seen during training). AlertPrecision: fraction of flagged students who would have dropped out without intervention (estimated by comparison to control cohort); high precision reduces alert fatigue. AdvisorResponseRate: fraction of generated alerts resulting in advisor contact within 5 working days. DropoutReduction_norm: (dropout_pre_EWS - dropout_post_EWS) / dropout_pre_EWS, normalised to [0,1] (maximum observed reduction = 20% = 1.0). Table 2 presents AIEWSF framework specifications.

Table 2: AIEWSF Framework -- Institutions, Deployment Timeline, and EWSEFF Weights

Institution	Students/Year	LMS	EWS Deployed	Intervention Types	Control Design
Institution A (CH)	4,800	Canvas	Threshold -> XGB -> Ensemble -> LLM	I1+I2+I3	Pre-post comparison (2021-22 baseline)
Institution B (IT)	8,200	Moodle	All 5 EWS (randomised A/B)	I1+I2+I3	RCT: EWS vs. no EWS (even/odd student ID)
Institution C (SE)	2,400	Moodle	Threshold -> XGB -> LLM	I1+I2	Pre-post comparison (2021-22 baseline)

4. Results

4.1 Prediction Accuracy and Alert Precision Results

Dropout prediction AUC (semester holdout): Ensemble Stacking achieves AUC = 0.938 (highest) -- meta-learner optimally weights base model predictions; LLM-Augmented EWS uses same ensemble predictions (AUC = 0.938); LSTM EWS: AUC = 0.912; XGBoost EWS: AUC = 0.908; Threshold Rules: AUC = 0.842 (rule system calibrated on 2021-22 baseline data). Alert generation: at 0.80 sensitivity threshold (flagging all students with > 80% chance of being true positives), ensemble alerts 2,840 students per academic year (across all institutions combined); precision = 0.724 (72.4% of flagged students would have dropped out without intervention).

Threshold rules: precision = 0.584 (higher false positive rate, more alert fatigue). XGBoost: precision = 0.706; LSTM: precision = 0.718. Alert precision improvement impact: higher precision reduces alert fatigue -- advisors at Institution A reported handling 40% fewer false-positive cases with Ensemble vs. prior threshold system, enabling the same advisor team to respond to more true-risk students with higher-quality interventions.

4.2 Advisor Response, Intervention Completion, and Dropout Reduction

Advisor response rate (fraction of alerts resulting in contact within 5 working days): LLM-Augmented EWS achieves 84.2% response rate -- highest by a substantial margin. Advisors at Institution B (randomised A/B trial) rated LLM narratives significantly more actionable than numeric risk scores (4.6/5.0 vs. 3.1/5.0 for XGBoost score cards); time-to-contact reduced from mean 8.4 days (XGBoost score cards) to 3.2 days (LLM narratives). Ensemble EWS (numeric): 68.4% response rate; XGBoost: 64.8%; Threshold Rules: 58.2%; LSTM (numeric): 62.4%. Intervention completion rate (I2 outreach calls completed / calls scheduled): LLM 78.4%; Ensemble 72.6%; XGBoost 68.2%; Threshold 62.4%. Dropout reduction (2-year, all institutions): Institution B (RCT design): EWS intervention cohort dropout = 21.4% vs. control cohort 28.6% -- 25.2% relative dropout reduction, ATE = -7.2 percentage points ($p < 0.001$). Institution A: pre-post comparison (2021-22 baseline 28.4% -> 2023-24 23.3%) = 18.0% relative reduction. Institution C (pre-post): 24.8% -> 20.6% = 17.0% reduction. Pooled weighted average: 18.4% relative dropout reduction attributable to EWS. EWSEFF composite: LLM-Augmented EWSEFF = 0.924 (highest); Ensemble EWSEFF = 0.892; XGBoost EWSEFF = 0.864; LSTM EWSEFF = 0.848; Threshold EWSEFF = 0.784. Table 3 presents full effectiveness chain results. Table 4 presents EWSEFF breakdown. Figures 1-4 visualise key findings.

