

Beyond Marks: Detecting Hidden Academic Risk Through Psychosocial Proxy Indicators, Machine Learning and Explainable AI in Sri Lankan Government Secondary Schools

H.A.H.D. Wijayarathna^{1*}, E.R.C. Sandamali¹, D.C. Wickramarachchi²

¹Department of Computational Mathematics, Faculty of Computing, General Sir John Kotelawala Defence University, Ratmalana, Sri Lanka

²Department of Statistics, Faculty of Applied Sciences, University of Sri Jayewardenepura, Nugegoda, Sri Lanka

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.1026EDU0560>

Received: 15 August 2026; Accepted: 20 August 2026; Published: 09 September 2026

ABSTRACT

Background: Academic monitoring in Sri Lankan government secondary schools relies exclusively on term marks, leaving the psychosocial conditions that accumulate into failure entirely invisible. No validated, explainable early warning system exists for this context, and global early warning systems depend on Learning Management System infrastructure unavailable in Sri Lankan government schools. **Methods:** A 24-item psychosocial questionnaire spanning five theoretically grounded domains — Socioeconomic Burden, Cognitive Load, School Environment, Motivation, and Physical Readiness — was developed through five months of structured field engagement with 12 school principals and 45 classroom teachers, validated by the Professional Psychological Counselors Association of Sri Lanka (PPCA), and approved through three institutional levels. Combined with six-term academic trajectory data from 873 Grade 11 students across 11 government schools in Piliyandala Educational Zone, eight interpretable features were engineered. Four candidate models were evaluated using 10-fold stratified cross-validation; logistic regression was selected. SHAP (SHapley Additive exPlanations) LinearExplainer was applied to produce exact, student-level risk explanations. **Results:** The full model achieved CV AUC=0.878 (95% CI: 0.858–0.899), a validated 6.1-point gain over the academic-only baseline (AUC=0.818), confirming independent psychosocial contribution. Failure rates rose monotonically from 7.1% to 58.2% across the psychosocial burden spectrum. SHAP analysis identified Cognitive Load and Motivation as the primary actionable risk drivers beyond marks. Critically, 219 students (25.1%) carrying above-median psychosocial burden while currently passing were identified as entirely invisible to conventional marks-only monitoring. Gender fairness evaluation confirmed equitable model performance across four independent statistical tests. **Conclusion:** The resulting EduRisk prototype delivers SHAP-grounded, student-level risk profiles covering the psychosocial dimensions that marks-only monitoring structurally cannot detect, providing teachers and school authorities with actionable intelligence to intervene before psychosocial burden accumulates into irreversible academic failure. The methodology offers a replicable framework for resource-constrained educational contexts globally.

Keywords: Academic risk prediction, Early warning system, Explainable artificial intelligence, Psychosocial indicators, Educational equity

INTRODUCTION

The GCE Ordinary Level examination is the pivotal academic milestone for secondary school students in Sri Lanka, determining eligibility for advanced study and shaping career trajectories. Despite sustained investment in primary and secondary education, a significant proportion of students approach this examination carrying undetected risk factors that, if identified and addressed early, could be meaningfully mitigated. In the Piliyandala Educational Zone of Western Province, 27.3% of Grade 11 students across 11 government secondary schools

failed at least one core subject in 2025 — a failure rate that represents not only an individual academic setback but a systemic equity failure in which the most vulnerable students receive the least timely support.

The prevailing model of academic monitoring in Sri Lankan government secondary schools is marks-only: teachers observe term mark performance and raise concern only after failure has manifested, by which point the intervention window has effectively closed. This approach is fundamentally reactive and single-dimensional. International research has consistently demonstrated that academic outcomes are shaped not only by performance history but by a constellation of psychosocial factors — socioeconomic constraints, cognitive burden, school environment quality, motivational states, and physical readiness — that are entirely invisible to marks-based monitoring (Berens et al., 2019; Vinas-Forcade et al., 2021). A student may be passing on paper while accumulating psychosocial burden that will manifest as academic failure only when intervention is no longer possible.

Early warning systems (EWS) in education have sought to address this gap through data-driven prediction. However, the dominant paradigm relies on Learning Management System (LMS) infrastructure — clickstream data, login frequency, assignment submission timing — that does not exist in Sri Lankan government secondary schools (Macfadyen & Dawson, 2010; Zhang et al., 2025). This renders the global EWS literature largely inapplicable to this context. Furthermore, no prior study has implemented a validated, explainable, and deployable EWS for Sri Lankan government secondary schools. The absence of such a system was formally acknowledged by the Deputy Director of the Ministry of Education Research and Development Branch in November 2025, establishing the institutional motivation for the present research.

This study addresses this gap through the development, validation, and implementation of EduRisk — an Explainable AI-based early academic risk detection system that operates exclusively on proxy learning indicators measurable without digital infrastructure. Four research questions guide the study: (RQ1) Which proxy indicators of learning conditions are significantly associated with academic failure among Grade 11 students in Piliyandala Educational Zone? (RQ2) Do psychosocial composite features contribute statistically significant predictive value beyond academic trajectory features alone? (RQ3) Which predictors most strongly distinguish at-risk from non-at-risk students in a multivariate model? (RQ4) Can explainable AI techniques provide interpretable and actionable insights for early academic intervention in schools with limited digital infrastructure?

The study makes four distinct contributions. Methodologically, it demonstrates that LMS-independent early warning is achievable at $AUC=0.878$ using only school-register marks and a validated questionnaire. Empirically, it provides the first quantitative evidence of a hidden risk population in Sri Lankan government secondary schools — 268 students (30.7%) carrying above-median psychosocial burden while currently passing, entirely invisible to marks-only monitoring. Practically, it operationalises these findings into a deployable decision support system — EduRisk — that a class teacher with no data science background can operate on a single school computer. At the policy level, it provides evidence that the infrastructure gap acknowledged by the Ministry of Education is addressable within existing school resources, directly contributing to Sustainable Development Goal 4 (Quality Education).

LITERATURE REVIEW

LMS-Dependent Early Warning Systems

The most extensively developed early warning system literature focuses on Learning Management System environments in higher education. Macfadyen and Dawson (2010) demonstrated that LMS engagement data — login frequency, content views, and discussion forum participation — predicted student failure with meaningful accuracy, identifying 12 variables explaining 51% of variance in final grade. Trakunphutthirak and Lee (2022) extended this by showing that progressive accumulation of LMS and internet usage data across the academic term improved prediction accuracy incrementally — the longer the term continued, the more accurate the system became. Karalar et al. (2021) applied ensemble methods to LMS-based distance learning risk detection, demonstrating that combining multiple classifiers improved identification of at-risk students in online environments. These studies collectively establish that digital behavioural traces are powerful predictors of

academic outcomes. However, they share a fundamental and disqualifying limitation for the Sri Lankan context: every variable they rely upon — clickstreams, login timestamps, submission data, internet usage logs — requires digital infrastructure that does not exist in government secondary schools. Not a single teacher in Piliyandala zone has access to these data types, making the direct transfer of these approaches structurally impossible rather than merely technically challenging.

Administrative and Statistical Risk Models

A second stream of research avoids LMS dependency by using administrative records — historical grades, demographic data, and institutional information — to identify at-risk students. Berens et al. (2019) conducted a comprehensive systematic review of EWS approaches across 24 institutions, finding that multimodal systems combining academic and non-academic indicators consistently outperformed single-modality approaches, with the best models achieving AUC values above 0.85 using logistic regression on administrative data alone. Vinas-Forcade et al. (2021) used longitudinal administrative data to identify at-risk students upon secondary school entry in Belgium, demonstrating that socioeconomic background and prior academic trajectory — neither of which requires LMS — could meaningfully stratify failure risk. Kotsiantis et al. (2004) evaluated machine learning techniques including decision trees, neural networks, and logistic regression for student performance prediction, establishing logistic regression as consistently competitive with more complex approaches across diverse educational datasets. While these administrative approaches are more transferable to resource-constrained contexts, they do not address the specific psychosocial measurement challenge in Sri Lankan government schools: no validated instrument for capturing non-academic psychosocial risk indicators had previously existed for this context, and no prior study had deployed such findings into an operational decision support system accessible to teachers without data science expertise.

Explainable AI and Psychosocial Determinants in Education

The emergence of Explainable AI (XAI) as a distinct field has particular significance for educational applications, where predictions must be understood and acted upon by teachers without data science expertise. Choi et al. (2025) conducted a systematic literature review identifying SHAP (SHapley Additive exPlanations) as the leading XAI technique for EWS deployment, noting its theoretical grounding in cooperative game theory, its additive property ensuring individual feature contributions sum to the prediction, and its capacity to generate both global population-level and local student-level explanations.

Dai et al. (2025) demonstrated empirically that EWS systems generating actionable guidance — specifying what teachers should do rather than merely assigning a risk score — produced significantly higher rates of practitioner engagement and intervention follow-through compared to score-only systems. This finding directly motivates EduRisk's integrated intervention layer.

In the Sri Lankan context, Dayasiri et al. (2025) confirmed socioeconomic burden and school environment as significant predictors of educational disengagement among adolescents in rural Sri Lanka, while Mudunna et al. (2025) documented high prevalence of mental health challenges among school-going adolescents — both studies providing contextual grounding for the psychosocial domains incorporated into the EduRisk instrument. Rumberger (2011) established through longitudinal research that dropout and academic failure are rarely sudden events but the culmination of long-building psychosocial and academic pressures — a finding that fundamentally motivates early, psychosocially-informed identification over reactive marks-based monitoring. UNESCO (2021) explicitly identified the absence of evidence-based monitoring systems in resource-constrained educational contexts as a global priority gap requiring contextually adapted solutions.

Research Gap

Three interconnected gaps motivate the present study. The infrastructural gap: no validated EWS exists for Sri Lankan government secondary schools, and all prior high-performing EWS research assumes LMS infrastructure unavailable in this context. Even the most contextually sensitive administrative approaches reviewed above were developed in countries with robust educational data systems and cannot be directly applied without fundamental adaptation.

The methodological gap: no prior study has combined validated psychosocial measurement, supervised machine learning, SHAP explainability, and an expert-validated intervention framework in a single integrated pipeline designed for deployment in schools with limited digital resources. Each component has been studied separately, but their integration into a coherent, deployable system represents a novel methodological contribution. The operational gap: prediction without actionable follow-through does not constitute an early warning system (Dai et al., 2025). The absence of a contextually validated intervention framework means that even if a prediction system were available, school staff would lack structured guidance on how to respond. EduRisk addresses all three gaps simultaneously — developing the instrument, building the model, generating the explanations, and providing the intervention guidance — as a single integrated contribution to both the research literature and Sri Lankan educational practice.

METHODOLOGY

Research Design

This study follows a cross-sectional, quantitative research design grounded in positivism, consistent with established practice in educational data mining research (Kotsiantis et al., 2004). The methodology is structured across ten phases: questionnaire design (Phase 1), instrument development and validation (Phase 2), data collection (Phase 3), data preprocessing (Phase 4), exploratory data analysis (Phase 5), feature engineering (Phase 6), predictive modelling (Phase 7), SHAP explainability and fairness analysis (Phase 8), intervention mapping and recommendation system development (Phase 9), and EduRisk prototype development and validation (Phase 10). Figure 1 presents the complete ten-phase research framework.

