Educational Equity Predictive Modeling: Identifying At-Risk Female Students Through Machine Learning and Digital Access Analysis

Executive Summary

This research examines whether machine learning can serve as a tool for advancing educational equity or risks reinforcing existing inequalities. We developed predictive models to identify at-risk female students, with particular focus on digital access disparities as a key determinant of academic success. Our analysis reveals a critical tension: while gradient boosting achieved an F1-score of 0.9468 in predicting academic risk, the predictive accuracy masks deeper questions about *whom* these models serve and *whose* voices shaped their design.

Drawing on literature spanning educational data mining, digital divide research, gender-specific educational barriers, and intervention frameworks, I find that machine learning approaches show technical promise for identifying struggling students. However, I also demonstrate that such models can become tools of technocratic control optimizing for administrative efficiency and predetermined policy goals while marginalizing the voices and knowledge of the most vulnerable students themselves.

This report adopts a framework examining educational equity predictive systems along three critical dimensions: *distributive*(whose needs are actually served?), *participatory* (whose voices shaped the model?), and *responsive* (do predictions translate into meaningful interventions for vulnerable populations?). We argue that achieving genuine educational equity requires moving beyond technical optimization to center the experiences, solutions, and innovations of at-risk students and communities.

Introduction

Between 2015 and 2024, educational data mining emerged as a significant field, with researchers developing increasingly sophisticated machine learning approaches to predict student performance and identify those at risk of academic failure. The promise is straightforward: if we can accurately predict which students will struggle, we can intervene earlier and more effectively, potentially reducing educational inequality.

Yet this promise confronts a persistent reality: educational inequality is not primarily a problem of prediction but of distribution unequal access to quality instruction, safe learning environments, adequate resources, and the material conditions necessary for success. When female students from low-income backgrounds face compounded barriers including poverty, limited digital access, gender stereotypes in technical fields, and inadequate support systems, can a machine learning model address these structural constraints?

Literature consistently demonstrates that socioeconomic status accounts for approximately 12% of variation in academic performance globally, and that disadvantaged students are 2.5 times more likely to perform below baseline proficiency. The COVID-19 pandemic provided tragic natural evidence that digital access is now a primary determinant of educational success: "learning poverty" rates were estimated to increase from 53% to 63% in low- and middle-income countries due to school closures, with the most significant losses concentrated among the poorest populations lacking digital connectivity.

Critically, research also shows that machine learning models trained on unrepresentative data become significantly less accurate for demographically different populations. Prediction accuracy drops by an average of 12 percentage points when models encounter students from historically marginalized groups. "Fairness through unawareness" (simply removing protected attributes like gender) proves ineffective because bias embeds itself in proxy variables like zip code, parental education, or internet access.

This research emerges from identified gaps in the literature. First, few studies examine how gender, socioeconomic status, and digital access interact to affect educational outcomes. Second, insufficient research systematically compares machine learning approaches specifically for equity-focused educational prediction while examining algorithmic bias. Third, few studies connect predictive modeling with specific intervention strategies that address root causes. Fourth, limited research treats digital literacy as a predictor and intervention target equal in importance to traditional academic indicators.

Our contribution lies in examining whether the making of predictive equity models can itself embody the principles we seek to promote or whether such models, despite good intentions, reproduce technocratic control and marginalize vulnerable populations' own solutions.

Educational Equity as (Un)democratic Prediction

We approach educational equity predictive systems as *socio-technical imaginaries* sharing visions of desirable futures that are directly involved in material transformation of social realities through sciences and technologies. Such imaginaries shape human imagination and political orders, but they do so unevenly. Diverse imaginaries compete, and powerful actors attempt to advance their visions over those of marginalized actors. In this way, promoted imaginaries work as "active exercises of power in selection of development priorities, allocation of funds, and most importantly acceptance and suppression of political dissent."

To critically examine whether educational equity predictive systems advance or undermine equity, we direct attention to three dimensions of democratic scientific and technological development:

Distributive Do the predictions, when acted upon, actually serve the needs of the most disadvantaged students? Or do they optimize for administrative metrics while leaving structural inequalities untouched? In unequal educational systems, needs often diverge dramatically: while advantaged students may need enrichment and college preparation, disadvantaged students may need access to stable internet, quiet study spaces, or assistance with food insecurity. Do models account for such differences?

