Journal of Information Technology Education: Research

An Official Publication of the Informing Science Institute InformingScience.org

JITEResearch.org

Volume 25, 2026

# STUDENT DROPOUT PREDICTION IN HIGHER EDUCATION: A SYSTEMATIC REVIEW OF MACHINE LEARNING METHODS AND RISK FACTORS

| Gloria L. Lopez-Muñoz * | Universidad del Cauca, Popayán, Colombia | gllopez@unicauca.edu.co |
|---|---|---|
| Carolina Gonzalez Serrano | Universidad del Cauca, Popayán, Colombia | cgonzals@unicauca.edu.co |
| Camilo Sanchez-Ferreira | Universidad del Cauca, Popayán, Colombia | csanchez@unicauca.edu.co |

* Corresponding author

## ABSTRACT

**Aim/Purpose** This study systematically reviews machine learning techniques for predicting undergraduate student dropout in higher education, identifies related risk factors, explores methodological gaps in Machine Learning (ML)-based dropout prediction, and outlines directions for future research and institutional implementation.

**Background** Student dropout remains a persistent global challenge with significant academic and socioeconomic implications. Although many studies report high predictive accuracy for dropout models, there is still limited understanding of how these models operationalize educational theory, incorporate psychosocial and equity-related factors, and translate predictions into empirically validated interventions in real institutional settings.

**Methodology** The review followed Kitchenham's methodology and PRISMA 2020 guidelines, using a PICO-C framework focused on undergraduate higher education. Searches from 2019 to 2024 across eight databases yielded 301 records. After screening using predefined inclusion/exclusion criteria, double review with Cohen's kappa, and a 16-criterion weighted quality assessment (including automated keyword checks), 75 studies met all quality and relevance thresholds.

Accepting Editor Dimitar Grozdanov Christozov │ Received: December 21, 2025 │ Revised: April 15, 2026 │ Accepted: April 25, 2026.

Cite as: Lopez-Muñoz, G. L., Gonzalez Serrano, C., & Sanchez-Ferreira, C. (2026). Student dropout prediction in higher education: A systematic review of machine learning methods and risk factors. *Journal of Information Technology Education: Research, 25,* Article 20. https://doi.org/10.28945/5776

| | |
|---|---|
| Contribution | This systematic review synthesizes 75 quality-assessed studies on ML-based undergraduate dropout prediction, clarifying which algorithms currently dominate the field, which risk factors and theoretical frameworks are most frequently used, and where critical gaps remain in generalizability, equity, explainability, temporal modeling, and intervention validation. It is distinctive for its explicit quality-weighting process, its focus on the actionability gap, and its mapping of emerging trends and outstanding needs in the field. |
| Findings | The systematic review shows rapid growth in dropout prediction research from 2019 to 2024, reflecting increased institutional awareness. Academic performance and engagement indicators are the most used risk factors in prediction models, whereas psychosocial factors (including self-efficacy, sense of belonging, motivation, and resilience) remain significantly underrepresented. Ensemble methods (particularly Random Forest and XGBoost) dominate the algorithm landscape, consistently achieving accuracies of around 86–88%. However, most studies rely on single-institution datasets, do not account for temporal changes, and rarely evaluate the impact or fairness of interventions triggered by predictions. These limitations reduce the generalizability, equity, and practical utility of ML-based predictive models in higher education. |
| Recommendations for Practitioners | Machine learning models should be viewed as decision-support tools, not standalone solutions. By combining predictive risk assessments with academic, financial, and engagement data, institutions can design multi-dimensional retention strategies, implement early-stage monitoring protocols, and establish clear intervention procedures that are evaluated for effectiveness and fairness. |
| Recommendations for Researchers | Researchers should move beyond single-institution, static models toward longitudinal, cross-institutional designs that incorporate temporal dynamics, transfer learning, and domain adaptation. Future studies should integrate psychosocial, motivational, and equity-oriented variables; systematically apply fairness-aware machine learning and explainability (XAI) techniques; and empirically evaluate whether predictive systems improve retention through controlled intervention studies. |
| Impact on Society | By clarifying how machine learning can reliably identify students at risk of dropout and exposing current limitations in fairness and implementation, this review supports the design of more equitable and effective retention policies in higher education. Improved early-warning systems and evidence-based interventions can reduce economic losses, mitigate social inequality, and enhance graduation rates, particularly in regions with historically high attrition. |
| Future Research | Future research should focus on cross-institutional benchmark datasets, longitudinal cohort studies, and models that adapt to diverse institutional and regional contexts. There is a pressing need for studies that integrate psychosocial and structural factors, evaluate fairness across vulnerable groups, apply explainability frameworks as standard practice, and link predictive models to experimentally tested intervention strategies, moving the field from predictive accuracy toward demonstrable educational impact. |

| | |
|---|---|
| Keywords | student dropout, machine learning, higher education, educational data mining, systematic review |

# INTRODUCTION

Student dropout in higher education is a persistent, multi-dimensional concern with profound social, economic, and educational consequences. Although machine learning (ML) methods have demonstrated high predictive accuracy (86-88%), a critical theory-practice gap persists: most models fail to operationalize educational theory, underrepresent equity-related and psychosocial factors, and lack empirical validation that their predictions translate into effective institutional interventions. This disconnect creates a contradiction: technically sophisticated models that explain little and influence less. A systematic reassessment of the current evidence is essential to understand how ML-based dropout prediction can move from isolated research success to responsible, equitable, institutional practice.

The magnitude of student dropout is substantial and widespread. Global evidence documents dropout rates as high as 50% in developing regions and in specific high-demand disciplines. Latin America faces particularly high attrition, with Mexico reporting 38% dropout rates across multiple disciplines (Urbina-Nájera & Méndez-Ortega, 2022), while Colombia and Chile document similarly elevated figures. Engineering and STEM disciplines are affected across multiple national contexts, including Germany, where institutional studies highlight attrition as a significant concern, particularly for first-year students (Berens et al., 2019; Wild., 2023), reflecting structural challenges that strain both institutional and social capacity. The economic implications are severe: a Chilean state university estimated dropout-related losses of USD 23.4 million over four years (Améstica-Rivas et al., 2021), while a U.S. multi-institutional study reported USD 16.5 billion in annual tuition revenue lost to attrition across 1,669 four-year colleges (Raisman, 2013). These figures underscore the urgency and scale of the problem.

Institutional recognition of dropout as a policy priority has grown significantly. Colombia's Ministry of Education has established SPADIES (Sistema Para la Prevención de la Deserción en la Educación Superior) (SPADIES – Estadísticas de deserción, n.d.), exemplifying the shift of dropout prevention from academic research to a governance-level priority. This institutional engagement reflects the growing consensus that prevention requires systematic, data-driven strategies rather than ad hoc interventions.

In response to this crisis, machine learning has emerged as a dominant analytical approach. Research has shifted from traditional classification methods (Decision Trees, Bayesian Networks) that achieve 74–83% accuracy to ensemble techniques (Random Forest, XGBoost) and neural networks, which now dominate the field, achieving higher accuracy. Concurrently, a growing body of work addresses explainability, fairness, and actionability, recognizing that accuracy alone is insufficient for responsible implementation. This methodological maturation suggests both promise and incompleteness.

Despite this progress, three persistent gaps limit the field's ability to address the core challenge. First, a theory-algorithm disconnect exists. Educational persistence theory (Tinto's Integration Model, Bean's Behavioral Model, engagement frameworks) offers rich, evidence-based explanations of dropout mechanisms. However, most ML studies cite these theories post hoc to justify including academic variables like GPA rather than operationalizing theory to generate novel, theoretically motivated features (e.g., social integration indices, academic momentum trajectories), creating a methodology-theory misalignment. Second, equity and psychosocial underrepresentation persist: Current models rely greatly on available administrative metrics (academic performance, test scores, attendance). Psychological factors (self-efficacy, sense of belonging, motivation, resilience), social dimensions, and equity-related variables capturing intersectional vulnerabilities remain marginal (appearing in fewer than 10% of studies). This narrow feature selection risks both prediction bias and model blindness to the mechanisms through which students actually drop out. Third, an actionability gap remains unaddressed. Approximately 71% of reviewed studies stop at prediction without empirically validating whether recommendations improve retention. Practitioners lack evidence that predictive models, when deployed, genuinely enhance institutional decision-making or student outcomes. The gap between research validation and institutional implementation remains largely uncharacterized.

These gaps are increasingly critical at the post-COVID-19 transformation of higher education. Independent reviews document a sharp expansion of online and hybrid learning and a growing reliance on digital platforms and learning analytics tools in higher education, especially after 2020 (Doo et al., 2023; M. A. Khan et al., 2023). The rapid shift toward hybrid and fully online learning has fundamentally altered the data landscape: institutions now generate rich digital traces of student engagement (discussion participation, resource access, platform navigation patterns) alongside traditional academic metrics. Simultaneously, structural inequities in digital access have created new forms of educational vulnerability. This broader shift suggests that models trained exclusively on pre-pandemic, face-to-face data may not capture post-pandemic dynamics or adequately serve evolving student populations. Although the present review does not formally stratify all results by pre-versus post-2020 datasets, its 2019-2024 corpus spans both periods. It therefore reflects this broader transformation only indirectly through temporal and contextual patterns reported in the included studies. A systematic reassessment of evidence from 2019-2024 is both timely and necessary to ensure that dropout prediction research remains relevant, fair, and actionable across transforming educational contexts.

This systematic review is narrowly scoped to enable precise, rigorous analysis according to the PICO-C (Population, Intervention, Comparison, Outcome, Context) framework. Population: undergraduate students in higher education institutions, excluding K-12, vocational training, and postgraduate education. Intervention: machine learning-based predictive models and computational approaches to identify at-risk undergraduate students and determine dropout causes. Comparison: different machine learning algorithms, data sources, theoretical frameworks, and implementation contexts presented in the reviewed literature. Outcome: (1) predictive algorithms employed and their reported performance metrics (accuracy, precision, recall) and comparative effectiveness; (2) main risk factors and features used in prediction models, with emphasis on the frequency and representation of psychosocial and equity-related variables; (3) theoretical frameworks cited in studies and their explicit operationalization in algorithmic design; (4) generalizability evidence across institutional and geographic contexts; (5) actionability and intervention validation approaches; and (6) equity implications and representation of diverse student populations. Context: peer-reviewed research published 2019–2024, global scope, across disciplinary contexts (engineering, social sciences, health, humanities) where available. This review investigates the following research questions, each directly addressing one or more of the three critical gaps and PICO framework components identified above:

**RQ1**: What is the temporal and geographic distribution of ML-based dropout prediction research from 2019 to 2024?
**RQ2**: What theoretical and conceptual frameworks have been used in undergraduate dropout prediction research, and how explicitly do studies operationalize these theories into algorithmic design?
**RQ3**: What are the main risk factors and features used in dropout prediction models?
**RQ4**: What data analysis techniques, predictive models, and algorithmic approaches have been most applied to predict undergraduate student dropout?
**RQ5**: What are the main methodological gaps, limitations, and knowledge gaps identified in the literature?