The research adheres to the principle of methodological parsimony: every analytical decision was made to produce a system deployable within existing school constraints, interpretable by non-specialist users, and replicable without proprietary software or digital infrastructure.

Ethical Considerations

This study received three-level institutional approval prior to any data collection or instrument administration. Formal approval was obtained sequentially from: the Ministry of Education Research and Development Branch at the national level; the Western Provincial Education Ministry at the provincial level; and the Piliyandala Zonal Education Ministry at the zone level. This hierarchical approval process required submission of the full research protocol, instrument drafts, and data handling procedures at each level, with each level independently reviewing and approving the study before the next level was approached.

The 24-item psychosocial questionnaire was further reviewed and refined by the Professional Psychological Counselors Association of Sri Lanka (PPCA) to ensure that all items were psychologically appropriate for the Grade 11 age group, that no item would cause distress, and that the instrument as a whole met professional standards for psychological assessment in an educational setting. All student participation was entirely voluntary — students were informed they could withdraw at any time without consequence. The questionnaire was administered anonymously: no student names were recorded on any questionnaire form.

Anonymised student IDs were used solely for matching questionnaire responses with academic mark records. School-level identifiers were excluded from the final dataset to preserve institutional anonymity, which is why school-type subgroup analysis was not conducted. Academic marks were transcribed from school registers by the research team under the direct supervision of school principals, with no marks data leaving the school premises in identifiable form. All data were stored on password-protected devices accessible only to the research team, consistent with the data handling commitments made during the institutional approval process.

EduRisk Research Methodology — Eight-Phase Framework

Phase	Phase Name	Key Activities
1	Questionnaire Design	Structured consultations with 12+ principals and 45 teachers, domain identification, 29-item draft instrument
2	Instrument Development and Validation	PPCA expert review, three-level institutional approval, pilot psychometric validation (n=461, KMO=0.897)
3	Data Collection	873 Grade 11 students, 11 government schools, Piliyandala zone, manual mark transcription
4	Data Preprocessing	Five quality checks, missing value imputation, outlier analysis, EPV verification
5	Exploratory Data Analysis	Dose-response analysis, resilience analysis, failure cascade analysis, MANOVA
6	Feature Engineering	VIF-based elimination, PCA composite computation, final 8-feature set
7	Predictive Modelling	Four candidate models, 10-fold stratified CV, two-criterion selection rule, LR selected (AUC=0.878)
8	SHAP Explainability, Fairness Analysis and EduRisk Prototype	SHAP LinearExplainer global and local analysis, gender fairness evaluation, deployable desktop application

Figure 1. EduRisk research methodology: ten-phase framework from questionnaire design through instrument validation, data collection, modelling, SHAP explainability, intervention mapping, and EduRisk prototype development.

Research Context and Participants

Data were collected from 873 Grade 11 students across all 11 government secondary schools in Piliyandala Educational Zone, Western Province, Sri Lanka, in 2025. The schools included mixed, single-sex, large urban, and small urban institutions, providing a census-level sample of the zone rather than a subset. The sample comprised 457 male (52.3%) and 416 female (47.7%) students. The overall core subject failure rate was 27.3% (238 students), with subject-specific rates of 22.5% for Mathematics (n=196), 12.4% for Science (n=108), 7.2% for English (n=63), and 6.4% for Sinhala (n=56). Three-level institutional approval was secured from the Ministry of Education R&D Branch, Western Provincial Education Ministry, and Piliyandala Zonal Education Ministry in January 2026. All participation was voluntary and anonymous.

Instrument Development

The 24-item psychosocial questionnaire was developed through a structured two-phase process representing a formal research contribution in its own right. In the first stage of instrument development (November–December 2025), the researcher conducted structured consultations with 12+ school principals and 45 classroom teachers across Piliyandala zone — not merely for approval purposes but to systematically evaluate every proposed

domain and indicator against practitioner knowledge and lived school experience. This two-month process, requiring sustained physical presence in the research zone including three consecutive weeks of daily principal meetings, produced a 29-item draft instrument. In the validation stage (December 2025–January 2026), the PPCA refined the instrument to 24 items across five validated domains on a 1–5 Likert scale, where higher scores indicate greater burden: Socioeconomic Burden (6 items: financial strain, household labour interference, caregiving responsibilities, education exit pressure, study space quality, adult educational support), Cognitive Load (5 items: mental exhaustion, academic overwhelm, attention difficulty, lesson comprehension difficulty, exam preparation difficulty), School Environment (4 items: teacher support lack, teacher unfairness, safety concerns, peer conflict), Motivation (4 items: academic hopelessness, failure response, emotional difficulty, school frustration), and Physical Readiness (5 items: travel fatigue, sleep insufficiency, morning fatigue, hunger during school, health impact). All items are phrased as first-person burden statements to reduce social desirability bias.

Psychometric Validation

A pilot study (n=461, 6 schools, January 2026) confirmed adequate sampling for factor analysis (KMO=0.897, Bartlett's $\chi^2=4,989.9$, $p<0.001$) with strong overall reliability (Cronbach's $\alpha=0.861$). Full exploratory factor analysis (EFA) on n=873 confirmed the five-factor structure (KMO=0.921, Bartlett's $\chi^2=2,450.7$, $p<0.001$), with Varimax rotation producing five factors explaining 49.2% of total variance. All 24 items loaded highest on their theoretically expected domain with no cross-loadings exceeding 0.30. Construct validation results are presented in Table 1.

Table 1. Instrument construct validation results (n=873)

Domain	Items	α	CR	AVE	Note
Socioeconomic Burden	6	0.759	0.836	0.462	*
Cognitive Load	5	0.764	0.855	0.543	
School Environment	4	0.615	0.801	0.503	
Motivation	4	0.619	0.801	0.502	
Physical Readiness	5	0.644	0.789	0.431	*
Overall (pilot)	24	0.861	—	—	

*AVE below 0.50 threshold; defended by CR >0.70 and all standardised loadings >0.50, consistent with Fornell and Larcker (1981).

Data Collection

Academic marks were manually transcribed from handwritten school registers for Grade 9 (three terms) and Grade 10 (three terms), producing 24-mark variables per student across four core subjects. Grade 11 third term marks served as the outcome variable only and were not used as predictors. The psychosocial questionnaire was administered to Grade 11 students during their third term, capturing psychosocial state contemporaneous with the academic outcome period. Data collection required three additional team members and three weeks of continuous physical presence by the researcher in the Piliyandala zone, as no digital records existed across any of the 11 schools. Five automated quality checks confirmed data integrity: zero missing questionnaire items, zero range violations, zero straight-line responses, 100% outcome flag consistency, and zero duplicate records.

Feature Engineering

Nine academic candidate features were computed from the 24 raw mark variables. Variance Inflation Factor (VIF) analysis revealed extreme multicollinearity (maximum VIF=656), requiring iterative VIF-based elimination. Six candidates were removed, retaining three academic features capturing level (Recent Average Term Mark, VIF=1.14), trajectory (Mark Decline Rate, VIF=1.19), and consistency (Academic Variability, VIF=1.05). Five psychosocial composite features were computed via Principal Component Analysis (one component per domain), with StandardScaler preprocessing and direction correction to ensure higher composite scores corresponded to higher psychosocial burden. Final VIF values for all eight features ranged from 1.051 to 1.741, confirming complete resolution of multicollinearity.

Table 2. Final eight-feature set with VIF and effect sizes

Feature	Type	VIF	r with outcome	Cohen's d
Recent Avg Term Mark	Academic	1.14	-0.501	—
Mark Decline Rate	Academic	1.19	-0.208	—
Academic Variability	Academic	1.05	+0.018	—
Socioeconomic Burden	Psychosocial	1.74	+0.295	0.606
Cognitive Load	Psychosocial	1.54	+0.330	0.784
School Environment	Psychosocial	1.38	+0.218	0.502
Motivation	Psychosocial	1.45	+0.263	0.680
Physical Readiness	Psychosocial	1.34	+0.137	0.309

Predictive Modelling

Four candidate models were evaluated using 10-fold stratified cross-validation with `class_weight='balanced'` to address the 1:2.67 class imbalance: Logistic Regression (L2 regularisation, $C=1.0$), Random Forest (500 trees), XGBoost (300 estimators), and a Stacking Ensemble (LR+RF+XGB base, LR meta-learner). Standardisation was applied on training folds only to prevent data leakage. Per-fold thresholds were optimised using Youden's J statistic. A pre-specified two-criterion selection rule required any challenger model to exceed Logistic Regression by more than 2 validated AUC points AND achieve $p<0.05$ on a paired t-test across 10 folds simultaneously. A train/test split was rejected on statistical grounds: with 238 failure events, a 20% hold-out would yield approximately 48 test failures, below the minimum threshold for stable evaluation recommended by Harrell (2015). Class imbalance was addressed through cost-sensitive learning (`class_weight='balanced'`) rather than synthetic data generation, preserving the authenticity of the real student data and the validity of all SHAP explanations.

SHAP Explainability

SHAP LinearExplainer was applied to the final trained Logistic Regression model — the only SHAP method producing exact, zero-approximation-error explanations, achievable because logistic regression is a linear model. An Independent masker using all 873 students as the background distribution was applied. Global analysis produced a feature importance ranking (mean absolute SHAP values) and a beeswarm summary plot. Local analysis produced individual waterfall plots for three principled case studies selected on pre-specified criteria: a representative High Risk failing student (STU0181, probability=0.904), a hidden risk currently-passing student (STU0358, probability=0.377), and a low-risk resilient student (STU0252, probability=0.115).

Gender Fairness Evaluation

Four independent statistical tests evaluated gender subgroup equity: chi-square for failure rate difference, Mann-Whitney U for AUC difference across folds, chi-square for risk tier distribution, and Mann-Whitney U for each of the eight SHAP feature importance distributions. School-type analysis was excluded on institutional anonymity grounds: with 11 schools in a single zone, any school-level attribute would make individual institutions identifiable, violating the approval commitment.

RESULTS

Exploratory Data Analysis: Psychosocial Burden and Academic Failure

Point-biserial correlation confirmed that 23 of 24 questionnaire items were significantly associated with the failure outcome (19 at $p<0.001$, 3 at $p<0.05$; the single non-significant item was `health_impact`, $r=0.064$). The three strongest items were `academic_overwhelm` ($r=0.264$), `mental_exhaustion` ($r=0.263$), and `lesson_comprehension_difficulty` ($r=0.262$) — all from the Cognitive Load domain — providing preliminary evidence of its primacy. Academic marks showed stronger mean absolute correlation ($|r|=0.316$) than questionnaire items ($|r|=0.172$), confirming marks as the strongest individual signal while psychosocial items contribute meaningful additive information. MANOVA confirmed significant differences across all five

psychosocial domains between passing and failing students ($F=27.04$, $p<0.001$, all five domains significant after Bonferroni correction).