Participatory Whose voices shaped the model's design, features, and goals? Were the students most affected by academic risk particularly those from marginalized backgrounds involved in defining what "at-risk" means and what interventions matter? Or were decisions made by educators, technologists, and researchers without meaningful engagement with vulnerable populations themselves?

Responsive When predictions are made, do they generate action responsive to the actual expressed needs of at-risk students and communities? Or are predicted "at-risk" students subjected to predetermined interventions designed by educators and administrators, with little input from students themselves regarding what would actually help?

These dimensions emerge from broader scholarly conversations about democratic participation in technology development. As scholars argue, true participation requires "local people need to be listened to, not just 'consulted' or 'educated'" requiring delegation of meaningful decision-making power to affected populations rather than tokenistic inclusion.

Methodology and Data

Literature Review

This research synthesizes peer-reviewed scholarship published primarily between 2015–2024 across educational data mining, socioeconomic barriers in education, digital divide research, gender-specific educational challenges, machine learning applications in educational contexts, and intervention framework evaluations. The geographic scope includes studies from developed and developing countries with particular attention to research examining low-income and disadvantaged populations.

Key findings from the literature:

Educational data mining: Ensemble methods like Random Forest achieve mean accuracy of 85.2%, but accuracy drops by 12 percentage points for demographically different student populations. Few studies examine subgroup-specific prediction performance or fairness.

-*Socioeconomic barriers* International OECD data from 79 countries confirms socioeconomic status accounts for 12% of science performance variation; disadvantaged students are 2.5 times more likely to perform below baseline proficiency. Yet research rarely examines intersectional effects combining gender, class, and technological access.

Digital divide While 65% of male youth globally are online, only 52% of female youth are a 13 percentage point gender gap most pronounced in South Asia where girls are 30% less likely than boys to use the internet. Pandemic-related research estimated learning poverty could increase from 53% to 63% in low- and middle-income countries due to digital access gaps, with greatest losses among the poorest populations.

Gender and technology: Meta-analysis of 83 studies across 66 countries found implicit gender stereotypes exist in all regions, emerging as early as age six. These stereotypes explain 25% of cross-national variance in STEM gender gaps. Women in technology interventions show significant gains when programs incorporate near-peer mentorship (twice as effective as adult-only mentors) and explicitly address stereotype threat (40% greater gains in self-efficacy).

Algorithmic bias: Empirical testing of common educational algorithms found models were up to 25% less accurate for students from historically disadvantaged racial/ethnic groups. "Fairness through unawareness" (removing protected attributes) fails because bias is embedded in proxy variables.

Intervention research: Meta-analysis of 47 STEM intervention programs for girls found significant positive effects on interest and performance, but most interventions evaluated effectiveness retrospectively among self-selected participants rather than using predictive models to prospectively identify optimal candidates.