## LITERATURE REVIEW

### *THEORETICAL FOUNDATIONS*

Educational theories of student persistence have evolved substantially over four decades, from Tinto's (1975, 1997) foundational Student Integration Model to contemporary frameworks that incorporate psychological, organizational, and equity dimensions. Tinto's model emphasizes academic and social integration as central mechanisms of retention, while Bean's (1980) Behavioral Model incorporates psychological commitment and organizational factors. Cabrera et al.'s (1992) integrated framework synthesizes these perspectives, and Expectancy-Value Theory (Eccles & Wigfield, 2002) positions motivation and self-concept as mediators of persistence. In online and hybrid contexts, the Community of Inquiry (CoI) model (Garrison et al., 2000) and Moore's (1997) theory of transactional distance emphasize cognitive presence, social presence, and pedagogical dialogue as mechanisms that promote engagement and retention.

These theoretical frameworks offer rich, actionable insights. They argue that dropout occurs not merely because students lack academic ability but because they fail to integrate academically and socially (Tinto, 1975), lose psychological commitment to the institution (Bean, 1980), experience low self-efficacy or expectancy (Expectancy-Value), or encounter barriers to presence and dialogue (CoI, transactional distance). Critically, none of these theories assumes that GPA and test scores alone determine persistence; each identifies mechanisms (social integration, psychological commitment, motivation, cognitive presence) that go beyond traditional academic metrics.

However, a theory-algorithm disconnect emerges when examining how the 75 reviewed studies operationalize these frameworks. Most studies citing Tinto's model, for example, do so to justify the inclusion of academic performance variables already in their datasets, rather than to develop new, theoretically driven features such as social integration indices or academic momentum trajectories (Villarreal-Torres et al., 2024). In this systematic review, we found that approximately 68% of the studies cite Tinto but do so retrospectively to justify existing academic metrics. In comparison, only about 6% explicitly operationalize theory into algorithmic design through novel, theoretically grounded feature construction. Similarly, while Bean's Behavioral Model is frequently cited as a theoretical justification for variable selection, psychological variables (self-efficacy, motivation, commitment) appear in fewer than 15% of models (Ferrándiz et al., 2022). Theoretical frameworks such as the Community of Inquiry (CoI) model and Moore's theory of transactional distance are largely absent from the machine learning literature on dropout prediction, despite being well-established in online learning contexts.

The consequence is a field populated with technically mature predictive systems that, in theory, are shallow. In a review of 75 empirical studies, models consistently report accuracies of 85% or higher when predicting dropout using administrative proxies such as academic performance, attendance records, and enrollment status. These metrics dominate the field. About 38% of reviewed studies rely primarily on GPA, and close to 24% incorporate attendance patterns, and similar proportions use enrollment-related indicators. Nevertheless, this empirical focus on administrative data masks a fundamental limitation: models rarely operationalize the theoretical mechanisms identified by educational theories as central to persistence. Most instead deploy readily available institutional metrics post hoc, justified by theoretical frameworks. This disconnect produces a paradox – accurate prediction without explanation. Models may classify students as 'at-risk' with high accuracy yet provide educators with no insight into whether interventions should target academic integration, social connection, motivation, or organizational factors. The implications are significant: this theory-algorithm gap limits both model interpretability for practice and the design of theoretically informed interventions.

## COMPUTATIONAL AND DATA-DRIVEN FRAMEWORKS

Alongside educational theories, data-driven computational frameworks have gained prominence. EDM (educational data mining), LA (learning analytics), CRISP-DM (cross-industry standard process for data mining), and ML (machine learning) methodologies emphasize reproducibility, scalability, and predictive accuracy, all of which align with educational theory (H. Singh et al., 2024; H. P. Singh & Alhulail, 2022).

The field has shown substantial algorithmic maturation. Early work employed logistic regression and decision trees, achieving accuracies of 74–83% (Prasanth & Alqahtani, 2023). Contemporary research has shifted toward ensemble methods. Random Forest appears in about 81% of dropout prediction studies, while XGBoost appears in about 45% (Mahawar & Rattan, 2024). These ensemble approaches leverage multiple weak learners to produce robust predictions; a design principle borrowed from statistical learning theory and now standard in educational ML.

Neural network architectures, including LSTMs (long short-term memory networks) and CNNs (convolutional neural networks), achieve 82–86% accuracy and enable temporal analysis of student persistence patterns (Ananthi Claral Mary & Arul Leena Rose, 2023; Z. Khan et al., 2024). This methodological evolution reflects advances in predictive power. Approximately 24% of recent studies employ explainable AI techniques such as SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), methods designed to reveal which features drive individual predictions (Albreiki et al., 2022; Delen et al., 2024; Padmasiri & Kasthuriarachchi, 2024), remaining as a minority practice. The dominant approach still focuses on static classification based on institutional snapshots and remains empirically opaque. For institutional deployment, educators and administrators require both accuracy and explainability; prediction without interpretation limits actionability.

## *EQUITY AND STRUCTURAL GAPS*

Despite high accuracy, a systemic equity problem persists. While educational theory increasingly recognizes that dropout is not merely an individual failure but a response to institutional and systemic barriers, including poverty, discrimination, disability-access gaps, and cultural disconnection, machine learning models remain largely equity-blind.

Among the 75 studies, academic performance measures (GPA, course failures, admission scores) remain the primary predictors in most studies, with behavioral engagement factors (attendance, submission patterns) serving as secondary variables (Borrella et al., 2022). In contrast, psychosocial factors, such as self-efficacy, sense of belonging, motivation, resilience, and engagement, are included in only a few studies and are rarely integrated into predictive models (Grimaldo & Manzanares-Medina, 2023; King et al., 2024).

Variables capturing intersectional vulnerabilities, such as Indigenous or minority status, disability, first-generation student status, and the combined effect of low income and minority identity, appear in fewer than 10% of the reviewed studies (Yu et al., 2021). Theories and frameworks focused on structural disadvantages and cultural assets are nearly absent. This imbalance is structural, not arbitrary. Academic and behavioral variables are routinely available through institutional information systems; equity variables require deliberate, often sensitive data collection or external linkages. Consequently, models optimized for accuracy, using whatever variables are readily available, systematically exclude equity dimensions. The result is predictive accuracy that may coexist with algorithmic bias: a model achieving 86% overall accuracy may systematically under- or overpredict dropout risk for particular demographic subgroups (Vasquez Verdugo et al., 2022).

## *ACTIONABILITY OF PREDICTION MODELS*

The literature increasingly emphasizes the gap between predictive accuracy and institutional usefulness. While high accuracy matters, it is not enough if models cannot be interpreted, trusted, or effectively deployed by educators and administrators. H. Singh et al. (2024) argue that explainability is essential for responsible deployment. Prasanth and Alqahtani (2023) show that early-warning systems need not only accurate risk scores but also actionable signals delivered with enough lead time for intervention.

Additionally, the evidence indicates that approximately 71% of the reviewed studies end after reporting classification metrics (accuracy, precision, recall) or cross-validation results. They do not empirically test whether their models, when deployed in an actual institutional setting, improve retention outcomes (Olaya et al., 2020).

Accordingly, most ML studies prioritize predictive accuracy, measured against held-out test sets, because accuracy is scientifically standardizable and publishable. Implementation effectiveness requires follow-up studies, institutional partnerships, and longer timescales, all of which are less amenable to academic publication cycles. Besides, implementing a model requires addressing organizational change: educators must trust and understand the model, institutions must design interventions aligned with risk profiles, and evaluation requires careful measurement of counterfactuals. These implementation challenges are beyond the technical scope of most studies.

A final structural limitation emerges from the geography and context of the studied populations. The distribution of study settings shows important contextual differences. Research in traditional face-to-face universities primarily relies on administrative records such as GPA, enrollment data, and attendance records. In contrast, research on MOOCs and fully online environments uses detailed behavioral trace data, such as clickstreams, video completion rates, and forum participation, making engagement the primary focus of prediction (Gardner et al., 2023; Zhidkikh et al., 2024). Studies in distance education programs fall somewhere in between. This institutional heterogeneity makes generalization difficult: features that are highly predictive in one setting may be absent or irrelevant in another. A model designed for a traditional face-to-face university might perform poorly on an online-only MOOC platform, and vice versa. Developing context-sensitive, adaptable models remains a main methodological challenge in the field and still needs to be addressed (Gardner et al., 2023; Masood et al., 2025; Zhidkikh et al., 2024).

Approximately 68% of reviewed studies focus on single institutions, and 64% of high-quality studies explicitly acknowledge that their models may not transfer reliably to other educational contexts. This generalizability void is particularly problematic because regions with the highest dropout rates, parts of Latin America, Africa, and Asia, are underrepresented in the research literature. Approximately 67% of studies originate in the Americas and Europe; only 6.7% from the Middle East and Africa, despite higher attrition in these regions. This geographic imbalance means that predictive models, which learn patterns from available data, are biased toward the problems and solutions prevalent in well-studied regions. This systematic review explicitly maps the temporal and geographic distribution of research, identifying where evidence is concentrated and where models are absent.

Taken together, these patterns reinforce a central contradiction in the field: technically sophisticated models that explain little and influence less in actual institutional practice.

## METHODOLOGY

This study conducts a systematic literature review of undergraduate dropout prediction in higher education, following the methodological procedures proposed by Kitchenham (2004) for evidence-based reviews and reported in accordance with the PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Kitchenham, 2004; Page et al., 2021). The Kitchenham protocol provides structured, reproducible procedures for planning, executing, and synthesizing systematic reviews in technology-related research, while PRISMA 2020 ensures transparent reporting of study identification, screening, and inclusion. Together, these frameworks are well established in educational technology research and support rigor and replicability.