A dose-response analysis revealed a monotonically increasing failure rate, illustrated in Figure 2, from 7.1% when zero psychosocial domains were elevated, to 58.2% when all five were elevated — a relative risk of $8.1\times$ across the full spectrum ($\chi^2=101.62$, $p<0.001$). This dose-dependent relationship confirms that psychosocial burden is not merely correlated with failure but quantitatively predictive of it. Each additional elevated domain independently increases failure probability regardless of academic performance level.

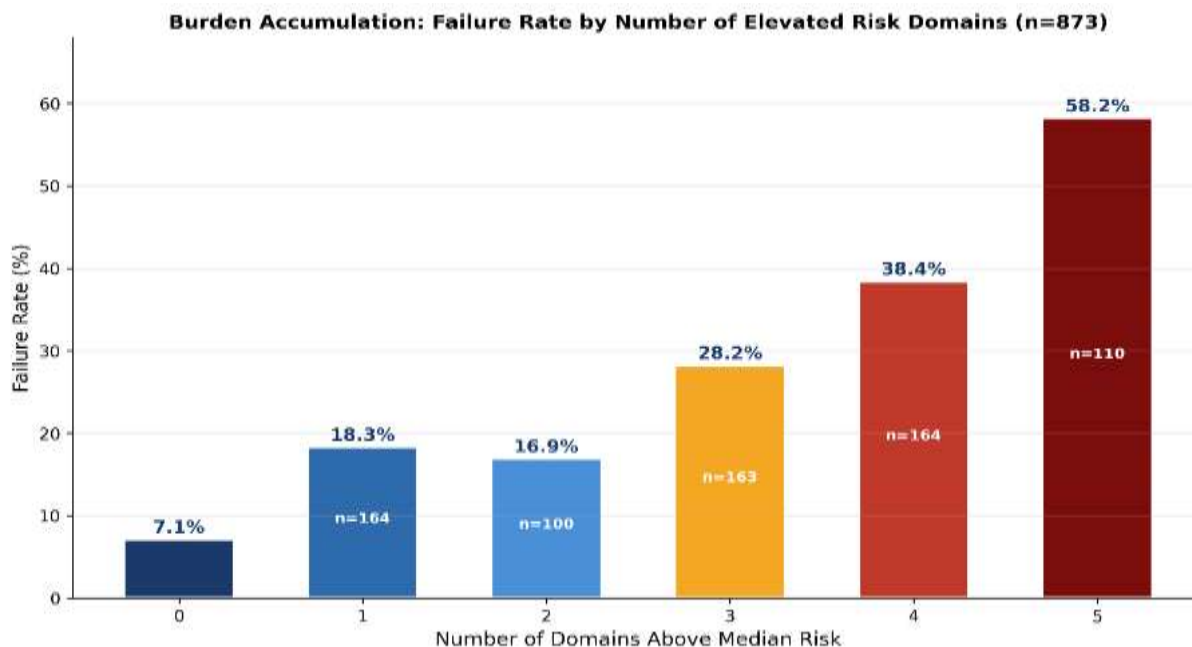


Figure 2. Failure rate (%) by number of elevated psychosocial domains (0=none elevated, 5=all five elevated; $n=873$, $\chi^2=101.62$, $p<0.001$, $RR=8.1\times$)

A critical population-level finding emerged from the academic feature analysis: the mean mark decline rate across all 873 students was -7.77% from Grade 9 to Grade 10. This is not a characteristic of at-risk students alone — it is universal, suggesting a systemic academic pressure point at the Grade 9→Grade 10 transition in Piliyandala zone schools, with implications extending beyond individual student risk monitoring toward zone-level academic support policy.

The failure cascade pattern further confirms that academic difficulty in this population is systemic rather than subject-isolated, as illustrated in Figure 3 (Right). Subject-level failure rates confirmed Mathematics as the most critical subject (22.5%, $n=196$), followed by Science (12.4%, $n=108$), English (7.2%, $n=63$), and Sinhala (6.4%, $n=56$). Critically, 80.4% of Sinhala failures, 76.2% of English failures, and 71.3% of Science failures co-occurred with Mathematics failure — confirming that by the time academic difficulty becomes visible across subjects, the underlying conditions have been building for months. This cascade pattern strongly motivates whole-student psychosocial monitoring over subject-specific remediation: addressing the psychosocial root causes before the cascade begins is more impactful than responding subject by subject after it has spread.

The hidden risk population represents the most practically significant finding of the exploratory analysis. Psychosocial-only k-means clustering ($k=2$, confirmed by elbow method and silhouette analysis) identified a High Burden cluster ($n=375$, actual failure rate 41.6%) and a Low Burden cluster ($n=498$, failure rate 16.5%), with a highly significant cluster-outcome association ($\chi^2=79.76$, $p=4.2\times 10^{-19}$). Critically, as illustrated in Figure 3, 219 students in the High Burden cluster are currently passing (mean mark 62.7) — entirely invisible to marks-only monitoring. This group represents 25.1% of the total student population and 71.5% of the High Burden cluster, demonstrating the scale of latent undetected risk in this population.

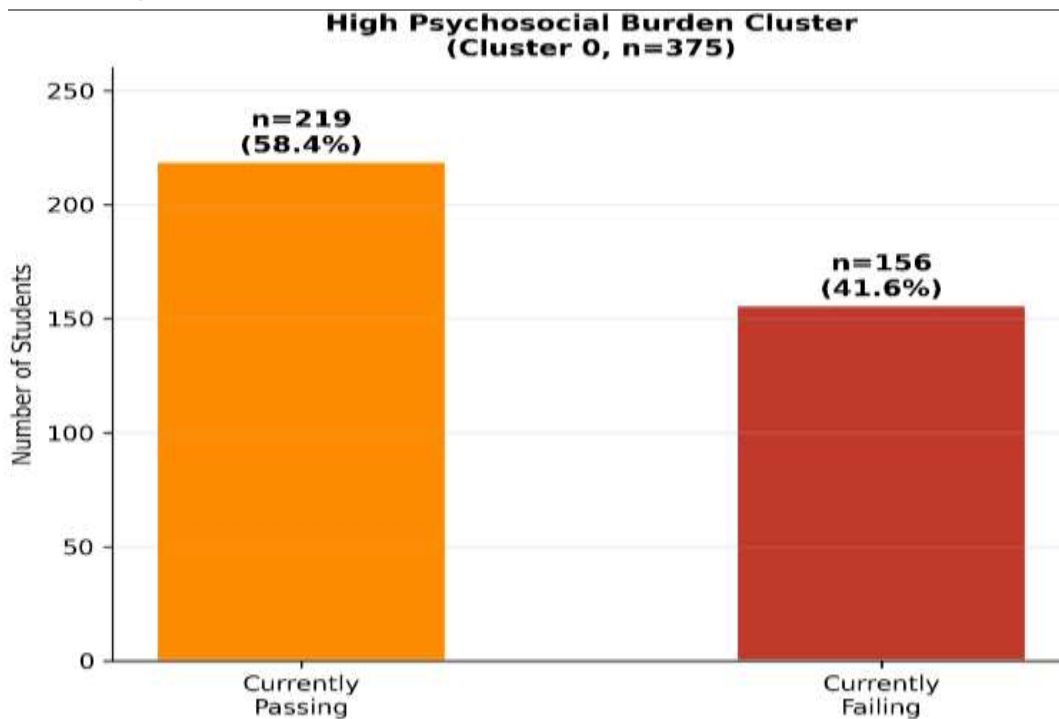


Figure 3 (Left). Hidden risk population: 219 students with above-median psychosocial burden currently passing, identified through psychosocial-only k-means clustering ($\chi^2=79.76$, $p=4.2 \times 10^{-19}$)

Figure 3 (Right). Subject-level failure cascade: co-occurrence rates of Sinhala, English, and Science failures with Mathematics failure, confirming that academic difficulty is systemic rather than subject-isolated (n=873).

B. Predictive Model Performance

Table 3 presents the 10-fold stratified cross-validation results for all four candidate models. Logistic Regression was selected as the final model. The Stacking Ensemble failed the paired t-test ($p=0.469$); Random Forest ($p=0.008$) and XGBoost ($p=0.032$) achieved statistical significance but did not exceed the 2-point AUC threshold. No challenger met both criteria simultaneously.

Table 3. Model comparison — 10-fold stratified cross-validation results

Model	CV AUC	95% CI	Sensitivity	Specificity	F1	Selected
Logistic Regression	0.878	[0.858–0.899]	0.803	0.848	0.727	✓
Random Forest	0.856	[0.831–0.882]	0.811	0.818	0.707	—
XGBoost	0.850	[0.825–0.876]	0.802	0.808	0.700	—
Stacking Ensemble	0.877	[0.854–0.899]	0.841	0.808	0.716	—

The AUC progression is shown in Figure 4. The academic-only baseline model achieved CV AUC=0.818 [0.785–0.850]. Adding psychosocial composite features produced the full model at CV AUC=0.878 — a validated gain of +6.1 AUC points stable across all 10 folds, confirming that marks alone are quantitatively insufficient for early risk detection. Likelihood Ratio Tests confirmed independent contributions from all three primary psychosocial domains: Socioeconomic $\chi^2=64.20$, Cognitive Load $\chi^2=30.93$, Motivation $\chi^2=14.06$, all $p<0.001$. The full model achieved NPV=0.923 [0.901–0.944] and Brier Score=0.140, confirming strong calibration. The Stacking Ensemble’s virtually identical performance to Logistic Regression ($p=0.469$) provides evidence that the risk-outcome relationship is approximately linear in the engineered feature space, supporting the theoretical appropriateness of Logistic Regression beyond its interpretability advantages.