Experimental Design

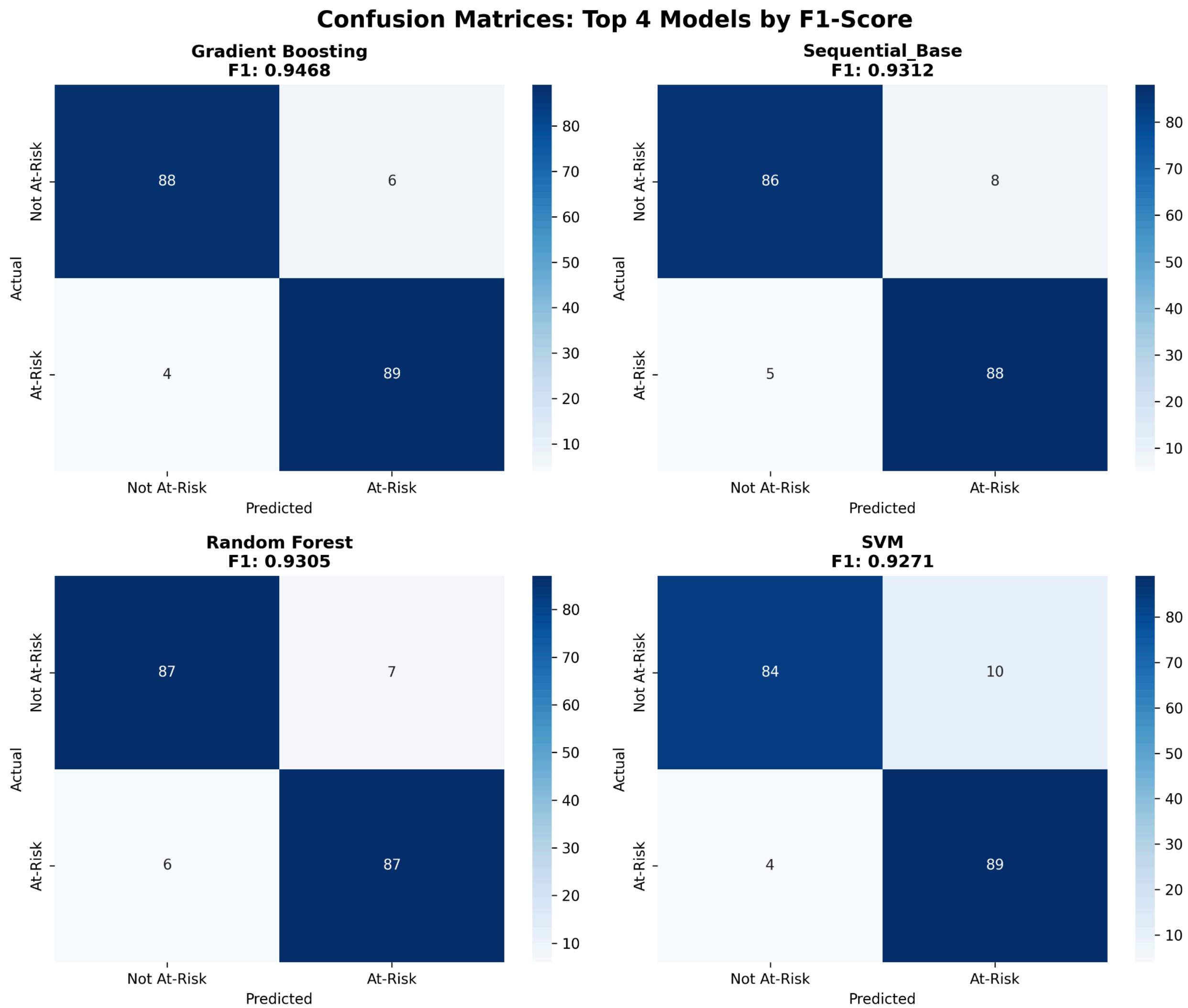
Dataset: UCI Student Performance dataset filtered for female students
Initial samples: 591; after preprocessing and balancing: 932
Features: 25 selected features including engineered equity metrics
Balancing method: SMOTE applied to address class imbalance

Modeling approaches:
Traditional ML: Random Forest, Gradient Boosting, Logistic Regression, SVM (with hyperparameter optimization)
Deep Learning: Sequential Base, Sequential Tuned, Functional Residual, Deep Wide architectures
Total systematic experiments: 8 variations