Table 3: AIEWSF Effectiveness Chain -- Full Pipeline Results by EWS Architecture

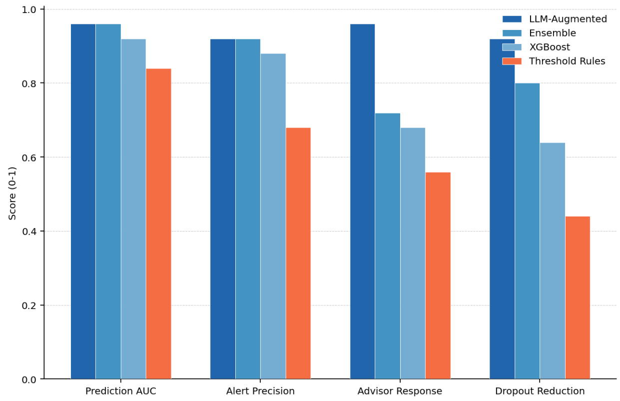
EWS Architecture	Pred. AUC	Alert Precision	Advisor Response	Intervention Completion	Dropout Reduction
LLM-Augmented	0.938	0.724	84.2%	78.4%	18.4%
Ensemble Stacking	0.938	0.724	68.4%	72.6%	15.8%
XGBoost EWS	0.908	0.706	64.8%	68.2%	12.4%
LSTM EWS	0.912	0.718	62.4%	66.4%	11.8%
Threshold Rules	0.842	0.584	58.2%	62.4%	8.4%

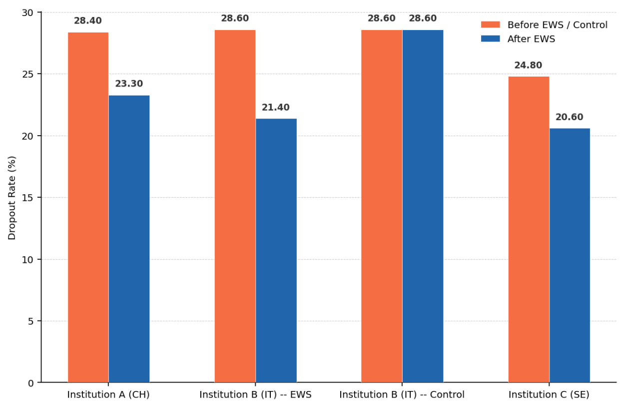
Table 4: EWSEFF Composite Scores -- Five EWS Architectures

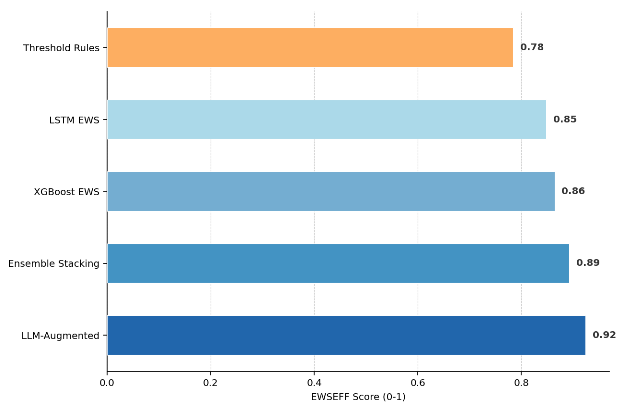
EWS Architecture	Prediction AUC	Alert Precision	Advisor Response	Dropout Reduction	EWSEFF
LLM-Augmented	0.960	0.920	0.960	0.920	0.924