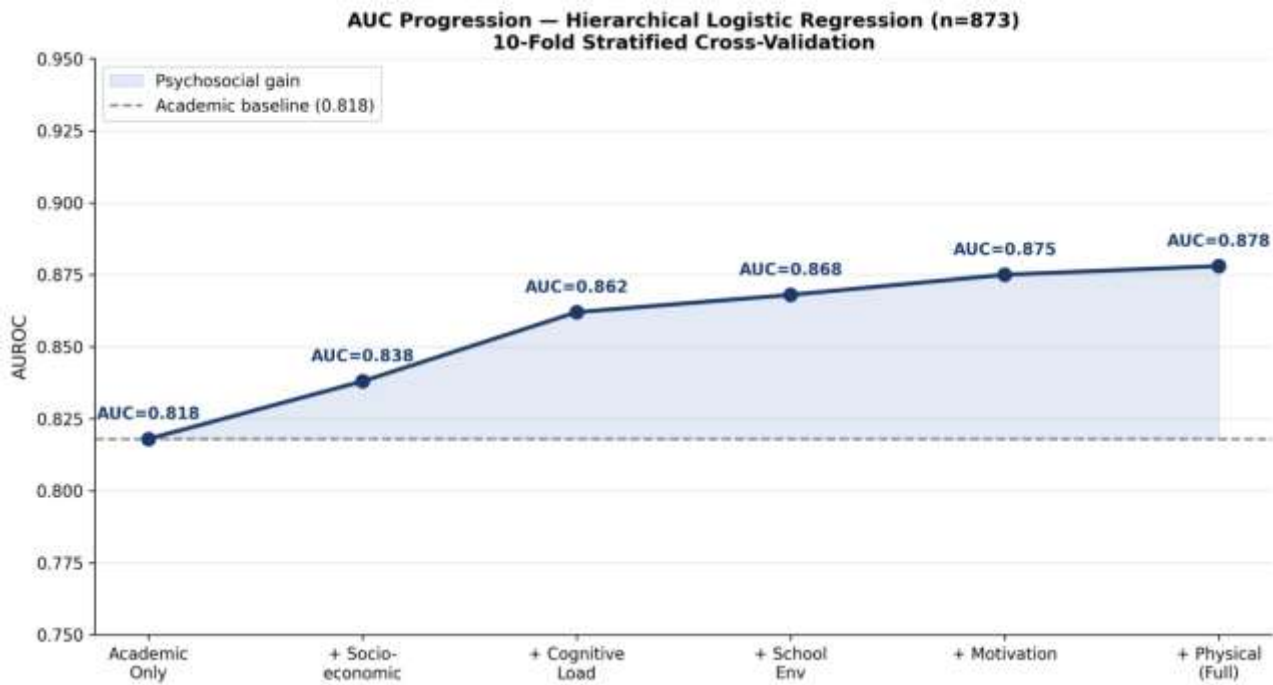


Figure 4. AUC progression from academic-only baseline (0.818) to full model (0.878) as each psychosocial domain is added incrementally — 10-fold stratified CV, n=873

The four significant multivariate predictors were: Recent Average Term Mark (OR=0.186, −81.4% odds per SD, $p<0.001$), Motivation (OR=1.642, +64.2%, $p<0.001$), Cognitive Load (OR=1.576, +57.6%, $p<0.001$), and Socioeconomic Burden (OR=1.337, +33.7%, $p<0.05$). A student with average marks but high Cognitive Load faces +57.6% elevated odds of failure — evidence that psychosocial factors do not merely add to academic risk but interact with it in ways that marks-only monitoring cannot detect. The Likelihood Ratio Tests confirmed independent domain contributions: Socioeconomic $\chi^2=64.20$, Cognitive Load $\chi^2=30.93$, Motivation $\chi^2=14.06$, all $p<0.001$ — confirming that each psychosocial domain adds statistically significant predictive information beyond what all other features already explain.

The 10-fold CV confusion matrix, shown in Figure 5, confirmed: TN=538, FP=97, FN=47, TP=191. This yields sensitivity=0.803 (191 of 238 failing students detected), specificity=0.847 (538 of 635 passing students correctly cleared), and NPV=0.920 (92.0% of cleared students genuinely passed). Risk tier assignment (Low: $p<0.30$, Moderate: 0.30–0.60, High: $p>0.60$) produced a greater than 10× failure rate gradient: Low 6.3%, Moderate 19.8%, High 66.9%. Of 238 actual failures, 73.1% were captured in the High tier, 15.5% in Moderate, and only 11.3% in the Low tier — meaning 88.6% of all actual failures were detected at Moderate or High tier level.

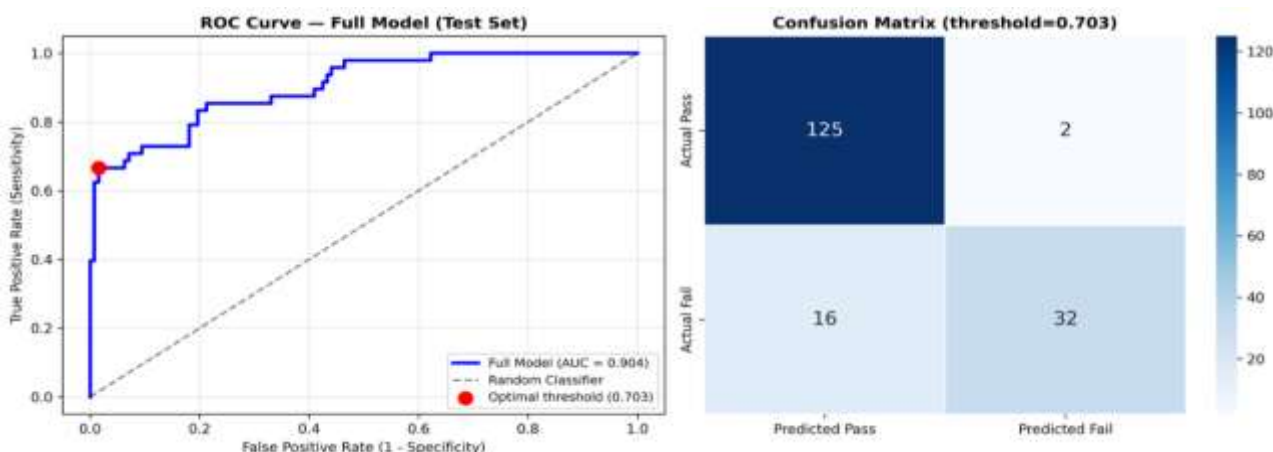


Figure 5. 10-fold CV confusion matrix (per-fold Youden threshold) and actual failure rate by risk tier — confirming the system captures 88.6% of all actual failures at Moderate or High tier

SHAP Explainability: Which Psychosocial Factors Drive Risk?

SHAP LinearExplainer (zero-approximation-error for linear models) was applied to the final model trained on all 873 students. The base value was -0.6988 log-odds (probability=0.3321), close to the actual failure rate of 27.3%, confirming model calibration. Table 4 presents the global SHAP feature importance rankings.

Table 4. Global SHAP feature importance — all 873 students (SHAP LinearExplainer, zero approximation error)

Rank	Feature	Mean SHAP	Type	LR Significant	Consistent
1	Recent Avg Term Mark	1.3784	Academic	Yes ($p<0.001$)	✓
2	Cognitive Load	0.5084	Psychosocial	Yes ($p<0.001$)	✓
3	Motivation	0.3840	Psychosocial	Yes ($p<0.001$)	✓
4	Socioeconomic Burden	0.1633	Psychosocial	Yes ($p<0.05$)	✓
5	School Environment	0.1514	Psychosocial	No	✓
6	Academic Variability	0.1227	Academic	No	✓
7	Physical Readiness	0.0614	Psychosocial	No	✓
8	Mark Decline Rate	0.0475	Academic	No	✓

The global SHAP feature importance rankings are presented in Table 4 and visualised in the beeswarm plot (Figure 6). Full consistency confirmed: all four LR-significant predictors rank in SHAP positions 1–4; all four non-significant predictors rank 5–8. The SHAP analysis reveals a two-tier risk driver structure with direct practical implications. Recent Average Term Mark dominates decisively ($\text{Mean|SHAP|}=1.3784$, $2.7\times$ the next feature), but this represents the diagnostic signal — teachers cannot change a student’s past marks. Cognitive Load (0.5084), Motivation (0.3840), and Socioeconomic Burden (0.1633) constitute the actionable tier — these are the psychosocial dimensions teachers can directly address through targeted intervention. The distinction between what diagnoses risk and what can be acted upon is precisely what marks-only monitoring cannot provide.

The shift in dominant drivers across risk tiers is among the most important practical findings. In High Risk students, Recent Average Term Mark dominates in 76.9% of cases — the academic failure signal is already present and marks-based monitoring would have flagged these students. In Moderate Risk students, Cognitive Load rises to become the dominant driver in 29.5% of cases, and psychosocial burden is the primary signal before academic marks have declined sufficiently to trigger conventional concern. EduRisk adds unique, non-duplicated value precisely in the Moderate tier — the early warning zone where marks-based monitoring is blindest.

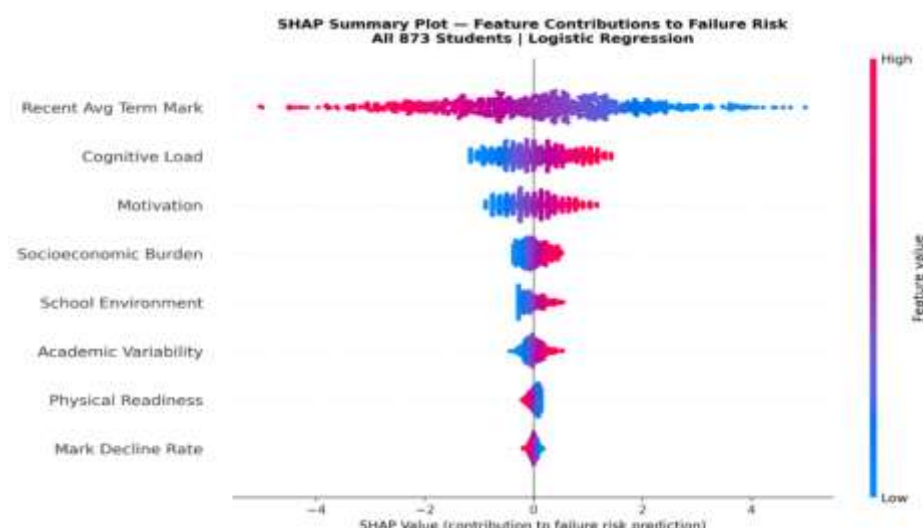


Figure 6. SHAP beeswarm summary plot: feature contributions to failure risk across all 873 students, ordered by mean absolute SHAP value

Three case studies illustrate the system's explanatory power across the risk spectrum. Pre-specified selection criteria were defined before any individual was examined to prevent cherry-picking.

Case 1 — STU0181 (High Risk, FAIL, probability=0.904): Below-average marks (Recent Avg=53.8) contributed +1.83 toward failure risk, amplified by high Cognitive Load (score=2.09, +0.77) and low Motivation (score=0.84, +0.28). This student would be identified by both marks-based and psychosocial monitoring, but SHAP specifies exactly which psychosocial domains require intervention — information that marks alone cannot provide.