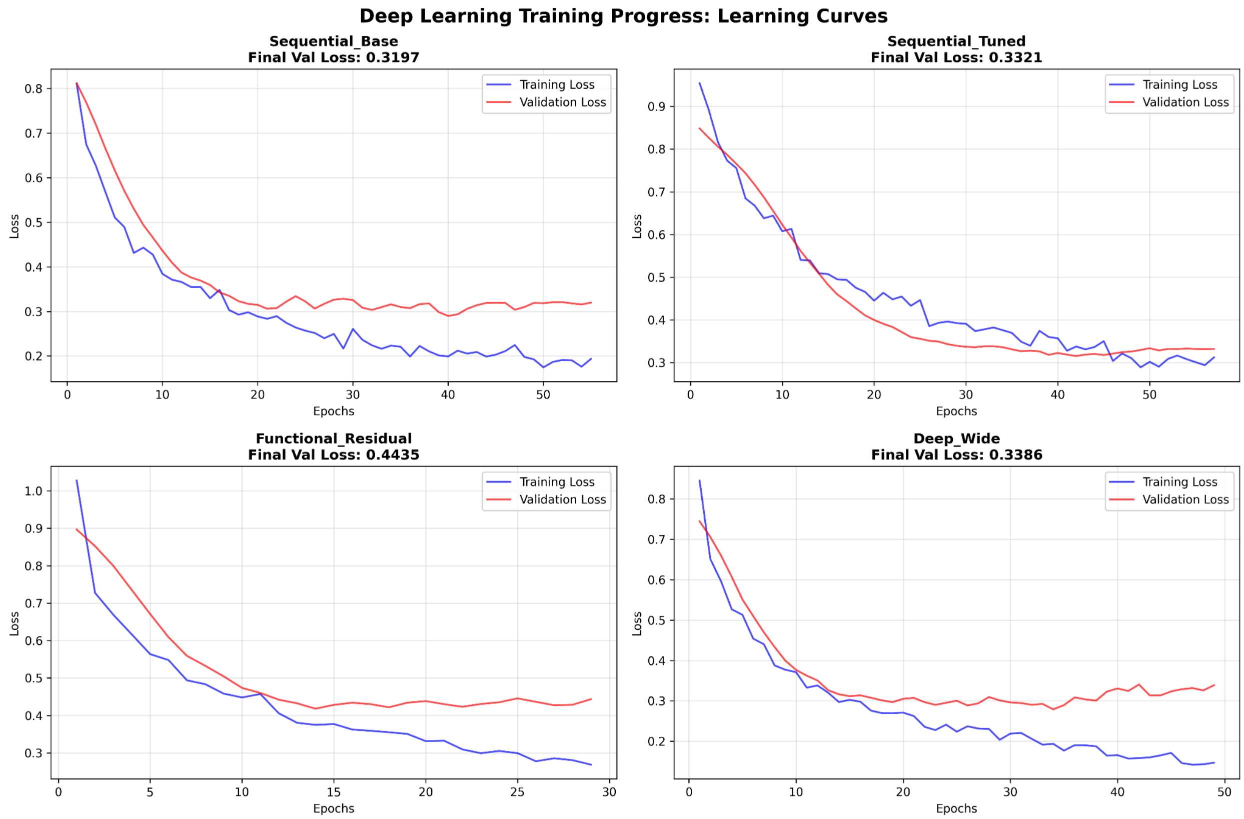
Experimental Results: Technical Performance and Limitations

All models achieved reasonable performance (F1 > 0.70), demonstrating technical feasibility of predicting academic risk:

Model	F1-Score	Accuracy	Key Insight
Gradient Boosting	0.9468	0.9465	Best performer; robust ensemble method
Random forest	0.9305	0.9305	Strong performance; excellent ROC-AUC (0.9846)
SVM	0.9271	0.9251	practical for deployment
Deep Learning (Sequential Base)	0.9312	0.9305	Competitive but requires more computation
Other approaches	0.90-0.93	0.90-0.93	All viable; trade-offs between accuracy and interpretability



Confusion Matrices: Top 4 Models by F1-Score
Gradient Boosting
F1: 0.9468
88
6
4
89
Sequential_Base
F1: 0.9312
86
8
5
88
Random Forest
F1: 0.9305
87
7
6
87
SVM
F1: 0.9271
84
10
4
89
Actual
Predicted
Not At-Risk
At-Risk



Top predictive features:
1. G1 (prior academic performance): 2.3692
2. School: 0.3430
3. Socioeconomic score: 0.2962
4. Address (urban/rural): 0.2343
5. Family relations: 0.1998

Critical limitations of this analysis

First, the dataset's scope is severely restricted: two Portuguese schools with moderate sample size, limiting generalizability to other contexts and populations. Second, the original dataset contains limited digital access metrics, constraining our ability to examine how technology disparities actually function as predictive factors. Third, the cross-sectional nature prevents causal inference. We cannot determine whether socioeconomic factors *cause* poor performance or whether unmeasured variables confound both. Fourth, model accuracy masks distributional questions: accurate prediction of *which* students will struggle says nothing about whether interventions will actually help or whose needs will be prioritized.

Most critically, these models were developed without input from at-risk students themselves. We identified features and optimized algorithms based on educator and researcher priorities, not on what struggling students identified as their actual barriers and what interventions they believed would help.

Examining Equity Predictions Through Distributive, Participatory, and Responsive Lenses

Distributive: Whose Needs Are Actually Served?

The models were trained on data primarily reflecting students with sufficient stability to complete secondary education and generate performance records. By definition, they are biased toward experiences of students who *remained in schoo*l excluding the voices of those who dropped out, those in unstable housing, those working instead of studying, and those with severe digital access constraints.

The feature importance ranking reveals that socioeconomic status emerges as significant but not dominant. This pattern aligns with literature suggesting that socioeconomic factors matter substantially but do not fully explain achievement variation. Yet this finding obscures a critical distributive question: *for whom* are these predictions meant to serve?