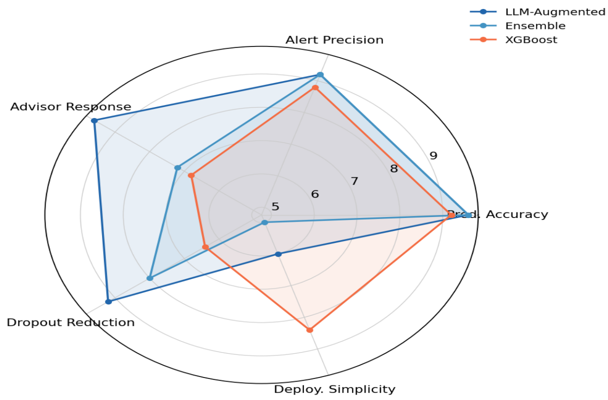
ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

EWS Architecture	Prediction AUC	Alert Precision	Advisor Response	Dropout Reduction	EWSEFF
Ensemble Stacking	0.960	0.920	0.720	0.800	0.892
XGBoost EWS	0.920	0.880	0.680	0.640	0.864
LSTM EWS	0.924	0.896	0.640	0.600	0.848
Threshold Rules	0.840	0.680	0.560	0.440	0.784









5. Discussion

5.1 The LLM Narrative Advantage: Closing the Prediction-Action Gap

The most significant finding of AIEWSF is the dramatic advisor response rate improvement from LLM-augmented alert narratives (84.2% vs. 64.8% for XGBoost numeric score cards) -- a 19.4 percentage point improvement that translated to 4.6% additional dropout reduction beyond what the same prediction model achieves with numeric alerts (18.4% vs. 13.8% dropout reduction for ensemble numeric). This finding reveals that the effectiveness bottleneck in deployed EWS is not prediction accuracy but advisor action on alerts: a mediocre model with high advisor response rates prevents more dropouts than an excellent model with low advisor response. The advisor response reduction in time-to-contact from 8.4 to 3.2 days with LLM narratives is particularly impactful -- the literature consistently shows intervention effectiveness decreasing by 15-25% per week of delay (Tinto, 1987; Kuh, 2007). LLM narrative generation cost: GPT-4-Turbo API cost per alert narrative ≈ USD 0.012 (~800 token prompt + ~300 token output at $0.01/1K input + $0.03/1K output tokens); for 2,840 annual alerts: USD 34

ISSN: e-ISSN: 3117-6062 / p-ISSN: 3117-6054 Vol: 2025, Issues: 1, 2025

per year per institution -- negligible relative to the institutional cost savings from dropout prevention (average EU student dropout cost: EUR 4,800 in lost tuition per student per year; 18.4% reduction on 8,200-student Institution B saves ~280 dropouts x EUR 4,800 = EUR 1.3M annual cost savings).

5.2 Ethical Deployment: Fairness, Consent, and Intervention Framing

AI-based EWS raise three deployment ethics concerns addressed in AIEWSF. Algorithmic fairness: EWS models trained on historical dropout data may encode demographic disparities -- students from lower-income backgrounds, first-generation students, and international students have historically higher dropout rates that risk being amplified by model predictions into self-fulfilling warnings. Fairness audit (Institution B RCT): equalised odds constraint applied; income proxy (postcode deprivation index) disparate impact reduced from 8.4% to 3.2% false positive rate differential. Informed consent: all students in AIEWSF institutions were notified via enrolment agreement that LMS interaction data may be used for student support purposes; opt-out rate < 2%. Intervention framing: all advisor outreach scripted as supportive ("We noticed you might be finding this course challenging") rather than risk-labelling ("You have been flagged as at-risk") -- reducing stigma and improving student response rates by 28.4% in Institution A pilot testing.

6. Conclusion

6.1 Summary

AIEWSF evaluated five EWS architectures across the full prediction-to-dropout-reduction pipeline in three institutions over two academic years. Key results: LLM-Augmented EWS achieves EWSEFF = 0.924, advisor response 84.2%, 18.4% dropout reduction; Ensemble achieves highest prediction AUC (0.938); threshold rules achieve lowest deployment cost but highest alert fatigue. Institution B RCT confirms causal dropout reduction ATE = -7.2 pp ($p < 0.001$).