The review is organized into planning and execution phases and was designed to ensure transparency, reproducibility, and relevance for educational researchers and institutional stakeholders interested in the responsible adoption of machine learning–based predictive technologies in higher education. The corpus centers on ML-based prediction studies, while a small number of non-ML statistical studies were retained only when they contributed substantially to understanding theoretical frameworks or risk-factor structures relevant to dropout prediction.

### *PLANNING PHASE*

#### Research objective

The main objective of this systematic literature review (SLR) is to synthesize research published between 2019 and 2024 on predicting undergraduate student dropout in higher education institutions, identifying the predictive models and machine learning techniques used, the analytical frameworks and risk factors examined, and the methodological gaps and limitations in this field. The corpus focuses on ML-based prediction studies, while a small number of non-ML statistical studies are retained only when they contribute substantially to understanding theoretical frameworks or risk-factor structures relevant to dropout prediction.

## Research questions and PICO-C framework

To guide protocol development, the following research question was proposed: "How have machine learning methods and related predictive approaches been used to predict undergraduate student dropout and to determine its causes in higher education institutions from 2019 to 2024?" This question was framed using the PICO-C (Population, Intervention, Comparison, Outcome – Context) framework, summarized in Table 1. To answer the central research question, five specific research questions (RQs) were formulated to guide the SLR, following the PICO-C formulation (Table 1). The research questions are positioned here to ensure they are clearly linked to the study purpose before the methodological execution is described (Table 2).

**Table 1. PICO-C-based review context**

| Component | Definition |
|---|---|
| Population | Undergraduate university students in higher education institutions |
| Intervention | Machine learning–based predictive model, related data-driven approaches |
| Comparison | Different predictive models, data sources, and methodological frameworks |
| Outcome | Predictive techniques, conceptual frameworks, risk factors, and methodological gaps associated with student dropout |
| Context | Undergraduate higher education institutions; studies from 2019 to 2024; excludes K–12, vocational, and postgraduate contexts |

**Table 2. Research questions for systematic review**

| | Question | Why? |
|---|---|---|
| **RQ1** | What is the temporal and geographic distribution of studies on undergraduate student dropout prediction published from 2019 to 2024? | To verify the relevance of the topic based on the number and location of investigations on student dropout prediction and its causes. |
| **RQ2** | What theoretical and conceptual frameworks have been used in the study of undergraduate student dropout during this period, and how do they relate to ML-based prediction studies? | To determine whether there are methodological or theoretical frameworks that can be used or adapted within the proposal and to identify the degree of theoretical integration in ML-based studies. |
| **RQ3** | What are the main risk factors associated with undergraduate student dropout identified in recent literature? | To determine the importance of analyzing risk factors and identify trends in the main factors that must be considered when developing dropout prediction solutions. |
| **RQ4** | What data analysis techniques or predictive models have been most applied to predict undergraduate student dropout? | To identify the different methodologies applied in student dropout prediction and assess their comparative performance. |
| **RQ5** | What are the main methodological limitations and knowledge gaps identified in current studies on undergraduate student dropout prediction and its causes? | To identify current challenges and future research directions in this area of study. |

The research questions were designed not only to map existing predictive approaches but also to identify methodological gaps, theoretical limitations, and practical barriers affecting the adoption of machine learning models in real educational contexts.

## *EXECUTION PHASE*

### Search strategy

The search protocol was established with predefined inclusion and exclusion criteria (Table 3) and a selected set of databases and repositories (Table 4). The search string was developed by combining keywords and Boolean operators through iterative refinement:

(“student dropout” OR “academic attrition” OR “academic failure”) AND (“higher education” OR “university” OR “undergraduate”) AND (“predict model*” OR “machine learning” OR “statistical model”) AND (“risk factor*” OR “variable*”) AND (“framework” OR “conceptual model*” OR “education* model*” OR “social* model*”)

**Table 3. Inclusion and exclusion criteria for study selection**

| Inclusion criteria | Exclusion criteria |
|---|---|
| Peer-reviewed empirical studies and theses | Studies without peer review |
| Documents between 2019 and 2024 | Papers before 2019 |
| English/Spanish languages | Studies in other languages |
| Open-access or otherwise available full-text documents | No open access or complete text available |
| Government/Institutional technical reports | Unstructured data sources |
| Studies focusing on undergraduate student dropout | Studies focusing exclusively on K-12 or postgraduate education |

Applying this search string to the selected databases and repositories (Table 4), under the established inclusion and exclusion criteria, yielded a total of 301 records, as summarized in Table 5. Because the SciELO database returned duplicate records already retrieved from other databases, it was not considered separately in the final counts.

**Table 4. Databases and platforms used for the search**

| Category | Databases/platform |
|---|---|
| Education | EBSCO – ERIC – SciELO – Redalyc – Dialnet |
| Technology | IEEE Xplore – ACM Digital Library |
| Multidisciplinary | Scopus – Web of Science – Springer |
| Government/Statistics and repositories | UNESCO, OECD, Colombian Ministry of Education, DSpace, Colombian university repositories |

Educational databases and local/regional repositories were intentionally included to capture empirical studies grounded in education theory and to incorporate relevant literature published in Spanish contexts, which is often underrepresented in purely technical databases. This deliberate inclusion of Spanish-language sources (SciELO, Redalyc, Dialnet) was motivated by the significant volume of dropout prediction research conducted in Latin America, where dropout rates are among the highest globally.

**Table 5. Records retrieved from databases**

| Category | Database | Records | Percentage |
|---|---|---|---|
| Multidisciplinary | Springer | 107 | 54.15% |
| | Scopus | 46 | |
| | Web of Science | 10 | |
| Technology | IEEE Xplore | 34 | 18.60% |
| | ACM | 22 | |
| Education | Redalyc | 47 | 23.92% |
| | Dialnet | 9 | |
| | ERIC | 6 | |
| | EBSCO | 10 | |
| Repositories/ GOV Registers | Government, Institutional, etc. | 10 | 3.00% |
| **Total** | | **301** | |

## Screening and selection

The initial search yielded 301 records. Following deduplication (n = 24), 277 unique records were submitted to title and abstract screening by two independent researchers, who applied the predefined inclusion and exclusion criteria to ensure methodological rigor. Inter-rater reliability (IRR) was assessed using Cohen's kappa, a widely used coefficient for agreement between nominal ratings (Cohen, 1960; De Raadt et al., 2019), with each article classified as:

- No (N): Clearly irrelevant
- Related (R): Potentially relevant
- Yes (Y): Clearly relevant for full-text review
- Articles classified as R or Y were retrieved in full text

The binary classification (Table 6) indicated substantial agreement, with observed agreement of 91.3% (253/277 articles). Distribution of classifications: exclusion agreement 26.0% (72 articles), inclusion with high relevance 57.0% (158 articles), and inclusion as related 8.3% (23 articles). Of the 24 disagreements, 5 articles were excluded following consensus review.

**Table 6. Inter-rater reliability for title and abstract screening**

| Classification | Observed agreement | Cohen's $\kappa$ | 95% CI |
|---|---|---|---|
| (N/R/Y) | 73.3% (203/277) | 0.519 | [0.428, 0.603] |
| (Exclude/Include) | 91.3% (253/277) | 0.748 | [0.637, 0.837] |

The first screening stage (title and abstract) led to the total exclusion of 77 records for the following reasons:

- Not related to the scope of search (n = 18)
- Insufficient focus on student dropout prediction or risk factors (n = 49)
- Exclusively focused on K-12 or postgraduate education (n = 7)
- Outside the temporal scope (n = 3)

This process resulted in 200 articles eligible for full-text assessment. Full-text retrieval was unsuccessful for 20 records; these were excluded after attempts to retrieve them through repeated database access sessions, institutional repository searches, and inter-library loan requests. The comprehensive screening process identified 180 studies meeting all eligibility criteria, which were then included in the quality assessment.

## *QUALITY ASSESSMENT*

A comprehensive quality assessment framework was applied to the 180 studies that passed full-text screening, following the approach proposed by Kitchenham (2004) for systematic literature reviews. In a preliminary stage, purely descriptive articles (reviews without primary research, empirical validation, or a focus on dropout/academic performance prediction, or lacking an undergraduate focus) were eliminated. This preliminary assessment removed 43 articles, leaving 137. These 137 were assessed using 16 quality assessment (QA) criteria (Table 7) grouped into 7 dimensions (D1 – D7).

**Table 7. Quality assessment criteria**

| Criterion | Question |
|---|---|
| QA1 | Does the study relate to the prediction of student dropout in higher education? |
| QA2 | Does the study address features/factors associated with student dropout in higher education? |
| QA3 | Is there explanation of features or factors contributing to dropout prediction? |
| QA4 | Is the research design clearly described (e.g., experimental, longitudinal, descriptive)? |
| QA5 | Does the study follow an established methodological framework? |
| QA6 | Does the study include a comparison stage of models or methods? |
| QA7 | Is the dataset adequately described (source, size, time period)? |
| QA8 | Are limitations on data quality or acquisition mentioned? |
| QA9 | Are predictive algorithms or methods clearly specified? |
| QA10 | Is a validation method employed in the study? |
| QA11 | Are validation metrics reported? |
| QA12 | Are quantitative results clearly presented? |
| QA13 | Are results compared to other studies or baseline models? |
| QA14 | Are practical implications and recommendations for intervention discussed? |
| QA15 | Is there discussion of how findings contribute to theory or practice? |
| QA16 | Are study limitations, gaps, or biases clearly stated or discussed? |

The quality assessment criteria were designed to balance methodological rigor with educational relevance. Criteria QA1–QA13 primarily address technical unassailability, research design clarity, data integrity, predictive methods, and validation practices commonly expected in machine learning-based studies. QA14 and QA15 extend the assessment to consider interpretability, practical implications, and contributions to theory or practice, reflecting the educational technology focus of this review. QA16 ensures transparency by explicitly discussing limitations, gaps, and potential biases.