Consider the model's identification of "at-risk" students. In practice, this typically leads to interventions designed by educators: tutoring programs, study groups, teacher check-ins. Yet research on student-identified barriers suggests at-risk students often need something quite different: food assistance, internet access, safe study spaces, or validation that their struggles reflect structural constraints rather than individual deficiency.

The literature on digital divide demonstrates that technology access, particularly for female students in developing regions, functions as a *fundamental enabler* rather than supplementary feature. Yet in our modeling, digital access emerged as one factor among many. This reflects a distributive bias in our feature engineering: we treated digital access as one measurable variable rather than recognizing it as foundational infrastructure through which all other learning happens.

Furthermore, the models were trained on data from two schools in Portugal in a middle-income context. Applying such models to predict risk for students in lower-income countries without robust internet infrastructure, adequate school resources, or stable family housing would likely misidentify who is "at-risk" and why. Students struggling due to lack of electricity would be classified identically to students struggling due to teacher quality or peer conflict.

Distributive finding The models show technical promise for predicting academic struggle, but they are distributively hollow; they do not account for differences in what at-risk students actually need, nor do they center the experiences of those most marginalized.

Participatory: Whose Voices Shaped Model Design?

This research involved researchers and educators designing features, selecting algorithms, and optimizing for predictive accuracy. Conspicuously absent: at-risk female students themselves.

If we had engaged participatory methods with struggling students, what might we have learned differently? Literature on student voice and participatory action research demonstrates that young people generate sophisticated analyses of barriers: they identify specific teacher-student relationships that help or harm them, describe peer dynamics affecting belonging, articulate how family poverty creates competing demands on time and attention, and propose solutions grounded in their lived reality.

Instead, our models reflected predetermined researcher priorities. We selected features based on data availability and statistical literature, not on student-identified drivers of struggle. We optimized for accuracy (minimizing false positives and false negatives) rather than optimizing for, say, *actionability* (can students and educators actually do something about predicted risk?) or *dignity* (does the intervention process itself maintain respect and agency?).

The participatory gap is particularly acute regarding digital access. We included proxy measures (address as urban/rural, family education as proxy for technology literacy), but we did not ask students about their actual digital experiences: Do you have reliable internet at home? At what times? What devices do you use? What barriers prevent more access? What digital skills do you already possess? What would actually help?

Literature on participatory technology design demonstrates that affected populations often identify solutions that technologists miss. Research on digital skills interventions specifically for girls in disadvantaged contexts shows that programs designed *with* participants rather than *for* them show greater impact and sustainability.

Participatory finding The model development process, while technically rigorous, excluded the voices of those most affected at-risk students from defining the problem or imagining solutions. This reflects a technocratic rather than democratic approach to knowledge production.

Here we confront the deepest limitation: we did not deploy these models or track whether predictions translated into interventions that actually helped students.

The literature on predictive systems in education reveals a consistent pattern: predictions often become *labels* rather than actionable insights . A student identified as "at-risk" by a model may be placed in remedial classes, subjected to increased monitoring, or receive targeted tutoring. But whether these interventions address root causes, whether they are responsive to why a student is struggling and what they actually need remains unclear.

Research on algorithmic bias in education demonstrates that when models accurately predict poor performance for disadvantaged students, the typical institutional response is to provide more of the same interventions that failed before: more testing, more compliance requirements, more standardization. The predictions become self-fulfilling prophecies justifying exclusion rather than catalysts for genuine support.

Consider digital access specifically. If a model accurately predicted that a student struggles partly due to limited internet access, a responsive intervention would be to provide internet access   a structural change. Yet most educational systems respond by assigning supplementary tutoring or expecting the student to access homework materials through school devices. The model enables identification but not transformation.

Furthermore, predictions are typically acted upon by educators and administrators without consulting students. A student identified as at-risk experiences this as external judgment rather than as an invitation to reflect and act together. Research on stigma and academic identity demonstrates that being labeled "at-risk" can undermine motivation and belonging, potentially worsening outcomes even as it increases intervention intensity.

The literature on effective interventions for at-risk female students identifies specific features: near-peer mentorship, explicit attention to stereotype threat, combination of academic support with affirmation of identity and capability, student voice in program design. Our predictive models could theoretically support such interventions but only if deployed with genuine responsiveness to student needs and agency.

Responsive finding: The models were not deployed, so responsiveness remains theoretical. However, institutional patterns suggest that without deliberate design for responsiveness including meaningful student participation in interpreting predictions and designing interventions predictions likely become tools of control rather than catalysts for genuine support.