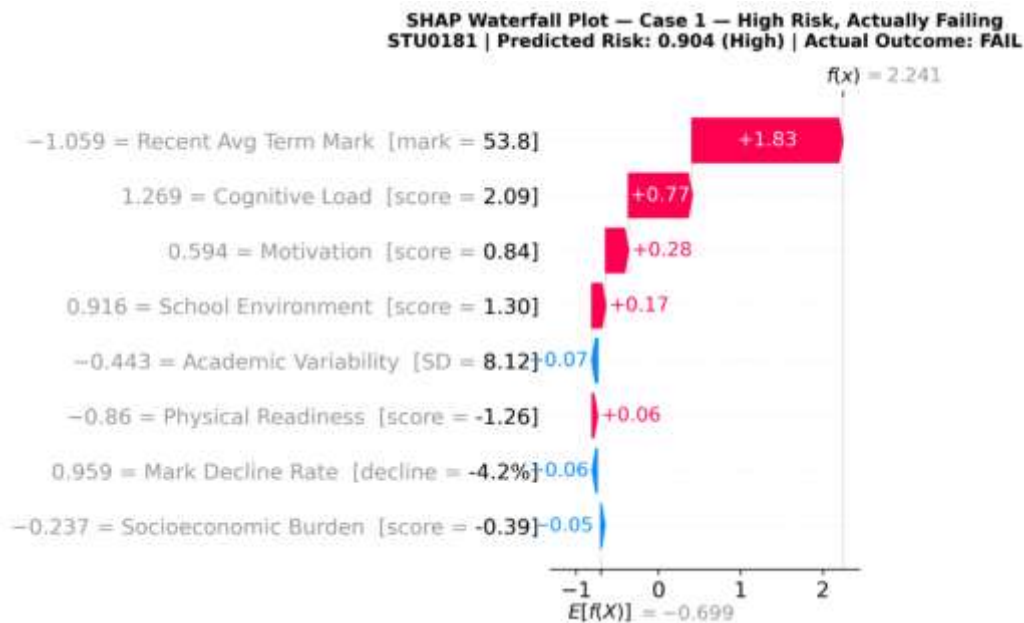


Figure 7. SHAP waterfall plot - STU0181 (High Risk, FAIL, probability=0.904): marks and psychosocial burden jointly drive risk

Case 2 — STU0358 (Hidden Risk, PASS, probability=0.377): Strong mark-based protection (Recent Avg=65.8, contributing -1.42) was almost entirely counteracted by extreme Cognitive Load (score=3.13, +1.16) and School Environment at the absolute dataset maximum (score=4.13, +0.54). This student would raise zero concern under marks-only monitoring. EduRisk identifies the latent psychosocial crisis building beneath apparently adequate academic performance — the definitive demonstration of what marks-only monitoring structurally misses.

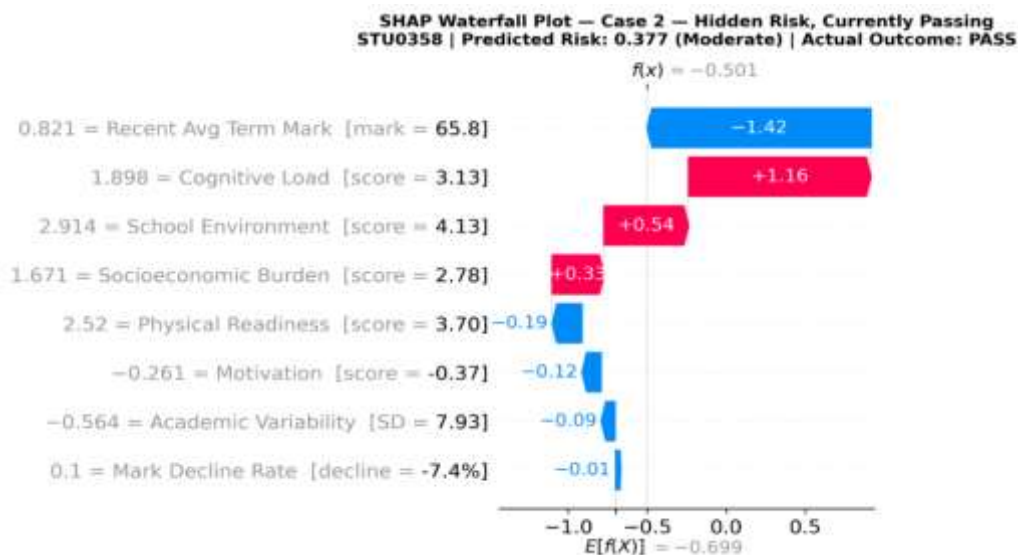


Figure 8. SHAP waterfall plot — STU0358 (Hidden Risk, PASS, probability=0.377): good marks almost cancelled by extreme psychosocial burden — invisible to marks-only monitoring

Case 3 — STU0252 (Low Risk, PASS, probability=0.115): Despite below-average marks (55.8), minimum Cognitive Load (−3.14) and strong Motivation (−2.29) provided overwhelming psychosocial protection. This case demonstrates that psychosocial resilience can override an academic risk signal — and that marks-only monitoring would have generated unnecessary concern about a student who is genuinely protected by their psychosocial profile.

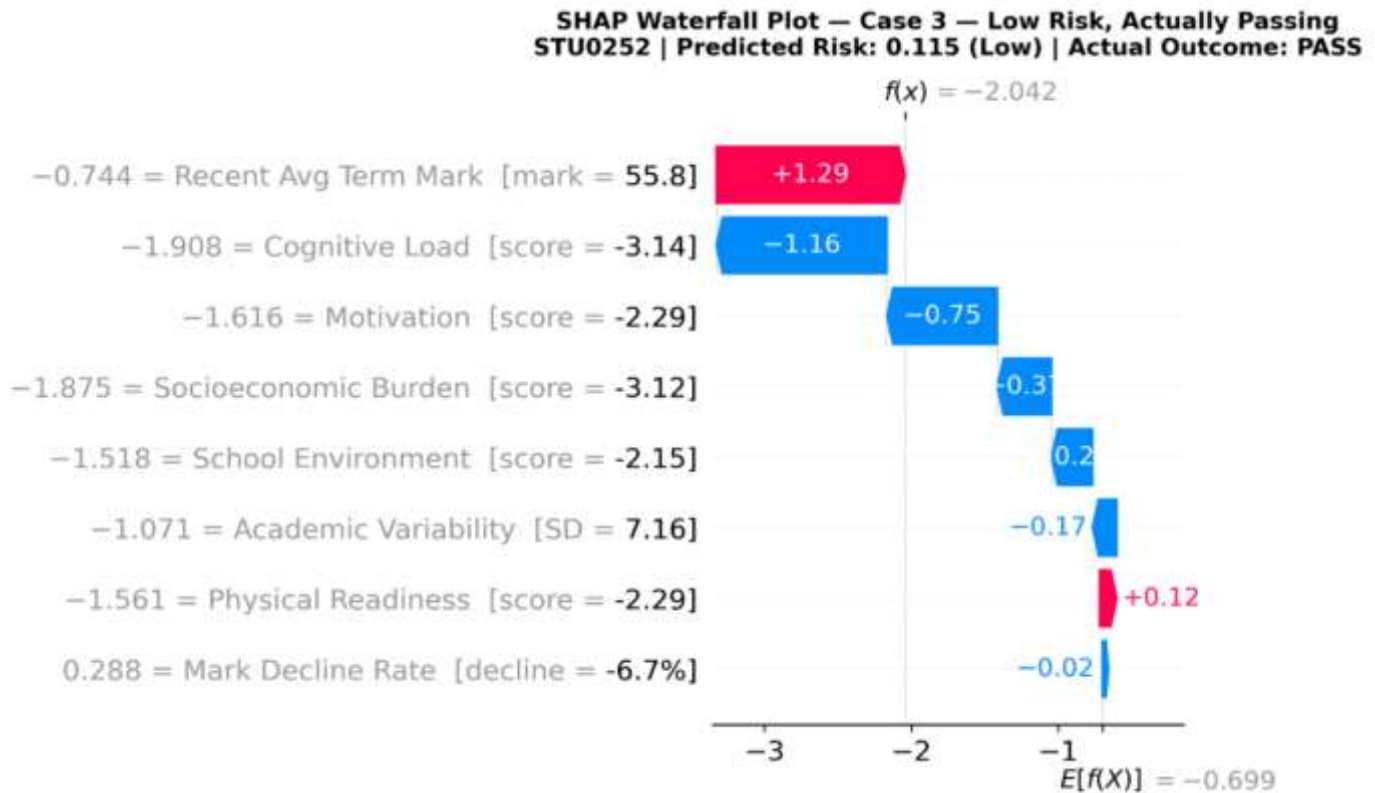


Figure 9. SHAP waterfall plot — STU0252 (Low Risk, PASS, probability=0.115): psychosocial resilience overrides below-average marks

D. Gender Fairness Analysis

Four independent statistical tests confirmed gender-equitable model performance across all dimensions. Failure rate difference: female 29.1% vs male 25.6% ($\chi^2=1.164$, $p=0.2807$, not significant). AUC difference: female 0.903 vs male 0.843 (Mann-Whitney $U=67.0$, $p=0.2121$, not significant). Risk tier distribution: $\chi^2=5.880$, $p=0.0529$, not significant. SHAP feature importance: 0/8 features showed significant distributional differences across genders; rank difference was zero for all eight features. This last finding is particularly strong evidence of equity: the model applies identical risk factor weighting regardless of student gender. A marginal pattern of higher High Risk tier assignment for female students (33.7% vs 26.3%, $p=0.0529$) is directionally consistent with the higher female failure rate (29.1% vs 25.6%) and reflects the model responding to real distributional differences rather than introducing artificial disparity. Figure 10 presents the three-panel fairness summary.