- **(D1) Relevance to Research Questions** (QA1–QA3): Does the study address undergraduate dropout prediction in higher education? Does it identify dropout factors? Are these factors explained?
- **(D2) Research Design and Methodology** (QA4–QA6): Is the research design clearly described? Does the study follow established methodological frameworks? Does it compare methods or models?
- **(D3) Data Integrity and Description** (QA7–QA8): Is the dataset adequately described (source, size, time period)? Are data quality limitations noted?
- **(D4) Predictive Methods and Validation** (QA9–QA11): Are the algorithms clearly specified? Is a validation method used? Are validation metrics reported?
- **(D5) Results and Metrics** (QA12–QA13): Are the quantitative results clearly presented? Are they compared with other studies or baseline models?
- **(D6) Interpretability and Implications** (QA14–QA15): Are practical implications and intervention recommendations addressed? Do the findings contribute to theory or practice?
- **(D7) Transparency and Limitations** (QA16): Are study limitations, gaps, and biases clearly described?

### Scoring system

For scoring the related articles, each criterion was rated on a three-point scale: Yes (1 point): criterion fully met; Partially (0.5 points): criterion partially met or incompletely addressed; No (0 points): criterion not met or not mentioned. A Python-based text analysis script was applied to each study's full text to ensure reproducible, consistent scoring. Scores were summed to obtain dimension-level scores, which were then converted into a weighted overall quality score. The weighting scheme is summarized in Table 8.

**Table 8. Quality dimension weights**

| Dimension | Criteria | Weight |
|---|---|---|
| D1 – Relevance | QA1–QA3 | 25% |
| D2 – Scope and Factors | QA2–QA3 | 20% |
| D3 – Research Design | QA4–QA6 | 15% |
| D4 – Data Validation | QA7–QA8 | 15% |
| D5 – Predictive Methods | QA9–QA11 | 15% |
| D6 – Results | QA12–QA13 | 5% |
| D7 – Transparency | QA16 | 5% |

Before quality classification, all studies were required to meet three mandatory criteria: QA1 ≥ 0.5 (relevance to undergraduate dropout prediction), QA7 ≥ 0.5 (adequate dataset description), and QA9 ≥ 0.5 (clear specification of predictive methods). Studies failing any mandatory criterion were excluded regardless of overall quality score (n = 24; 17.5%).

### Quality classification threshold

Studies passing the mandatory gates were classified as:

- High quality: ≥ 80%
- Moderate quality: 68 – 79%
- Low quality: 60 – 67%
- Insufficient quality: <60%

These calibrated thresholds were refined through iterative testing to achieve an optimal balance between methodological rigor and comprehensive coverage of the literature, following recommendations for systematic reviews in educational technology research (Kitchenham, 2004; Petticrew & Roberts, 2006).

As summarized in Figure 1, the PRISMA 2020 flow diagram illustrates the screening and selection process that led to the final set of included and excluded records. Studies classified as High, Moderate, or Low Quality were included in the synthesis (n=75, 54.7%), while Insufficient Quality studies were excluded (n=46, 33.6%) and those failing mandatory gates (n=16, 11.7%). Low-quality studies were included to ensure comprehensive landscape representation and trend mapping, but were interpreted with appropriate caution in synthesis, following systematic review best practices (Page et al., 2021; Petticrew & Roberts, 2006). Including low-quality studies allows the review to detect emerging methodological tendencies and underserved topic areas, while clear quality labeling in interpretation prevents their weaker evidence from being overweighted in conclusions.

The resulting 54.7% inclusion rate aligns with recommended ranges for systematic reviews (40–60%), providing sufficient corpus size for robust trend analysis while maintaining rigorous quality standards.

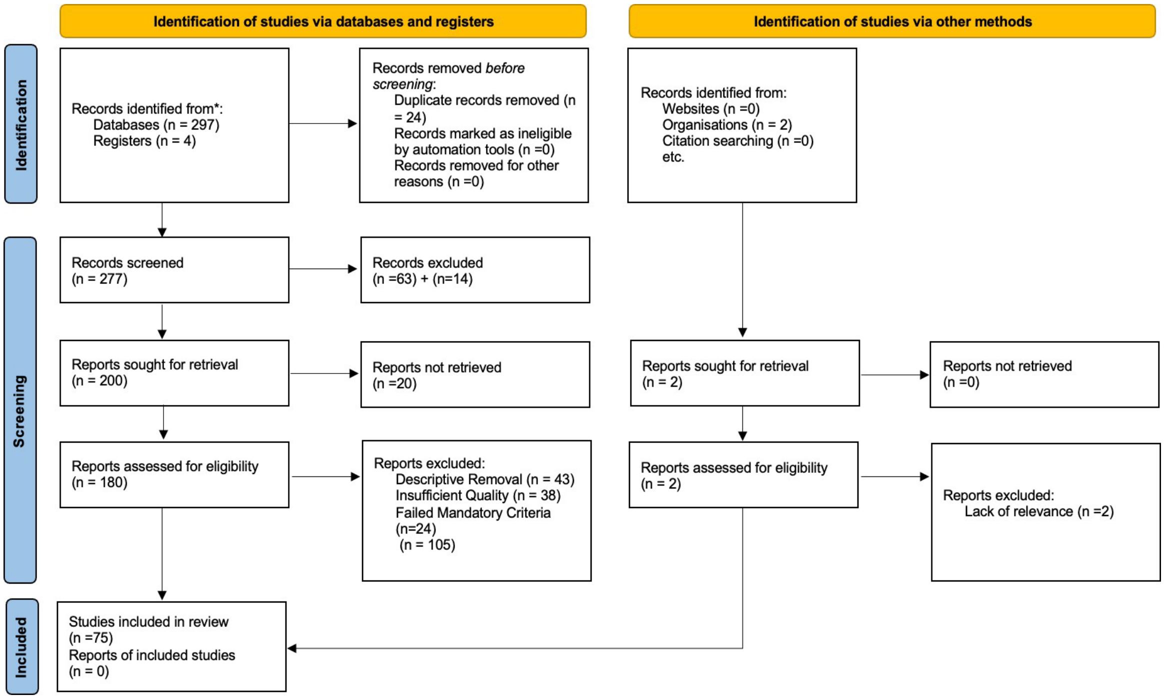

**Figure 1. PRISMA 2020 flow diagram (adapted from Page et al., 2021)**

## Quality assessment results

Of the 137 articles subjected to quality assessment, the distribution across quality categories demonstrated substantial variation, as shown in Table 9.

**Table 9. Quality classification results**

| Quality classification | Studies | Percentage | Mean score | Mean n score (%) | Range (%) |
|---|---|---|---|---|---|
| High Quality | 37 | 27.0% | 14.03 | 87.74% | 80.00 – 97.50 |
| Moderate Quality | 16 | 11.7% | 11.66 | 72.86% | 68.33 – 77.92 |
| Low Quality | 22 | 16.1% | 10.22 | 63.86% | 60.83 – 67.92 |
| Insufficient Quality | 38 | 27.7% | 8.42 | 52.63% | 32.92 – 59.58 |
| Failed Mandatory Criteria | 24 | 17.5% | 7.71 | 48.19% | — |
| **Total** | **137** | **100%** | **9.65** | **60.31%** | **32.92 – 97.50** |

Figure 2 summarizes the mean normalized QA scores by dimension, highlighting weaker performance in research design and transparency than in relevance and predictive methods. This selection process culminated in a final corpus of 75 articles, providing the empirical basis for the comprehensive state-of-the-art analysis presented in the next section.

Overall, the quality assessment reveals a heterogeneous methodological landscape within the reviewed literature. While a substantial subset of studies exhibits moderate to high methodological quality, a considerable proportion presents limitations related to research design clarity, transparency, and the explicit discussion of assumptions and limitations.

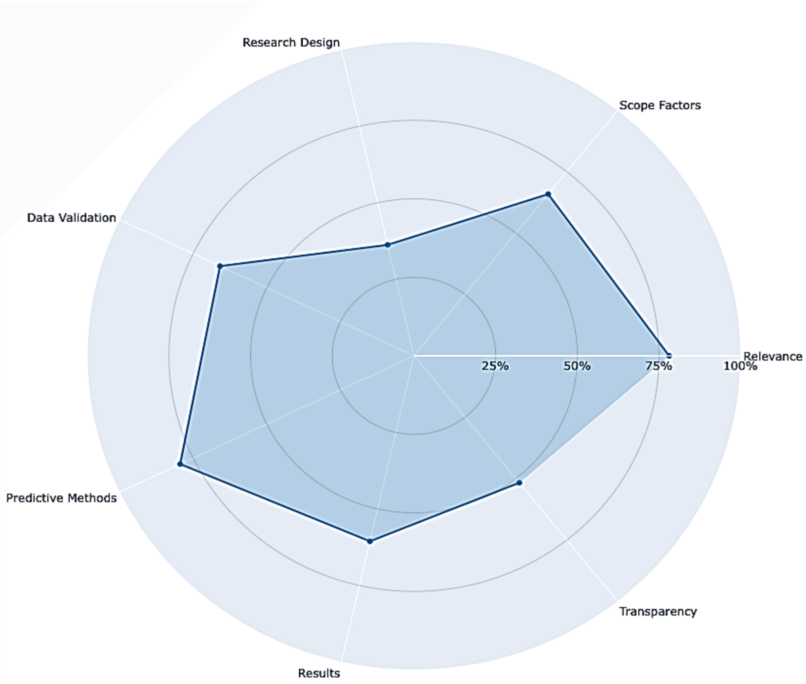

**Figure 2. Mean normalized quality assessment scores by dimension**

Predictive methods and validation practices are generally well developed, whereas weaknesses are more often found in methodological rigor, reproducibility, and cross-context evaluation. These findings indicate that although the field demonstrates technical maturity in predictive modeling, significant opportunities remain to strengthen the design, validation, and transferability of dropout-prediction models across institutional contexts.

# RESULTS AND DISCUSSION

This section presents a critical synthesis of the 75 included studies, evaluating current advancements and methodological gaps in student dropout prediction. The analysis is organized sequentially around the five research questions outlined in Table 2. Visual elements, such as bar charts comparing algorithm categories and their average performance, complement the narrative synthesis presented below and are recommended for the final publication version. Each subsection opens with an explicit link to the corresponding research question.

## *TEMPORAL AND GEOGRAPHIC DISTRIBUTION*

### RQ1: What is the temporal and geographic distribution of studies on undergraduate student dropout prediction published from 2019 to 2024?