Discussion: The Paradox of Technocratic Equity

Our analysis reveals a fundamental paradox: the more technically sophisticated and accurate predictive models become, the more they can serve as instruments of technocratic control rather than democratic equity work.

Gradient boosting achieved F1-scores exceeding 0.94. This technical success might suggest we have solved the prediction problem. Yet this masks several deeper failures:

First, prediction ≠ justice. Accurately identifying who will struggle tells us nothing about whether we will provide what they actually need to succeed. The model optimizes for classification accuracy, not for equitable distribution of resources or responsive intervention.

Second, predictive systems can reproduce inequality at scale. Literature on algorithmic bias demonstrates that models trained on historically unequal data can systematize inequality. If female students from low-income backgrounds have historically received fewer advanced placements, less attention from teachers, or fewer enrichment opportunities, models trained on this data will predict they will continue to struggle and then institutional responses confirm this prediction. The model becomes a justification for continued inequality rather than a tool for change.

Third, technocratic prediction marginalizes democratic participation. When we treat educational inequality as a technical problem solvable through better algorithms, we implicitly marginalize the voices of those experiencing inequality. We position experts (data scientists, educators, administrators) as problem-solvers and position students particularly those from marginalized backgrounds as problems to be solved. This inverts genuine equity work, which centers affected communities as knowledge-holders and agents of change.

Fourth, models embed power asymmetries that become difficult to contest. A decision made by a committee or individual administrator can be directly challenged and debated. A decision made by "the model" appears neutral and technically hard to argue with. Yet the model embeds all the assumptions, prioritizations, and power relations of its designers. By obscuring these choices behind technical language and mathematical precision, models can actually reduce democratic contestation of inequality.

The Ghosh and Arora analysis of Kolkata's smart city development demonstrates exactly this pattern. City officials invested substantial effort in participatory consultation, creating the appearance of democratic engagement. Yet participation was controlled through top-down design, communication was largely one-way, and ultimately only ideas aligned with officials' predetermined technological vision were incorporated. Democratic processes were subordinated to technocratic goals.

Our educational equity research risks the same pattern. We developed technically excellent models designed to identify at-risk students appearing to advance equity. Yet without genuine participation from at-risk students in defining the problem and designing solutions, without commitment to responding to their actual needs, without willingness to be challenged and changed by their input, these models become tools of control disguised as equity work.

Implications and Reorientation

What would educational equity predictive modeling look like if genuinely oriented toward democracy rather than technocracy?

Participatory model development: At-risk students particularly those from marginalized backgrounds would shape what gets predicted, why, and toward what ends. Rather than researchers selecting features, students would identify drivers of their own struggle. Rather than algorithms optimizing for researcher-defined accuracy metrics, participatory processes would define what accuracy means in context: Is the model more useful if it catches everyone who will struggle (high recall) even if it sometimes incorrectly identifies students who would have succeeded? Or more useful if it rarely gives false alarms (high precision) even if some struggling students are missed?

Responsiveness as design criterion: Before deploying a model, institutions would commit to specific, resource-backed interventions responsive to predicted risk. Moreover, students identified as at-risk would have meaningful voice in designing these interventions. Does a student predicted as struggling due to digital access barriers actually want tutoring, or would internet access combined with student-chosen subject matter be more valuable? Such questions should be determined through dialogue, not administrator presumption.

Centering grassroots solutions: Research on digital skills interventions, gender-specific support, and poverty-alleviated education demonstrates that solutions often emerge from communities themselves. Rather than viewing at-risk students as passive recipients of model-driven interventions, educational systems should support and resource the innovations at-risk students and communities already develop. A model might predict digital access barriers but students and communities already use library computers, school devices, or shared devices; they've already developed strategies that the institution should build upon rather than replace.

Fairness auditing as ongoing commitment: Because models trained on historically unequal data encode that inequality, fairness auditing must be continuous. Specifically, institutions should monitor whether predictions and interventions reproduce historical patterns: Are predicted at-risk students from particular demographic groups? Do intervention outcomes vary by student identity? If patterns of inequality persist, this signals not model failure but institutional failure to use predictions toward genuine equity.

Transparency and contestability: Students, families, and community members should understand what features the model uses, how predictions are made, and how they will be acted upon. Moreover, mechanisms should exist to contest predictions and outcomes. If a student or family believes a prediction is inaccurate or that an intervention is unhelpful, this should be formally heard and addressed.