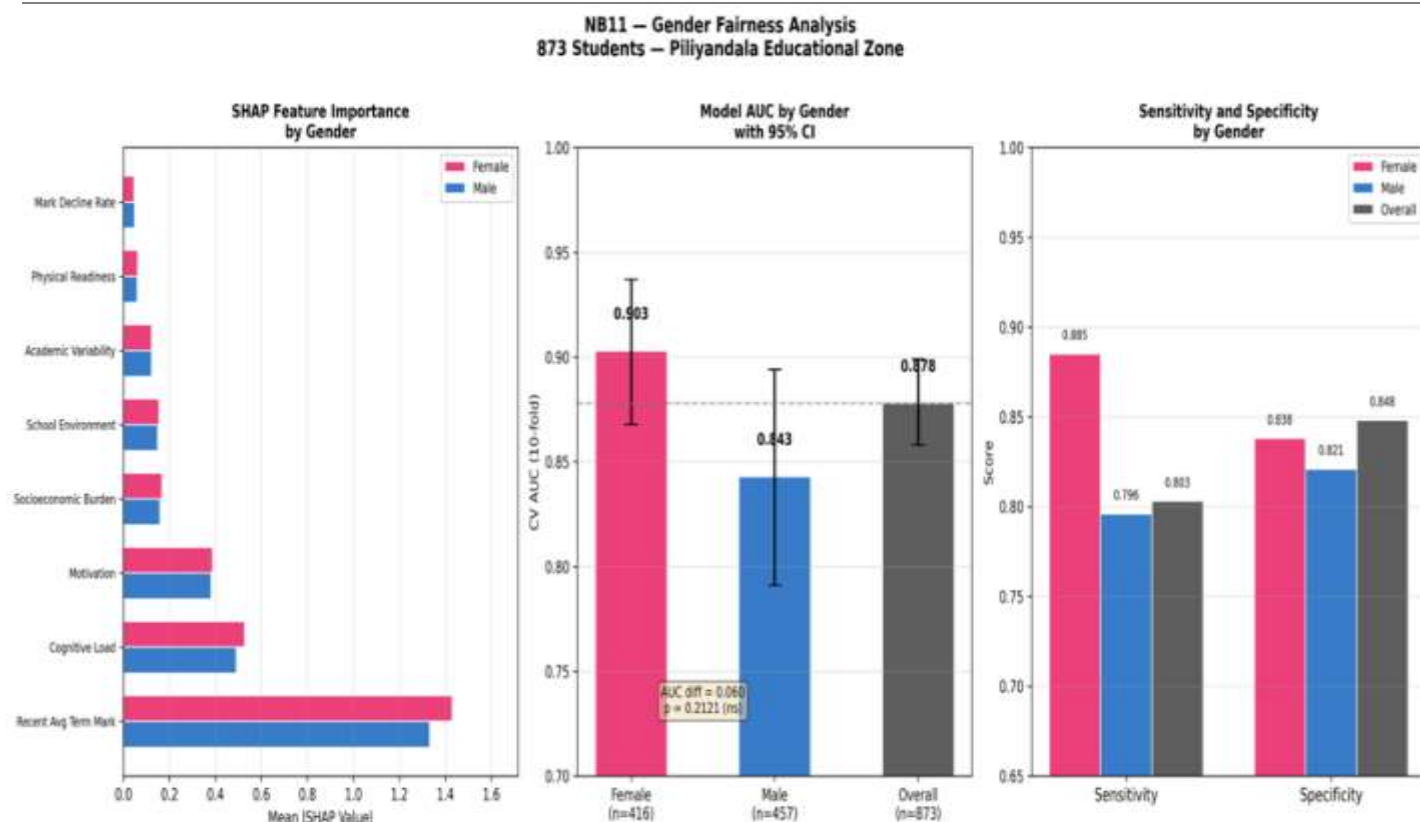


Figure 10. Gender fairness analysis: SHAP feature importance by gender (left), AUC comparison with 95% CI (centre), sensitivity and specificity by gender (right) — all four statistical tests non-significant

E. EduRisk Prototype

System Architecture

The EduRisk system architecture follows a three-layer design. The Data Layer handles ingestion of two data types: academic mark records transcribed from school registers and psychosocial questionnaire responses entered via the system's administrator module. A preprocessing pipeline applies the same StandardScaler objects and PCA transformers fitted during training — ensuring consistency between training-time feature computation and deployment-time feature computation without refitting on new data. The Intelligence Layer applies the pre-trained Logistic Regression model and SHAP LinearExplainer to produce, for each student: a predicted failure probability, a risk tier assignment (Low/Moderate/High), and a full vector of SHAP feature contributions. The intervention mapping module cross-references each student's SHAP contributions against the PPCA-validated 14-row intervention table using domain-specific P50 and P75 thresholds to determine which intervention rows are triggered. The Presentation Layer delivers role-specific views through a CustomTkinter desktop interface: the teacher sees only their assigned class, the principal sees the whole school with grade-level navigation, and the administrator manages data entry and system configuration. The offline mode renders validated intervention guidance directly from the table; the online RAG mode sends triggered rows to Groq API (Llama 3.3 70B) which synthesises personalised natural-language guidance grounded exclusively in the validated knowledge base. Figure 11 presents the complete system architecture diagram.

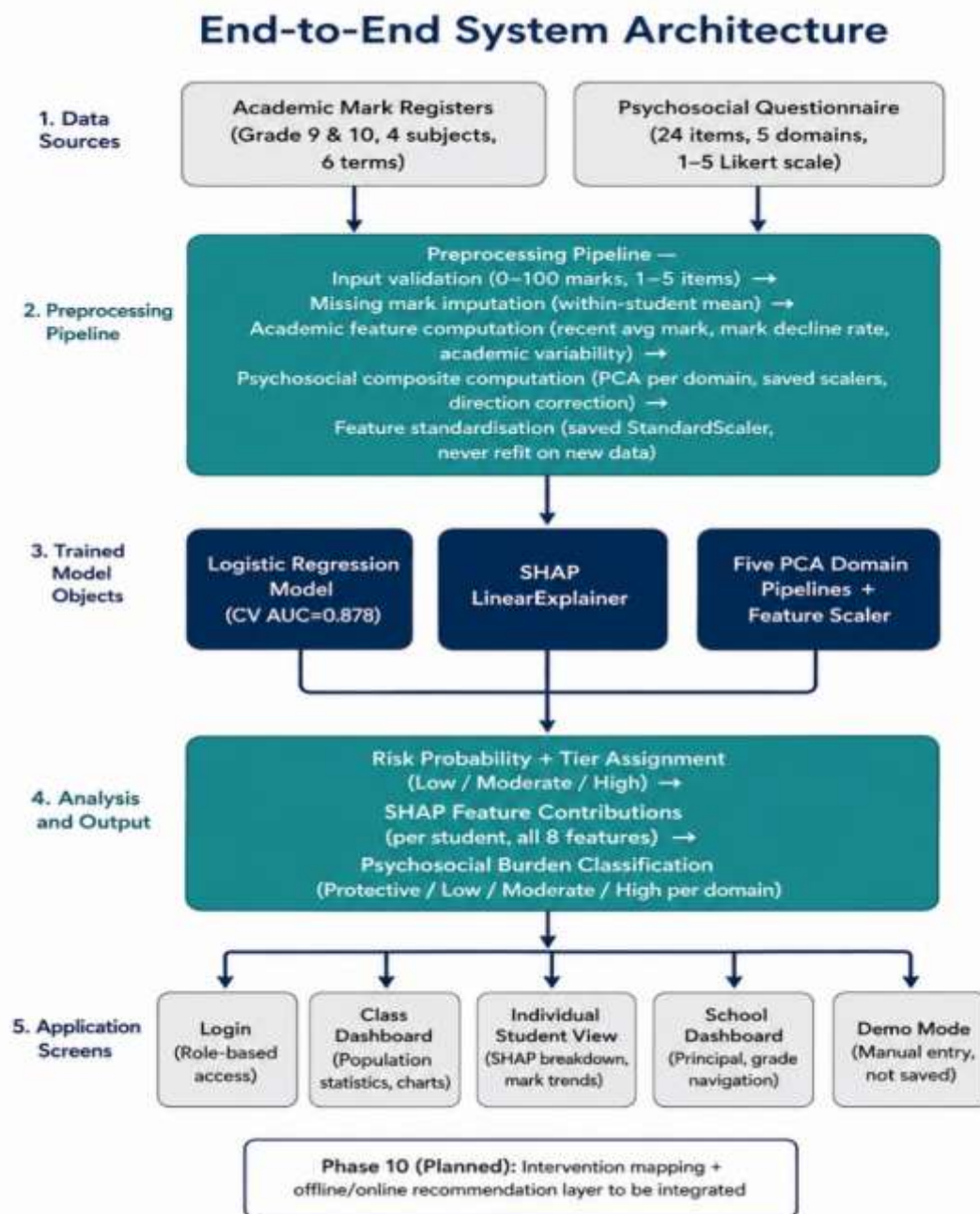


Figure 11. EduRisk system architecture: three-layer design from data ingestion through intelligence layer to role-based presentation, supporting offline and online RAG-based recommendation modes.

The EduRisk prototype operationalises the complete research pipeline as a deployable standalone desktop application built with CustomTkinter 6.0.0 (GUI), SQLite (database), scikit-learn and SHAP (inference), Matplotlib (visualisation), and the Groq API with Llama 3.3 70B (online recommendation layer). Three user roles are implemented with role-based access control: class teachers see only their assigned class, principals see the whole school with grade-level navigation, and administrators manage system configuration. This architecture mirrors how Sri Lankan school authority actually operates, increasing the likelihood of real-world adoption.

The system was validated on 128 held-out students from three Grade 11 classes (11A: 43, 11B: 46, 11C: 39) from a single Piliyandala zone school — data never used in any stage of model training, validation, or testing. The system processed all 128 students through the complete pipeline (preprocessing → feature computation → prediction → SHAP) in under one second, producing risk probabilities ranging from 0.001 to 0.997, tier assignments, and SHAP feature breakdowns directly readable by teachers with no data science background.

Figure 12 shows the class dashboard, providing population-level insights never previously available to Piliyandala zone teachers: risk tier distribution across the class, psychosocial domain averages, and six-term subject mark trends. The individual student risk profile view (Figure 13) displays predicted risk probability and tier prominently, alongside a full SHAP feature-contribution breakdown with each feature's raw value and directional contribution toward or away from failure risk. A validated 14-row, 7-domain, dual-severity intervention mapping table was developed from the questionnaire items, supplemented by educational psychology literature, and refined through structured consultation with school principals and the PPCA. It specifies observable signs, recommended actions, responsible parties, and urgency levels for each domain-severity combination. The offline recommendation mode delivers this validated intervention guidance directly within the system without internet connectivity. The online mode implements a Retrieval-Augmented Generation (RAG) architecture using the Groq API with Llama 3.3 70B — retrieving only the triggered intervention rows for that student and synthesising personalised, natural-language guidance grounded exclusively in the validated table content. The LLM cannot generate recommendations outside the expert-validated knowledge base, ensuring all guidance is faithful to the practitioner-reviewed intervention framework. The complete system was demonstrated to five Piliyandala Educational Zone school principals and the research supervisory team, confirming operational feasibility and practitioner accessibility. The school dashboard available to principals is shown in Figure 14. Figure 15 illustrates the online RAG-based recommendation output for a High Risk student, demonstrating the system's capacity to synthesise personalised, actionable guidance from the validated intervention knowledge base.