The analysis of the 75 included studies reveals distinct patterns regarding temporal evolution, geographic distribution, and institutional context, as shown in Tables 10 and 11.

**Table 10. Temporal distribution of publications**

| Temporal range | Studies | Percentage |
|---|---|---|
| 2019 – 2020 | 18 | 24.0% |
| 2021 – 2022 | 22 | 29.3% |
| 2023 – 2024 | 35 | 46.7% |

**Table 11. Geographical distribution of publications**

| Region | Studies | Percentage | Countries |
|---|---|---|---|
| USA | 12 | 16.0% | USA (12) |
| Latin America | 16 | 21.3% | Colombia (6), México (4), Brasil (3), Perú (2), Ecuador (1) |
| Europe | 22 | 29.3% | Spain (8), Germany (5), UK (4), Finland (3), Other (2) |
| Asia-Pacific | 20 | 26.7% | China (8), India (6), Indonesia (3), Australia (2), Other (1) |
| Middle East/ Africa | 5 | 6.7% | Bahrain (2), Sri Lanka (1), South Africa (1), UAE (1) |

The findings show a strong upward trend in research activity from 2019 to 2024, indicating growing academic and institutional interest in machine learning-based dropout prediction. This increase is particularly evident in 2023–2024, but the present review does not systematically stratify results by pre- vs. post-2020 datasets. Consequently, any potential post-pandemic changes in data sources (for example, greater reliance on platform-log data or online learning contexts) can only be inferred indirectly from the temporal distribution and external evidence, rather than demonstrated conclusively by this corpus.

The geographic distribution of studies, shown in Figure 3, remains highly imbalanced: the Americas and Europe account for roughly two-thirds of all research, while Africa and parts of the Asia-Pacific region remain notably underrepresented despite facing some of the highest dropout rates. This imbalance suggests a systemic lack of context-specific research and raises concerns about the global generalizability of current predictive models. Models developed in dominant regions may inadequately capture contextual, cultural, and institutional factors critical to effective early-warning systems in underrepresented settings.

Furthermore, 58% of the studies are conducted in public institutions, 28% in private institutions, 11% in online or MOOC contexts, and only 3% in mixed institutional models. This distribution indicates that multi-institutional or cross-system analyses remain rare, further constraining the external validity and transferability of the existing evidence base.

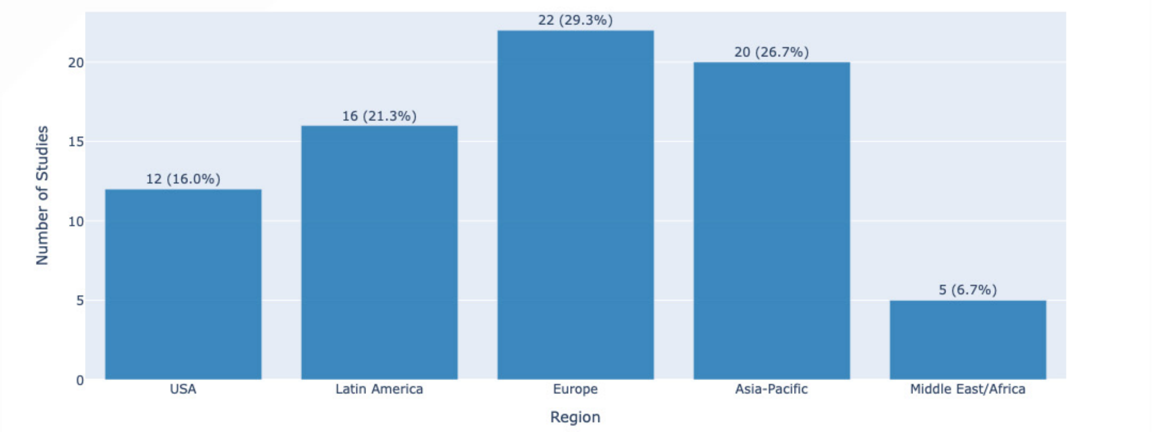

**Figure 3. Geographical distribution of publications**

## *THEORETICAL AND CONCEPTUAL FRAMEWORKS*

### RQ2: What theoretical and conceptual frameworks have been used in the study of undergraduate student dropout during this period?

Regarding theoretical and conceptual frameworks, the field remains largely reliant on traditional persistence theories, particularly Tinto's (1975) Student Integration Model, which appears in 45.3% of studies, emphasizing academic and social integration as key mechanisms for understanding student dropout. Bean's behavioral model, Cabrera's integrated framework, and Expectancy-Value Theory, although less frequently used, also highlight the psychological, motivational, and organizational dimensions of student persistence (Cabrera et al., 1992, 1993).

In parallel, many studies incorporate data-driven frameworks such as Educational Data Mining (EDM) (37.3%), Learning Analytics (LA) (25.3%), CRISP-DM (6.7%), or ML-specific methodologies (30.7%), indicating a shift toward computational modeling and large-scale data analysis, often prioritizing predictive performance over explicit theoretical grounding. However, only 8 studies (10.7%) explicitly combine persistence theories with EDM/LA frameworks, suggesting an emerging but still limited convergence between theoretical and algorithmic perspectives. This restricted integration indicates that many predictive pipelines operate technically effective but conceptually decoupled systems, reducing their explanatory capacity and their potential to contextualize predictive outcomes within educational research. This distribution is summarized in Table 12 and Figure 4.

**Table 12. Theoretical and conceptual frameworks**

| Framework type | Specific framework | Studies | Frequency |
|---|---|---|---|
| Educational persistence | Tinto's Student Integration Model | 34 | 45.3% |
| | Bean's Behavioral Model | 15 | 20.0% |
| | Cabrera et al. (1992, 1993) Integrated Model | 9 | 12.0% |
| Data-driven | Expectancy–Value Theory | 7 | 9.3% |
| | Educational Data Mining (EDM) | 28 | 37.3% |
| | Learning Analytics (LA) | 19 | 25.3% |
| | CRISP-DM | 5 | 6.7% |
| | ML-specific frameworks | 23 | 30.7% |
| Mixed approaches | (persistence theories with data-driven approaches) | 8 | 10.7% |
| Equity-oriented | Explicit equity frameworks | 6 | 8.0% |

It is important to distinguish between studies that genuinely operationalize theories (by embedding theoretical constructs directly into feature engineering and model interpretation) and those that merely cite theories superficially as background context. The majority of studies in this corpus fall into the latter category, citing Tinto's model or Bean's framework without deriving theoretically motivated features or using theory to interpret prediction results. Only a small subset of studies demonstrates genuine theoretical operationalization.

Evidence also indicates that equity-oriented frameworks are severely underrepresented: only 8% of studies employ explicit equity perspectives or disability-inclusion models, while cultural asset frameworks and structural theories of inequality are almost absent. Together with the fact that 26% of studies do not report any explicit theoretical articulation, this reveals a substantial gap in theoretically grounded and equity-focused approaches to dropout prediction, increasing the risk that predictive models reproduce existing structural inequities.

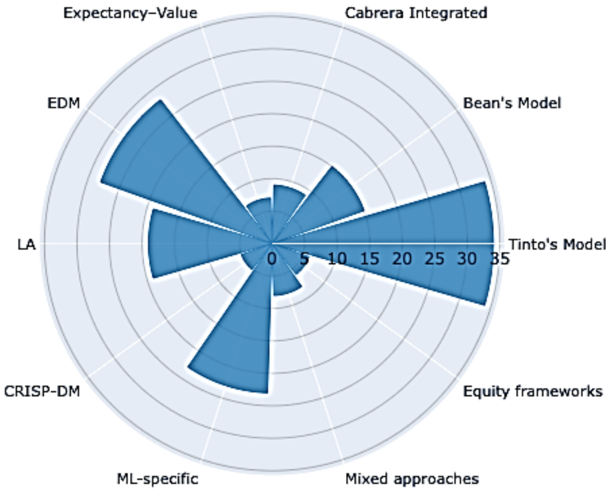

**Figure 4. Theoretical and conceptual frameworks cited in the review**

## *MAIN RISK FACTORS*

### RQ3: What are the main risk factors associated with undergraduate student dropout identified in recent literature?

Analysis of dropout risk factors across the 75 studies reveals a consistent pattern in which academic variables dominate as primary predictors, as shown in Table 13. First-semester GPA, course failures, admission test scores, and credit accumulation appear in more than three-quarters of the studies, reflecting both their strong predictive power and their ready availability through institutional and governmental information systems. Financial constraints, employment while studying, and scholarship status also frequently emerge, underscoring the combined influence of academic performance and socioeconomic pressures on dropout decisions.

**Table 13. Most frequently identified risk factors**

| Risk factor category | Factor | Studies | Frequency |
|---|---|---|---|
| Academic | First-semester GPA | 73 | 97.3% |
| | Course failures/low grades | 72 | 96.0% |
| | Admission test scores | 63 | 84.0% |
| | Credit accumulation rate | 59 | 78.7% |
| Behavioral/ Engagement | Class attendance/absences | 66 | 88.0% |
| | Course engagement | 63 | 84.0% |
| | Assignment submission patterns | 50 | 66.7% |
| Socioeconomic | Financial constraints | 61 | 81.3% |
| | Employment while studying | 55 | 73.3% |
| | Scholarship/aid status | 52 | 69% |
| Demographic | Age at enrollment | 47 | 62.7% |
| | Gender | 46 | 61.3% |
| Institutional | Program fit/major choice | 44 | 58.7% |
| | Initial institution reputation | 32 | 42.7% |
| Equity (Underexplored) | Indigenous/minority status | 6 | 8.0% |
| | Disability status | 4 | 5.3% |

The contextual variations are evident across different types of educational institutions. In traditional universities (51 studies), academic performance metrics are the most influential predictors, whereas in MOOCs and fully online environments (11 studies), engagement indicators become more representative. Similarly, in distance education programs (9 studies), engagement often substitutes for classroom attendance, highlighting that the most informative predictors strongly depend on the institutional context. This contrast between traditional and online/MOOC settings underscores the need for context-sensitive feature selection in dropout prediction research. Beyond the dominance of first-semester GPA and early academic performance as predictors, relatively few studies model dropout risk across the full degree trajectory. Most datasets and models focus on the first year, and nearly 42% of the reviewed studies explicitly lack temporal dynamics, relying on static snapshots rather than longitudinal trajectories. This early-stage focus limits understanding of delayed or cumulative dropout processes and constrains the robustness of causal interpretations.