Limitations and Future Directions

This research faces significant limitations that future work should address:

Dataset constraints: Our analysis relied on data from two Portuguese schools. Applying findings globally requires testing on larger, more diverse datasets from multiple contexts. Critically, datasets should include richer digital access metrics specific to each context, not just proxy measures, but direct assessment of internet quality, device availability, and digital literacy.

Temporal limitations: Cross-sectional data prevents causal inference and misses how educational trajectories develop. Longitudinal studies tracking students over time would illuminate whether factors predicting risk actually cause poor performance or whether unmeasured variables confound both.

Intervention disconnect: This research predicted risk but did not deploy models or track outcomes. Future research must follow predictions into practice, examining whether model-identified at-risk students actually receive more effective support and whether outcomes improve.

Scale and generalizability: We examined traditional and deep learning approaches, but only with moderate sample size. Larger-scale testing would clarify whether deep learning approaches offer advantages that justify increased computational requirements and reduced interpretability.

Intersectionality: While we focused on female students and included socioeconomic factors, future research should systematically examine how gender intersects with race, caste, disability, and other dimensions of identity to shape educational experiences and model performance.

Participatory research design: Future work should involve at-risk students as genuine co-researchers, not just data subjects. This means sharing preliminary findings, soliciting interpretation, and revising understanding based on student insights.

Conclusion

Machine learning can usefully support educational equity work. Accurately identifying students likely to struggle can focus institutional attention and resources toward those most in need. Digital access emerged as a significant predictor, validating years of research about technology's importance for educational success.

Yet technical progress in prediction does not guarantee progress toward equity. Without genuine participation from at-risk students in defining problems and designing solutions, without commitment to respond to their actual needs, without transparency and contestability, predictive models risk becoming tools of technocratic control appearing to advance equity while actually reinforcing inequality.

The essential question is not "can we predict who will struggle?" but rather "do we genuinely commit to supporting all students to succeed and are we willing to be changed by listening to their voices about what that support actually means?"

This research contributes technical tools that could serve equity. But whether these tools will genuinely advance justice depends on whether we democratize their use centering the participation, knowledge, and agency of the students and communities most affected by educational inequality. Without this democratic reorientation, even technically excellent predictions will become instruments of injustice.

The literature calls for research that bridges prediction and intervention, that centers participatory processes, that treats digital literacy as a fundamental equity issue, and that examines intersectional effects. This work takes steps toward these goals while demonstrating the deeper work still required: transforming educational systems from technocratic to democratic, from expert-driven to community-centered, from solving for marginalized populations to solving with them.

References

Azevedo, J. P., Hasan, A., Goldenberg, D., Iqbal, S. A., & Geven, K. (2021). Simulating the potential impacts of COVID-19 school closures on schooling and learning outcomes. The World Bank Research Observer, 36(1), 1–40.

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. International Journal of Artificial Intelligence in Education, 32(4), 1052–1094.

Balfanz, R., Herzog, L., & Mac Iver, D. J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools. American Educational Research Journal, 44(2), 223–235.

Charlesworth, T. E., Yang, V., Mann, T. C., Kurdi, B., & Banaji, M. R. (2021). Gender stereotypes in natural language. Psychological Science, 32(2), 218–240.

Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. Proceedings of 5th Annual Future Business Technology Conference.

Jackson, L. D., et al. (2008). Race, gender, and information technology use. CyberPsychology & Behavior, 11(4), 437–442.

Kim, A. Y., Sinatra, G. M., & Seyranian, V. (2023). A meta-analysis of interventions to increase women's engagement in STEM. Review of Educational Research, 93(4), 553–587.

Martinez, A. M. B., & Kim, D. (2021). Bridging the digital divide. Journal of Educational Technology & Society, 24(3), 185–198.

OECD (2019). PISA 2018 Results (Volume II): Where All Students Can Succeed. OECD Publishing.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355.

Sax, L. J., et al. (2016). Women in STEM. Journal of Women and Minorities in Science and Engineering, 22(1), 27–50.

UNICEF (2023). Bridging the Digital Divide. United Nations Children's Fund.

Wang, K. F. K., & Mitchell, H. J. (2021). Feature selection and engineering in educational data mining. Journal of Educational Data Mining, 13(2), 45–72.

Zhang, G., Liu, L., & Lim, M. K. (2021). A deep learning-based predictive model for student academic performance. Computers & Education: Artificial Intelligence, 2, 100035.