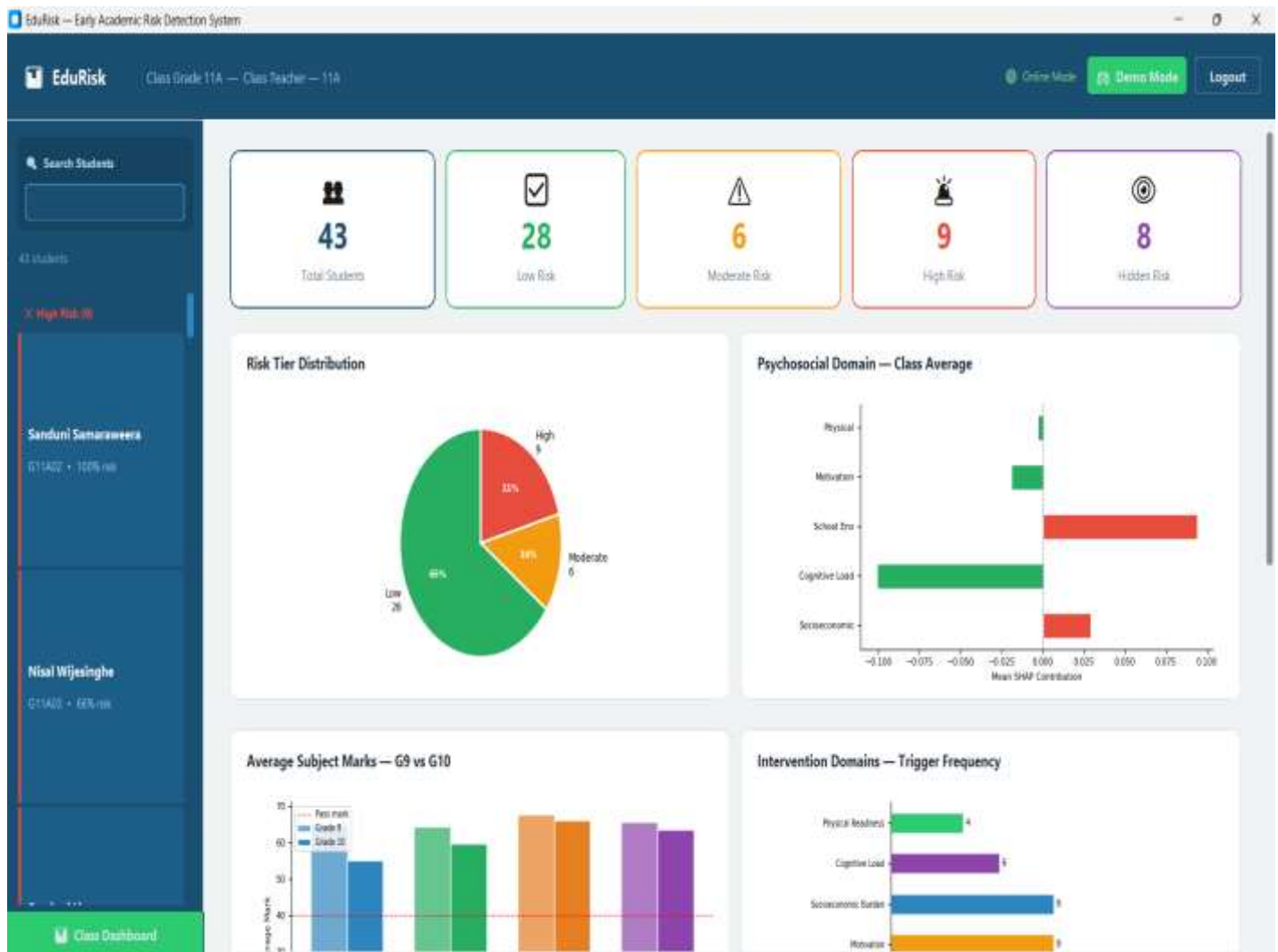


Figure 12. EduRisk class dashboard (teacher view): risk tier distribution, psychosocial domain averages, six-term mark trends, and priority student sidebar for Class 11A



Figure 13.1 EduRisk individual student risk profile: predicted risk probability and tier, SHAP feature-contribution breakdown with raw values

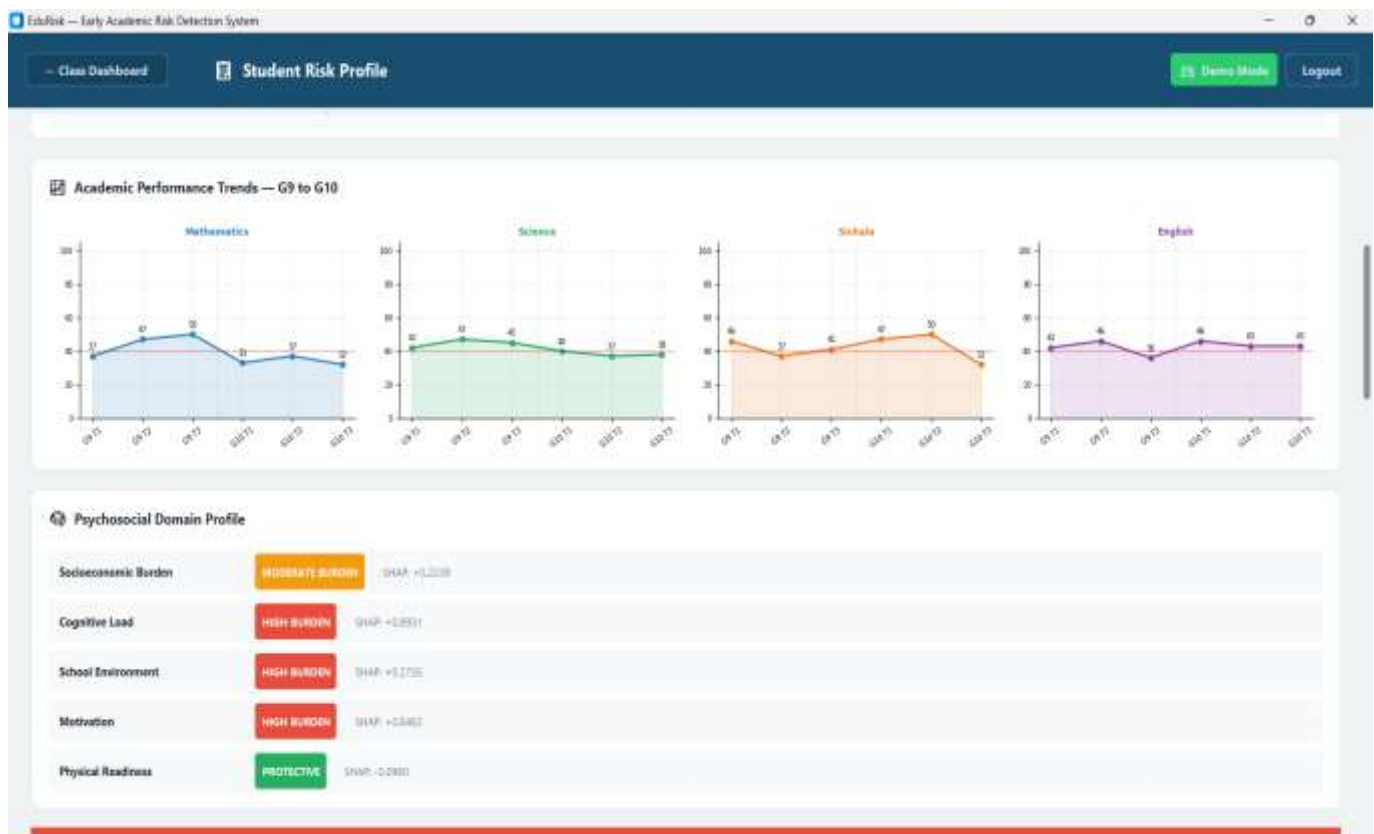


Figure 13.2 EduRisk individual student risk profile: academic performance trends and psychosocial domain burden classification

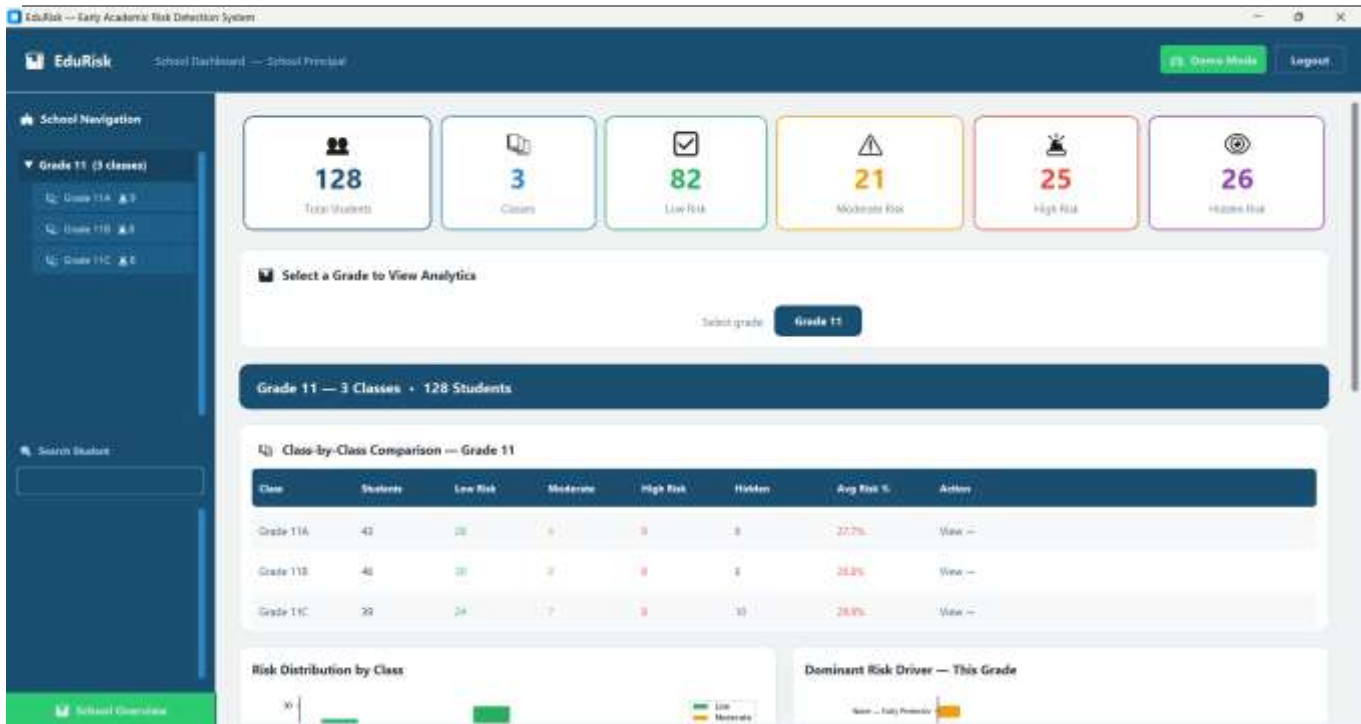


Figure 14. EduRisk school dashboard (principal view): whole-school risk distribution, grade-level navigation, and cross-class comparison.