Despite this wide range of quantitative indicators (Figure 5), psychosocial factors remain markedly underrepresented. Constructs such as self-efficacy, sense of belonging, motivation, resilience, and engagement (in its psychological rather than behavioral sense) appear in only a small fraction of studies and are rarely operationalized as predictive features. Similarly, equity-related factors are severely neglected: only a small minority of studies include variables such as Indigenous or minority status (8%) or disability (5.3%), and only 6% examine intersecting vulnerabilities, for example, the combined effect of first-generation, low-income, and minority status. Disaggregated analyses by demographic subgroups are rare. This narrow focus on performance and basic demographics risks overlooking structural disadvantages linked to geography, ethnicity, or cultural background and may lead to predictive models that reproduce existing inequities rather than reveal them.

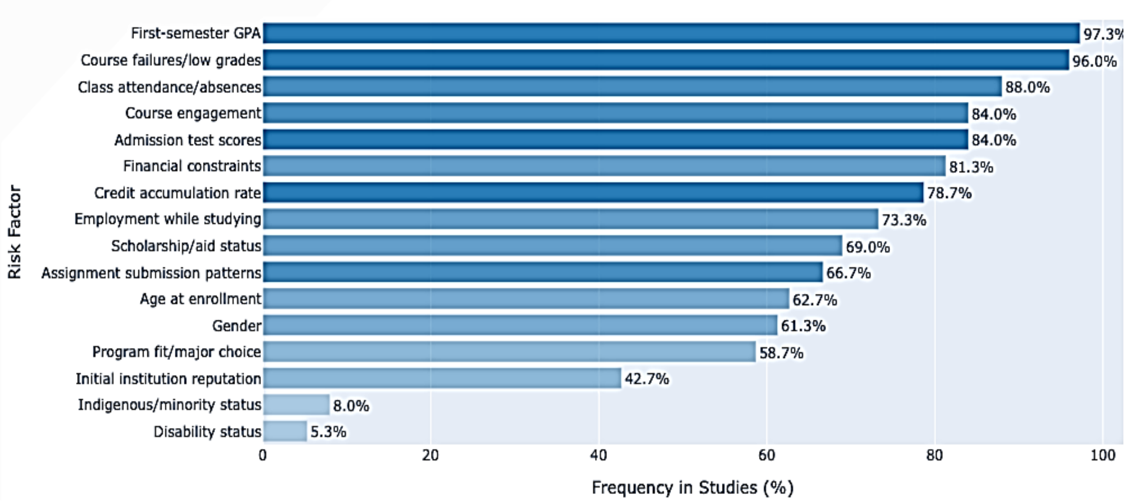

**Figure 5. Frequency of citation of dropout-associated risk factors**

## *PREDICTIVE MODELS AND ANALYTICAL TECHNIQUES*

### RQ4: What data analysis techniques or predictive models have been most applied to predict undergraduate student dropout?

The algorithmic landscape is dominated by ensemble methods, with Random Forest used in 61 studies and XGBoost in 34, typically reporting average accuracy values between 86% and 88%, as presented in Table 14. Traditional baseline algorithms such as logistic regression, decision trees, Naïve Bayes, and SVM remain widely adopted, typically used as reference models for performance comparison and yielding lower average accuracies in the 79–81% range. Neural networks, including feed-forward architectures and deep learning models such as CNNs and LSTMs, appear in 26 studies, achieving average accuracies of 82–86%.

Methodologically, ensemble and hybrid approaches are the predominant strategies for performance optimization in recent studies: stacking ensembles and hybrid models frequently outperform the best individual methods, with average accuracies of approximately 87% or higher. Only a small subset of studies, five in total, has employed advanced causal or time-to-event approaches, such as survival analysis to model dropout (Barragán et al., 2022) and uplift modeling. These methods demonstrate competitive performance and, more importantly, provide richer temporal or intervention-oriented insights than standard methods.

**Table 14. Algorithm distribution**

| Algorithm category | Algorithm | Studies | Frequency | Average accuracy |
|---|---|---|---|---|
| Ensemble Methods | Random Forest | 61 | 81.3% | 86.2% |
| | XGBoost | 34 | 45.3% | 87.4% |
| | Gradient Boosting | 21 | 28.0% | 84.5% |
| | Stacking Ensemble | 16 | 21.3% | 88.1% |
| | Voting Classifier | 8 | 10.7% | 85.3% |
| Traditional Methods | Logistic Regression | 47 | 62.7% | 79.3% |
| | Decision Trees | 43 | 57.3% | 81.4% |
| | Naive Bayes | 18 | 24.0% | 76.2% |
| | SVM | 19 | 25.3% | 80.1% |
| Neural Networks | Neural Networks (general) | 26 | 34.7% | 82.8% |
| | Deep Learning/ CNN | 14 | 18.7% | 85.6% |
| | LSTM/RNN | 7 | 9.3% | 83.4% |
| Advanced Methods | Survival Analysis | 3 | 4.0% | 84.2% |
| | Uplift Modeling | 2 | 2.7% | 81.5% |
| Hybrid Methods | | 12 | 16.0% | 87.3% |

Additionally, approximately 78% of studies rely on k-fold cross-validation (often 10-fold), and nearly 71% explicitly address class imbalance using resampling, class weights, or synthetic oversampling, indicating growing awareness of the need for rigorous validation. A graphical and numerical representation of the revised literature is presented in Figure 6 and Table 14, respectively.

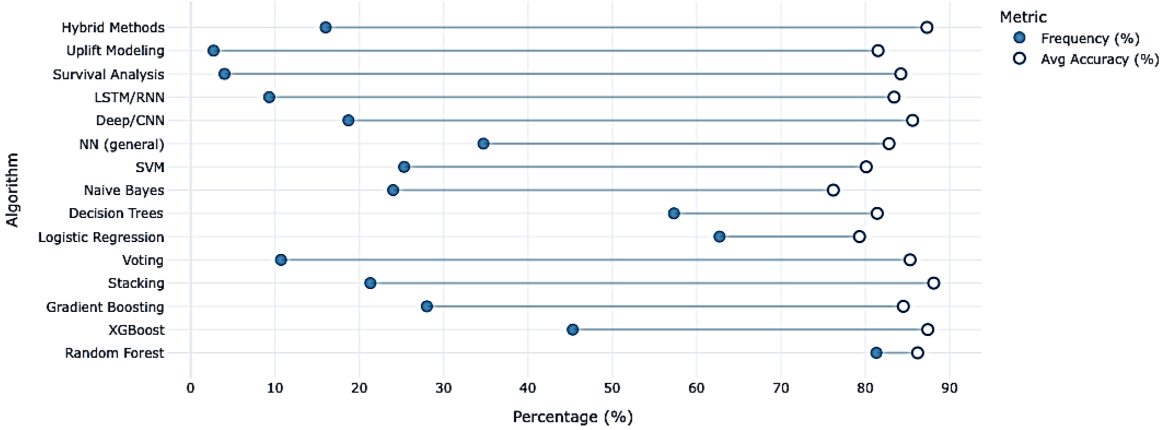

**Figure 6. Frequency vs accuracy by algorithms in the review**

Despite this technical maturation, important gaps remain. First, relatively few models are explicitly designed as adaptive or context-sensitive systems; most architectures are static classifiers trained on specific datasets and then applied without systematic adaptation to new cohorts, institutions, or modalities. Second, although recent approaches show a clear increase in the use of explainable AI (XAI) techniques, the majority of high-performance ensemble and complex models still operate as low-transparency "black boxes," constraining institutional trust. The tension between predictive performance and institutional interpretability is central: a model achieving 88% accuracy but unable to explain *why* a student is at risk provides limited utility to academic advisors who must act on its outputs. XAI tools such as SHAP and LIME are beginning to address this, but remain non-standard. Finally, precision, recall, and related metrics are often reported without systematic analysis of trade-offs; few studies explicitly optimize for the recall of true dropouts versus precision in classifying at-risk students, leaving open questions about how well these models balance early detection with the informational costs of false alarms in institutional decision-making contexts. These findings motivate the review's recommendation that explainability become a standard requirement for ML-based dropout prediction systems intended for deployment.

## *METHODOLOGICAL GAPS AND LIMITATIONS*

### RQ5: What are the main methodological limitations and knowledge gaps identified in current studies on the prediction and determination of causes of undergraduate student dropout?

Analysis of reported limitations across the 75 studies identifies critical barriers that currently constrain the transition from predictive modeling to effective educational practice (Table 15).

The most prevalent structural limitation concerns generalizability and context: 68% of studies rely on single-institutional datasets, and 64% of high-quality studies explicitly acknowledge that their models may not transfer reliably to other educational settings or student populations. This fragmentation suggests that, although local predictive accuracy is often high, the field lacks robust cross-institutional models that operate reliably across diverse socioeconomic and educational environments.

Equally critical is the gap in intervention effectiveness. Although prediction accuracy has improved, 71% of studies lack empirical validation of interventions, indicating limited evidence that identifying at-risk students is systematically linked to improved retention outcomes when models are deployed in practice. This disconnect is compounded by limited real-time dynamism: 42% of models ignore temporal dynamics, focusing on static snapshots rather than evolving longitudinal trajectories, and 68% lack the adaptive capabilities to generate timely insights during the academic term. The implications of this static vs. longitudinal distinction are significant: a first-year GPA snapshot captures a student's current performance. However, it cannot detect accelerating disengagement over time, predict delayed dropout in later years, or support just-in-time intervention. Longitudinal modeling frameworks, by contrast, could enable ongoing recalibration of risk and more responsive early-warning systems, as summarized in Figure 7.