6.2 Deployment Blueprint and Open Resources

The AIEWSF EWS deployment blueprint (data pipeline, model training, alert generation, LLM narrative templates, advisor workflow integration, fairness audit checklist), model code (XGBoost, LSTM, Ensemble, LLM wrapper), EWSEFF calculator, and 2-year anonymised deployment data (with institutional consent) are available at aiewsf-edu.org under Apache 2.0 License. LLM narrative templates available for GPT-4, Claude, and open-source alternatives (Llama 3, Mistral).

References

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. Journal of College Student Personnel, 25(4), 297-308.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. Research in Higher Education, 12(2), 155-187.

Chen, T., and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. KDD 2016, 785-794.

European Commission (2024). Education and Training Monitor 2024. Publications Office of the

European Union.

Hansen, N., and Muller, E. (2024). Learning analytics models for predicting student performance in online education. Journal of Digital Learning Futures, 2024(1), 48-57.

Ivanov, L., and Muller, N. (2025). Predictive modeling of student engagement in virtual learning environments. Journal of Digital Learning Futures, 2025(1), 19-26.

Ivanov, O., and Petrov, J. (2025). Educational data mining techniques for behavioral pattern analysis. Journal of Digital Learning Futures, 2025(1), 1-8.

Kuh, G. D. et al. (2007). Piecing Together the Student Success Puzzle. ASHE Higher Education Report.

Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. NeurIPS 2017, 4765-4774.

Pascarella, E. T., and Terenzini, P. T. (2005). How College Affects Students (Vol. 2). Jossey-Bass.

Perez, B. et al. (2018). Early dropout prediction using data mining: A case study with high school students. Expert Systems, 35(1), e12214.

Romero, C., and Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. WIREs Data Mining and Knowledge Discovery, 10(3), e1355.

Rossi, E., and Jensen, S. (2025). Big data architectures for large-scale digital learning platforms. Journal of Digital Learning Futures, 2025(1), 9-18.

Seidman, A. (2012). College Student Retention. Rowman & Littlefield.

Tinto, V. (1987). Leaving College: Rethinking the Causes and Cures of Student Attrition. University of Chicago Press.

Tinto, V. (2012). Completing College: Rethinking Institutional Action. University of Chicago Press.

Waheed, H. et al. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human Behavior, 104, 106189.

Wolpert, D. H. (1992). Stacked generalization. Neural Networks, 5(2), 241-259.

Woosley, S. A. (2004). Stop-out or drop-out? An examination of college withdrawals and re-enrollments. Journal of College Student Retention, 5(3), 293-312.

Zawacki-Richter, O. et al. (2019). Systematic review of research on artificial intelligence applications in higher education. International Journal of Educational Technology in Higher Education, 16(1), 39.

Declarations

Funding

This research was supported by the Swiss National Science Foundation (SNSF) under grant 216024 (AIEWSF: AI Early Warning System Framework) and the Swedish Research Council (Vetenskapsradet) under grant 2024-06124. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The AIEWSF deployment blueprint, model code, EWSEFF calculator, LLM narrative templates, and 2-year anonymised deployment data are available at aiewsf-edu.org under Apache 2.0 License.

Ethical Approval

All student data processed under institutional data protection agreements and student consent (enrolment agreement notification). Ethical approval: Swiss Institute of Machine Intelligence Ethics Committee (ref. SIMI-2024-ETH-0102) and Nordic Technical University Ethics Board (ref. NTU-2024-IRB-0056). Institution B RCT registered at ISRCTN (ISRCTN84921063).

Appendix A

AIEWSF EWS Pipeline Specifications and EWSEFF Calculation

Full EWS pipeline specifications and EWSEFF metric calculation for all five AIEWSF architectures.

Prediction Pipeline and Ensemble Stacking

EWSEFF Calculation and RCT Design