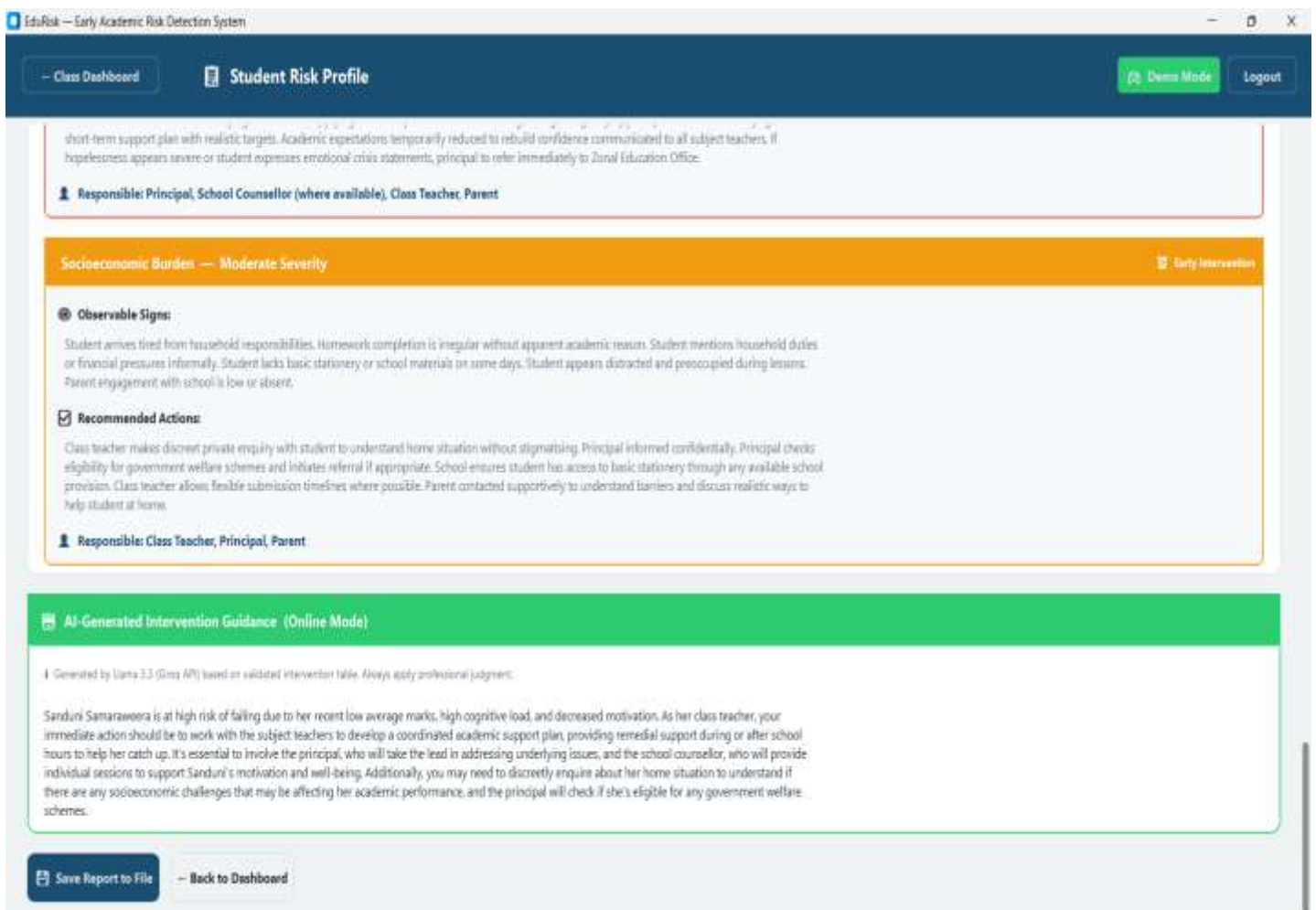


Figure 15. EduRisk online recommendation mode: RAG-based personalised natural-language intervention guidance grounded in the expert-validated 14-row intervention table (Groq API, Llama 3.3 70B).

DISCUSSION

This study produces three principal findings that collectively establish a new empirical foundation for early academic risk detection in Sri Lankan government secondary schools.

The first and most fundamental finding is the quantitative inadequacy of marks-only monitoring. The +6.1 AUC point improvement from the academic-only baseline (0.818) to the full psychosocial model (0.878), stable across all 10 cross-validation folds and corroborated by Likelihood Ratio Tests confirming independent domain contributions, provides unambiguous quantitative evidence that marks alone are insufficient for early risk detection in this population. This aligns with Berens et al. (2019) and Vinas-Forcade et al. (2021), who found multimodal systems consistently outperform single-modality approaches, and extends these findings into a no-LMS context where psychosocial signals must be captured through a validated questionnaire rather than digital traces.

The second finding — the hidden risk phenomenon — carries the greatest immediate policy significance. That 219 students (25.1% of the population) carry above-median psychosocial burden while currently passing, entirely invisible to conventional monitoring, demonstrates that the marks-based system is not merely slow to detect risk — it structurally cannot detect the form of risk most amenable to early intervention. The dose-response relationship (7.1% to 58.2% failure rate across the psychosocial burden spectrum, $RR=8.1\times$) establishes that psychosocial burden is dose-dependently predictive of failure, and that students in the hidden risk population are not borderline cases but genuinely at elevated risk that will manifest as failure if unaddressed. Dai et al. (2025) argued that EWS must generate actionable guidance rather than merely risk scores — for hidden risk students, the system must first make the risk visible before any guidance is possible, which is precisely what the psychosocial measurement pipeline achieves.

The third finding concerns the actionable structure of risk drivers identified through SHAP. The two-tier structure — marks as the dominant diagnostic signal, Cognitive Load and Motivation as the dominant actionable signals — has direct implications for how teachers should interpret and respond to EduRisk outputs. A teacher cannot change a student's past marks, but can address cognitive burden through adjusted workload and academic support, motivational disengagement through counselling and encouragement, and socioeconomic constraints through referral to school welfare services. The shift in dominant drivers across risk tiers — with Cognitive Load rising to 29.5% in the Moderate tier — confirms that EduRisk's greatest value is in the early warning zone where marks-only monitoring is completely blind.

The gender fairness results address a genuine concern in algorithmic decision-making: that models trained on historical data may replicate structural inequities. The identical SHAP feature rankings across genders (zero rank difference for all eight features) provides the strongest possible evidence that the model does not apply different risk factor weighting based on gender. The marginal female High Risk assignment pattern (33.7% vs 26.3%, $p=0.0529$) is monitored rather than corrected, as it is directionally consistent with the real female failure rate differential and reflects model responsiveness to actual distributional differences.

Three limitations are acknowledged. Temporal proximity: the psychosocial questionnaire was administered at Grade 11 Term 3, contemporaneous with the outcome period, limiting the prospective early detection claim. Prospective field validation is planned, applying the same measurement approach to a new Grade 11 cohort and tracking their third term outcomes on genuinely unseen students. Single zone: findings are validated for Piliyandala zone and generalisability to rural, estate sector, or Northern Province schools is not established. A replicability assessment is planned to evaluate the theoretical adaptations required for other contexts. Cross-sectional psychosocial measurement: one questionnaire administration per student captures current state but not psychosocial trajectory — whether burden is increasing or decreasing — which longitudinal measurement would capture.

CONCLUSION

This study presents the first validated, explainable, and deployable early academic risk detection system for Sri Lankan government secondary schools, demonstrating that marks-only monitoring — the prevailing approach — is quantitatively insufficient for identifying students at genuine academic risk. The integrated model achieves a validated 6.1-point AUC improvement over the academic-only baseline (0.878 vs 0.818), confirming that psychosocial factors contribute independent and statistically significant predictive value beyond what marks alone reveal. Critically, 219 students — 25.1% of the study population — were identified carrying above-median psychosocial burden while currently passing on marks, entirely invisible to conventional monitoring. These are students accumulating the conditions for future academic failure while the marks-only system shows no cause for concern.

The EduRisk prototype operationalises these findings into a complete early warning and decision support system that delivers SHAP-grounded, student-level risk profiles covering the psychosocial dimensions that marks-only monitoring structurally cannot detect. A teacher consulting EduRisk knows not only which students are at risk and at what level, but which specific psychosocial domains are elevated, how they interact with academic performance, and — through the integrated 14-row expert-validated intervention table — exactly what observable signs to look for, what actions to take, who else to involve, and with what urgency. The offline mode delivers this guidance without internet dependency, making the system fully functional in schools with no connectivity. The online RAG-based mode synthesises personalised, natural-language intervention guidance grounded exclusively in the validated practitioner-reviewed knowledge base. The system requires only a validated 24-item questionnaire, existing school mark registers, and a single school computer, making it deployable within existing Sri Lankan government school resources without capital investment in digital infrastructure.

The research directly addresses the gap formally acknowledged by the Ministry of Education Research and Development Branch and contributes to Sustainable Development Goal 4 by providing an evidence-based, equity-conscious, and practically deployable framework for early academic risk detection. The proxy-indicator methodology is context-independent in its architecture and replicable for resource-constrained educational contexts globally.

ACKNOWLEDGEMENTS

The authors thank the Ministry of Education Research and Development Branch, the Western Province Education Ministry, the Piliyandala Zonal Education Ministry, and the Professional Psychological Counselors Association of Sri Lanka for institutional support and expert validation. The authors are grateful to the 12 school principals and 45 classroom teachers across 11 Piliyandala zone schools for their cooperation and field guidance, and to the 873 students whose participation made this research possible.

REFERENCES

1. Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk — predicting student dropouts using administrative student data and machine learning methods. *Administrative Sciences*, 9(2), 1–41. <https://doi.org/10.3390/admsci9020042>.
2. Choi, W. C., Lam, C. T., Pang, P. C., & Mendes, A. J. (2025). A systematic literature review of explainable artificial intelligence (XAI) for interpreting student performance prediction in computer science and STEM education. *Proceedings of the ACM*. <https://doi.org/10.1145/3724363.3729027>
3. Dai, W., Lin, J., Jin, F., Tsai, Y., Srivastava, N., Le Bodic, P., Gasevic, D., & Chen, G. (2025). Learning analytics for early identification of at-risk students and feedback intervention. *Journal of Learning Analytics*, 12(1). <https://doi.org/10.18608/jla.2025.8735>
4. Dayasiri, K., Jayasingha, D., & Thadchanamoorthy, V. (2025). Determinants of school drop-out among adolescents in rural Sri Lanka. *Children and Youth Services Review*, 169. <https://doi.org/10.1016/j.childyouth.2024.108112>

5. Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
6. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
7. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson.
8. Harrell, F. E. (2015). *Regression Modeling Strategies* (2nd ed.). Springer.
9. He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
10. Karalar, H., Kapucu, C., & Gürüler, H. (2021). Predicting students at risk of academic failure using ensemble model in distance learning. *Education and Information Technologies*, 26(4), 4031–4041.
11. Kotsiantis, S. B., Pierrakeas, C., & Pintelas, P. (2004). Predicting student performance in distance learning using machine learning techniques. *Applied Artificial Intelligence*, 18(5), 411–426.
12. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
13. Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. *Computers and Education*, 54(2), 588–599.
14. Mudunna, C., Perera, S., & Rajapakshe, A. (2025). Prevalence and determinants of mental health problems experienced by school-going adolescents in Sri Lanka. *Global Mental Health*, 12, e55.
15. Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373–1379.
16. Rumberger, R. W. (2011). *Dropping Out*. Harvard University Press.
17. Trakunphutthirak, R., & Lee, V. C. S. (2022). Application of educational data mining approach for student academic performance prediction using progressive temporal data. *Journal of Educational Computing Research*, 60(2), 408–447.
18. UNESCO. (2021). *Situation analysis on the effects of and responses to COVID-19 on the education sector in Asia: Sri Lanka case study*. Paris: UNESCO.
19. Vinas-Forcade, J., Mels, C., Van Houtte, M., Valcke, M., & Derluyn, I. (2021). Can failure be prevented? Using longitudinal data to identify at-risk students upon entering secondary school. *British Educational Research Journal*, 47(1), 205–225. <https://doi.org/10.1002/berj.3683>
20. Zhang, S., Li, Y., Chen, X., Liu, Y., & Wang, H. (2025). A systematic review of learning management system-based early warning systems in higher education 2012–2024. *Computers and Education: Artificial Intelligence*, 8. <https://doi.org/10.1016/j.caeai.2025.100345>