**Table 15. Reported limitations**

| Research dimension | Specific limitation (Gap) | Studies | Frequency |
|---|---|---|---|
| Context & Generalizability (RQ1) | Single institutional focus (lacks cross-context validation) | 51 | 68.0% |
| Theoretical Grounding (RQ2) | Lack of explicit theoretical framework (purely data-driven/atheoretical) | 20 | 26.7% |

| Research dimension | Specific limitation (Gap) | Studies | Frequency |
|---|---|---|---|
| Data Quality & Risk Factors (RQ3) | Absence of equity/structural frameworks (ignore structural inequality theories) | 69 | 92.0% |
| | No demographic disaggregation (blind to subgroup bias) | 55 | 73.3% |
| | Missing psychosocial factors (ignores motivational dimensions) | 41 | 54.7% |
| Methodological Rigor (RQ4) | No temporal dynamics (static snapshots vs. longitudinal) | 32 | 42.7% |
| | Limited baseline comparisons | 35 | 46.7% |
| Practical Impact (RQ5) | No intervention validation (prediction without action) | 53 | 70.7% |

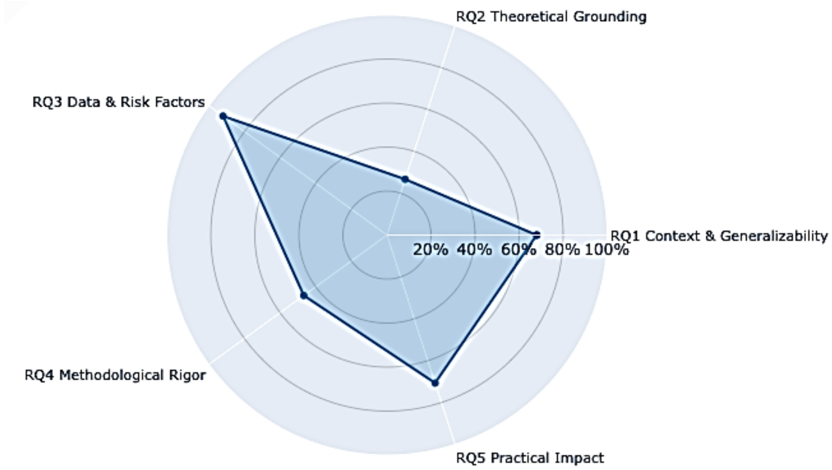

**Figure 7. Distribution of limitations by research dimension**

The comparison between traditional high education settings and online/MOOC environments highlights an additional gap: models trained in brick-and-mortar institutions often lack the engagement and trace-based features central to MOOC dropout prediction, and vice versa. Cross-context validation across these fundamentally different educational modalities remains largely unexplored.

Finally, equity and mechanistic understanding represent systemic blind spots. Disaggregated analysis is missing in 73% of studies, potentially masking performance biases across demographic subgroups. The underuse of subgroup analysis, the sparse inclusion of structural disadvantage variables (socioeconomic deprivation indices, geographic remoteness, first-generation status), and the absence of fairness auditing mean that many deployed models risk reproducing or amplifying existing inequities – misidentifying structurally disadvantaged students as high-risk due to their circumstances rather than their engagement or intent. Furthermore, while predictive performance is often high, 61% of studies report limited insight into the underlying dropout mechanisms, with psychosocial factors remaining underexplored (54%).

**Table 16. Integrative synthesis**

| RQ | Key finding | Supporting evidence | Strategic implication |
|---|---|---|---|
| **RQ1** | Research growth vs. geographic imbalance. Rapid acceleration of ML research, but highly concentrated in Western contexts. | 46.7% of studies published in 2023–2024; Americas and Europe account for ~67%; Africa under-represented. | Future research must prioritize inclusive contexts to ensure models are relevant to regions with the highest dropout rates. |
| **RQ2** | Theoretical disconnection. Persistence theories (Tinto) and data-driven frameworks operate in parallel with limited integration. | 34 studies cite Tinto; only 10.7% explicitly integrate theory with EDM/ML pipelines. | Need for hybrid frameworks that operationalize persistence theory into feature engineering to improve model interpretability. |
| **RQ3** | Predictor narrowness. Models rely heavily on academic performance (GPA), neglecting equity and intersectional factors. | 97% include GPA; only 8% analyze minority status; 42% lack temporal dynamics. | Shift focus from "performance-only" models to holistic, longitudinal profiles that include equity and intersectional risk factors. |
| **RQ4** | Algorithmic convergence vs. transparency. Ensemble methods dominate performance, but "black box" risks remain high despite XAI growth. | RF (81%) and XGBoost (45%) yield the best accuracy (86–88%); only 24% of recent studies use XAI. | Adoption of high-performance ensembles must be paired with systematic XAI (SHAP/LIME) to build institutional trust. |
| **RQ5** | Generalization and implementation gap. Models are rigid, single-institution tools that fail to adapt to new contexts or drive interventions. | 68% are single institutions; 64% admit poor transferability; 71% lack intervention testing. | Research must pivot to domain adaptation and transfer learning to create robust models that work across different institutions and populations. |

Collectively, these gaps demonstrate that contemporary research is technically mature but remains disconnected from equity imperatives and the practical requirements of real-world educational systems.

## CONCLUSIONS

### *CONTRIBUTIONS*

This systematic review of 75 quality-assessed studies (2019–2024) provides robust evidence that student dropout prediction has evolved into a technically mature field, characterized by the widespread adoption of high-performance ensemble methods and growing methodological sophistication. The review makes four distinct contributions to the literature. First, it provides a quality-controlled synthesis of a 75-study corpus that spans eight databases and both English- and Spanish-language sources, offering broader coverage than prior reviews limited to English-only technical databases. Second, it applies a structured 16-criterion quality assessment framework with weighted dimensions, enabling systematic differentiation of strong-, moderate-, and low-quality evidence. Third, it characterizes the actionability gap (the distance between predictive capability and validated institutional intervention) as the field's defining unresolved challenge. Fourth, it identifies the specific psychosocial factors most absent from current models (self-efficacy, sense of belonging, motivation, resilience, engagement constructs) and the equity-related blind spots most likely to produce models that reproduce rather than reveal structural inequities.

## IMPLICATIONS

*For theory:* The review demonstrates that the theory-algorithm disconnection is persistent. The field requires hybrid frameworks that genuinely operationalize persistence theory (Bean, 1980; Cabrera et al., 1992; Tinto, 1975) into feature engineering, hypothesis formulation, and model interpretation, rather than citing these theories superficially as background. Equity-oriented and structural frameworks must be incorporated as first-class analytical perspectives, not afterthoughts.

*For practice:* Machine learning models should be viewed as decision-support tools embedded within institutional retention strategies, not as standalone solutions. Institutions adopting predictive systems should establish clear intervention protocols, evaluate fairness across demographic subgroups before deployment, and require interpretable outputs (via XAI techniques such as SHAP or LIME) that advisors can act upon. The combination of predictive risk assessment with academic, financial, and engagement data enables multi-dimensional retention strategies that are more likely to be equitable and effective.

*For policy:* The geographic concentration of research in high-income Western contexts means that models designed for European and North American universities may be poorly suited to Latin American, African, or Southeast Asian institutions that face distinct socioeconomic and structural challenges. Policy frameworks for adopting predictive systems should require context-specific validation before deployment.

## LIMITATIONS

This review has several limitations that should be acknowledged. First, the corpus is limited to 2019-2024, excluding earlier foundational work. Second, although the corpus includes Spanish-language databases, it may underrepresent literature in other languages (e.g., Portuguese and Chinese), potentially underestimating contributions from Brazil and China. Third, the quality assessment framework, while systematic, involves judgment in partial scoring of criteria and may not fully capture the qualitative dimensions of methodological rigor. Fourth, the review focuses on predictive modeling research and does not systematically evaluate the separate literature on the effectiveness of retention interventions. Fifth, although the 2019–2024 corpus spans both pre-pandemic and post-pandemic periods, the analysis does not systematically stratify findings by pre-2020 vs. post-2020 datasets. As a result, potential shifts in predictors, algorithms, or theoretical frameworks across these periods are only partially captured and should be interpreted with caution.

## FUTURE RESEARCH DIRECTIONS

Future research must pursue a forward-looking agenda across five strategic directions. First, cross-institutional benchmarks: developing shared, multi-institutional datasets and standardized benchmark tasks would enable meaningful performance comparisons and reduce fragmentation caused by single-institution designs. Second, longitudinal modeling: research should shift from static first-year snapshots to models that track student trajectories throughout the degree, enabling detection of delayed and cumulative dropout processes. Third, fairness-aware machine learning: systematic integration of fairness auditing tools, subgroup performance analysis, and equity-oriented feature sets should become standard practice, ensuring that models do not reproduce or amplify structural inequities. Fourth, explainability as standard: XAI techniques (SHAP, LIME, counterfactual explanations) should be required components of any study reporting a predictive model intended for institutional deployment, ensuring that risk scores are interpretable and actionable. Fifth, intervention validation: the field must move from predictive accuracy toward demonstrable educational impact, prioritizing studies that empirically evaluate whether early-warning systems, when coupled with targeted interventions, produce measurable improvements in retention, particularly for vulnerable student populations.

In summary, while the technological foundations for predicting student dropout are well established, the central challenge now is aligning predictive models with the informational, contextual, and ethical requirements of real-world educational systems. Progress in this area will depend less on marginal gains in accuracy and more on developing predictive approaches that are adaptable, interpretable, equitable, and responsive to the diverse conditions under which educational decisions are made.

## FUNDING

This work was supported by the Colombian Ministry of Science, Technology, and Innovation (Minciencias) through a doctoral grant, Ref. Grant: 933 - Minciencias (COL).

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the academic and institutional support provided by the Computational Intelligence Research Group (GICO) at the University of Cauca, where this study was conducted.

## REFERENCES

Albreiki, B., Habuza, T., & Zaki, N. (2022). Framework for automatically suggesting remedial actions to help students at risk based on explainable ML and rule-based models. *International Journal of Educational Technology in Higher Education, 19,* Article 49. https://doi.org/10.1186/s41239-022-00354-6

Améstica-Rivas, L., King-Domínguez, A., Sanhueza Gutiérrez, D., & Ramírez González, V. (2021). Efectos económicos de la deserción en la gestión universitaria: El caso de una universidad pública chilena [Economic effects of dropout in university management: The case of a Chilean public university]. *Hallazgos, 19*, 209–228. https://doi.org/10.15332/2422409X.5772

Ananthi Claral Mary, T., & Arul Leena Rose, P. J. (2023). Ensemble machine learning model for university students' risk prediction and assessment of cognitive learning outcomes. *International Journal of Information & Education Technology, 13*(6), 948-958. https://doi.org/10.18178/ijiet.2023.13.6.1891

Barragán, S., González, L., & Calderón, G. (2022). Modelling student dropout risk using survival analysis and analytic hierarchy process for an undergraduate accounting program. *Interchange, 53*(3-4), 407-427. https://doi.org/10.1007/s10780-022-09463-7

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. https://doi.org/10.1007/BF00976194

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk – Predicting student dropouts using administrative student data from German universities and machine learning methods. *Journal of Educational Data Mining, 11*(3), 1–41. https://doi.org/10.5281/zenodo.3594771

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2022). Taking action to reduce dropout in MOOCs: Tested interventions. *Computers & Education, 179,* 104412179. https://doi.org/10.1016/j.compedu.2021.104412

Cabrera, A. F., Castañeda, M. B., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. *The Journal of Higher Education, 63*(2), 143–164. https://doi.org/10.2307/1982157

Cabrera, A. F., Nora, A., & Castañeda, M. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *The Journal of Higher Education, 64*(2), 123-139. https://doi.org/10.2307/2960026

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational & Psychological Measurement, 20*(1), 37-46. https://doi.org/10.1177/001316446002000104

Delen, D., Davazdahemami, B., & Rasouli Dezfouli, E. (2024). Predicting and mitigating freshmen student attrition: A local-explainable machine learning framework. *Information Systems Frontiers, 26*(2), 641-662. https://doi.org/10.1007/s10796-023-10397-3

De Raadt, A., Warrens, M. J., Bosker, R. J., & Kiers, H. A. L. (2019). Kappa coefficients for missing data. *Educational & Psychological Measurement, 79*(3), 558-576. https://doi.org/10.1177/0013164418823249

Doo, M. Y., Zhu, M., & Bonk, C. J. (2023). A systematic review of the research topics in online learning during COVID-19: Documenting the sudden shift. *Online Learning*, *27*(1), 15–36. https://doi.org/10.24059/olj.v27i1.3405

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Ferrándiz, D. Á., Corona, M. A., Castellón, E. G., & Cruz, M. F. (2022). Neurodidactic factors in the prediction of academic dropout in Andalusian university students: Preventive actions based on ICT. *Texto Livre*, *15*, e40502. https://doi.org/10.35699/1983-3652.2022.40502

Gardner, J., Yu, R., Nguyen, Q., Brooks, C., & Kizilcec, R. (2023). Cross-institutional transfer learning for educational models: Implications for model performance, fairness, and equity. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1664-1684). Association for Computing Machinery. https://doi.org/10.1145/3593013.3594107

Garrison, D. R., Anderson, T., & Archer, W. (2000). Critical inquiry in a text-based environment: Computer conferencing in higher education. *The Internet and Higher Education, 2*(2–3), 87–105. https://doi.org/10.1016/S1096-7516(00)00016-6

Grimaldo, M., & Manzanares-Medina, E. (2023). Variables intervinientes en el rendimiento académico en ingresantes de una universidad privada de Lima [Predictors of academic performance among entering freshmen at a private university in Lima]. *Revista Electrónica Educare, 27(1),* 1-14. https://doi.org/10.15359/ree.27-1.14283

Khan, M. A., Nabi, M. K., Khojah, M., & Tahir, M. (2023). Impact of the COVID 19 pandemic on online learning in higher education. *Frontiers in Education*, *8*, 1225834. https://doi.org/10.3389/feduc.2023.1225834

Khan, Z., Ali, A., Khan, D. M., & Aldahmani, S. (2024). Regularized ensemble learning for prediction and risk factors assessment of students at risk in the post-COVID era. *Scientific Reports, 14,* 16200. https://doi.org/10.1038/s41598-024-66894-1

King, R. B., Luo, Y., & Xie, M. (2024). Good begets good: The role of helping others on engagement and achievement among university students. *Research in Higher Education, 65*(5), 989-1006. https://doi.org/10.1007/s11162-023-09768-1

Kitchenham, B. (2004). *Procedures for undertaking systematic reviews.* (Technical Report TR/SE-0401 & NICTA 0400011T). Department of Computer Science, Keele University. https://www.inf.ufsc.br/~aldo.vw/kitchenham.pdf

Mahawar, K., & Rattan, P. (2024). Empowering education: Harnessing ensemble machine learning approach and ACO-DT classifier for early student academic performance prediction. *Education & Information Technologies*, *30*, 4639-4667. https://doi.org/10.1007/s10639-024-12976-6

Masood, S. W., Gogoi, M., & Begum, S. A. (2025). Optimised SMOTE-based imbalanced learning for student dropout prediction. *Arabian Journal for Science and Engineering, 50,* 7165-7179. https://doi.org/10.1007/s13369-024-09287-w

Moore, M. G. (1997). Theory of transactional distance. En D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). Routledge

Olaya, D., Vásquez, J., Maldonado, S., Miranda, J., & Verbeke, W. (2020). Uplift Modeling for preventing student dropout in higher education. *Decision Support Systems,* 134. 113320. https://doi.org/10.1016/j.dss.2020.113320

Padmasiri, P., & Kasthuriarachchi, S. (2024). Interpretable Prediction of Student Dropout Using Explainable AI Models. *2024 International Research Conference on Smart Computing and Systems Engineering (SCSE), 7*, 1-7. https://doi.org/10.1109/SCSE61872.2024.10550525

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ, 372*, n160. https://doi.org/10.1136/bmj.n160

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide.* Blackwell Publishing. https://doi.org/10.1002/9780470754887

Prasanth, A., & Alqahtani, H. (2023). Predictive Modeling of Student Behavior for Early Dropout Detection in Universities using Machine Learning Techniques. *2023 IEEE 8th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, 1-5. https://doi.org/10.1109/ICETAS59148.2023.10346531

Raisman, N. (2013). The Cost of College Attrition at Four-Year Colleges & Universities. Policy Perspectives. *Educational Policy Institute*. https://eric.ed.gov/?id=ED562625

Singh, H., Kaur, B., Sharma, A., & Singh, A. (2024). Framework for suggesting corrective actions to help students intended at risk of low performance based on experimental study of college students using explainable machine learning model. *Education & Information Technologies, 29(7)*, 7997-8034. https://doi.org/10.1007/s10639-023-12072-1

Singh, H. P., & Alhulail, H. N. (2022). Predicting student-teachers dropout risk and early identification: A four-step logistic regression approach. *IEEE Access, 10*, 6470-6482. https://doi.org/10.1109/ACCESS.2022.3141992

SPADIES – Estadísticas de deserción. (n.d.). *SPADIES: Estadísticas de deserción en la educación superior en Colombia.* Ministerio de Educación Nacional de Colombia. http://www.mineducacion.gov.co/sistemasdeinformacion/1783/w3-propertyvalue-68157.html

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45(1),* 89-125. https://doi.org/10.2307/1170024

Tinto, V. (1997). Classrooms as Communities: Exploring the Educational Character of Student Persistence. *The Journal of Higher Education, 68(6*), 599-623. https://doi.org/10.2307/2959965

Urbina-Nájera, A. B., & Méndez-Ortega, L. A. (2022). Predictive model for taking decision to prevent university dropout. *International Journal of Interactive Multimedia & Artificial Intelligence, 7*(4)*,* 205-213. https://doi.org/10.9781/ijimai.2022.01.006

Vasquez Verdugo, J., Gitiaux, X., Ortega, C., & Rangwala, H. (2022). FairEd: A Systematic Fairness Analysis Approach Applied in a Higher Educational Context. *12th International Learning Analytics and Knowledge Conference, LAK22,* 271-281. https://doi.org/10.1145/3506860.3506902

Villarreal-Torres, H., Ángeles-Morales, J., Marín-Rodriguez, W., & Cano-Mejía, J. (2024). Modelo de clasificación para la deserción estudiantil en las universidades públicas del Perú. *Revista de Ciencias Sociales, 30(1),* 452-469. Scopus. https://doi.org/10.31876/rcs.v30i1.41667

Wild, S. (2023). Trajectories of subject-interests development and influence factors in higher education. *Current Psychology, 42(15),* 12879-12895. https://doi.org/10.1007/s12144-021-02691-7

Yu, R., Lee, H., & Kizilcec, R. F. (2021). Should college dropout prediction models include protected attributes? *Proceedings of the Eighth ACM Conference on Learning at Scale (L@S'21)*, 91–100. https://doi.org/10.1145/3430895.3460139

Zhidkikh, D., Heilala, V., Van Petegem, C., Dawyndt, P., Järvinen, M., Viitanen, S., De Wever, B., Mesuere, B., Lappalainen, V., Kettunen, L., & Hämäläinen, R. (2024). Reproducing Predictive Learning Analytics in

CS1: Toward Generalizable and Explainable Models for Enhancing Student Retention. *Journal of Learning Analytics, 11(1),* 132-150. https://doi.org/10.18608/jla.2024.7979

## AUTHORS

**Mg. Gloria L. Lopez-Muñoz** (corresponding author) is a doctoral student in Computer Science at the University of Cauca, Colombia. She holds a degree in Physics Engineering (2017) from the University of Cauca and a master's degree (2011) in Mechatronic Systems from the University of Brasília (UnB), Brazil. She is an Assistant Professor in the Basic Sciences Department at the National Open and Distance University of Colombia (UNAD). Her research interests include visual servoing, robotics, and machine learning in educational applications.

**Dr. Carolina Gonzalez Serrano** holds a PhD in Computer Science from the University of Vigo (Spain) and a Master's degree in Computer Science from the University of Cauca (Colombia). She is a Full Professor in the Department of Systems at the Faculty of Electronics and Telecommunications Engineering at the University of Cauca and a Senior Researcher with the Computational Intelligence Research Group. Her research focuses on designing and applying intelligent systems in education, particularly in learning analytics and the early identification of student dropout risk in higher education using machine learning techniques. Her interests also include multi-agent systems, intelligent tutoring systems, and STEM/STEAM education.

**Dr. Camilo Sánchez Ferreira** graduated in Physical Engineering from the University of Cauca, Colombia (2008). He holds a master's degree (2012) and a PhD (2017) in Mechatronic Systems from the University of Brasília, Brazil. He is currently a full-time Assistant Professor in the Department of Physics at the University of Cauca. His research interests include analysis and control of dynamic systems, automation, robotics, digital image processing, computer vision, artificial intelligence, machine learning, optimization, and embedded systems (FPGAs).