

RESEARCH

Open Access



Explainable machine learning analysis of academic resilience among disadvantaged students in Türkiye using Programme for International Student Assessment 2022 data

Mehmethan Erduran¹ , Elif Buğra Kuzu Demir² and Kadir Demir^{1*}

*Correspondence:

Kadir Demir

kadir.demir@idu.edu.tr

¹Department of Management Information Systems, İzmir Demokrasi University, İzmir, Türkiye

²Department of Computer and Instructional Technologies Education, Dokuz Eylül University, İzmir, Türkiye

Abstract

Academic resilience is a context-specific pattern of positive academic adaptation among students facing socioeconomic disadvantage. This study examined academic resilience in Türkiye using 2022 data from the Programme for International Student Assessment (PISA). Students in the weighted bottom quartile of the national Index of Economic, Social and Cultural Status distribution formed the analytic sample ($n = 1885$). For each of the ten PISA mathematics plausible values (PVs), students reaching the corresponding weighted national top quartile were classified as resilient, yielding 204 to 230 resilient students across PV-specific replications. The survey-aware workflow used student weights, school-grouped validation, training-only preprocessing, and minority-class threshold tuning. Weighted logistic regression, Radial Basis Function Support Vector Machine (RBF-SVM), ExtraTrees, Extreme Gradient Boosting (XGBoost), and CatBoost were benchmarked. Although the primary PV-aware comparison favoured CatBoost on Precision-Recall Area Under the Curve (PR-AUC), a three-repeat school-grouped sensitivity analysis found the highest mean PR-AUC for ExtraTrees (0.332, 95% confidence interval [CI] 0.318 to 0.348); logistic regression obtained 0.302 (95% CI 0.241 to 0.340) and CatBoost 0.296 (95% CI 0.267 to 0.316). CatBoost was retained for transparent SHapley Additive exPlanations (SHAP) rather than as a uniquely optimal classifier. The results indicate a limited but detectable group-level signal. Information and Communication Technology (ICT) resources, digital self-efficacy, subject-related ICT use, mathematics anxiety, gender patterning, and digital-emotional interactions jointly characterized the modelled resilience profile. These associations are descriptive and non-causal.

Keywords Academic resilience, PISA 2022, Türkiye, SVM, CatBoost, SHAP, Educational inequality



1 Introduction

Educational inequality is often summarized through average differences between advantaged and disadvantaged students, but this leaves an important question: under what conditions do some disadvantaged students still attain strong academic outcomes? Academic resilience refers to positive academic adaptation despite social, economic, or educational risk [1, 2]. In Programme for International Student Assessment (PISA) research, it is operationalized through the combination of low socioeconomic status and high achievement, while remaining a context-specific classification rather than a fixed personal trait [3–5]. PISA 2022 is especially relevant because it captures learning after pandemic disruption, when unequal digital learning support became more visible [6, 7].

Türkiye provides a strong case for studying academic resilience. The Organisation for Economic Co-operation and Development (OECD) PISA 2022 Türkiye country note reports that approximately 12% of disadvantaged students reached the top national quarter of mathematics performance (see the country note section ‘Performance gaps within Türkiye’, ‘Socio-economic divides’); this is consistent with the 11.7% weighted prevalence observed in the present analytic cohort [8]. At the same time, inequalities in learning opportunities remain substantial, making the Turkish case scientifically and policy relevant.

Previous research identifies psychological conditions, family and school support, and learning opportunities as relevant to academic resilience [2, 4, 5, 9–12]. Digital access, digital self-efficacy, and educational uses of Information and Communication Technology (ICT) may be especially relevant after the pandemic, although their contribution is unlikely to be uniform across students or forms of use [7, 13].

Descriptive and regression approaches remain valuable for estimating specified associations, but academic resilience may also involve non-linear and conditional combinations of emotional, digital, family, and school conditions. Machine learning can model these combinations, while explainable artificial intelligence can make the resulting predictive structure inspectable [13, 14].

The study makes three connected contributions. Methodologically, it conducts a Türkiye-specific, plausible-value-aware comparison of weighted logistic regression, RBF-SVM, ExtraTrees, XGBoost, and CatBoost using student weights and school-grouped evaluation. Substantively, it distinguishes digital learning capacity, comprising resources, self-efficacy, and subject-related use, from ICT frequency. Interpretively, PV-aggregated SHAP tests whether digital resources and mathematics anxiety form a conditional profile and whether relational indicators are globally prominent or context-dependent [15–17]. The contribution therefore lies in the conditional and PV-robust structure revealed, rather than in the use of machine learning or SHAP alone. The research questions are:

1. Which student, family, school, and ICT-related variables are most closely associated with academic resilience among disadvantaged students in Türkiye?
2. How do tuned Support Vector Machine models and tree-based models compare when classifying resilient and non-resilient disadvantaged students?
3. What does SHAP reveal about the structure, direction, and combination of the most influential predictors?

2 Literature review and conceptual framework

2.1 Academic resilience: from developmental process to PISA-based operationalization

Academic resilience is generally located at the intersection of adversity and positive adaptation. In developmental research, resilience has long been understood not as a rare heroic quality or a fixed personal trait, but as a dynamic process through which individuals achieve relatively positive outcomes despite exposure to risk [1, 18]. This perspective is important for educational research because it prevents resilience from being reduced to innate talent, exceptional motivation, or individual willpower alone. Instead, resilience is shaped by the interaction between individual capacities and environmental conditions, such as social support, educational opportunities, and community resources.

In educational contexts, academic resilience refers more specifically to the capacity of students to achieve positive academic outcomes despite social, economic, or educational disadvantage. Martin and Marsh [2] showed that academic resilience is associated with academic self-beliefs, persistence, lower anxiety, and positive engagement. Their work is especially relevant because it connects resilience not only to performance outcomes, but also to the psychological and motivational conditions that allow students to continue learning under pressure. From this perspective, resilient students are not simply high achievers; they are students whose achievement is meaningful because it occurs under conditions of disadvantage.

In large-scale assessment research, especially studies based on the Programme for International Student Assessment (PISA), academic resilience is usually operationalized more narrowly. Students are often classified as resilient when they come from socioeconomically disadvantaged backgrounds yet achieve high academic performance [3–5]. This operational definition is practical and policy-relevant because it allows researchers to identify students who perform well despite structural disadvantage. However, it also creates an important conceptual tension. A PISA-based resilience label depends on how socioeconomic disadvantage is defined, which achievement domain is selected, and whether the performance threshold is national or international.

Academic resilience should therefore be treated as a valuable analytical framework rather than a definitive student category. Different thresholds can identify different groups of resilient students, so the present quartile-based definition is interpreted as a policy-relevant operationalization rather than an essential identity.

2.2 Protective factors in PISA-based academic resilience research

Previous PISA-based studies show that academic resilience is rarely associated with one single protective factor. Rather, resilient students tend to differ from other disadvantaged students across multiple domains, including psychosocial characteristics, family support, teacher support, school climate, and learning opportunities [3–5, 9, 10]. This multi-domain pattern is important because it suggests that resilience should not be interpreted as individual success detached from context. Even when resilience is measured at the student level, it is shaped by the environments in which students learn.

At the student level, psychosocial variables are often central. Prior research has linked academic resilience to confidence, engagement, persistence, sense of belonging, and lower levels of academic anxiety [2, 9]. These factors matter because disadvantaged students may face not only material constraints but also psychological pressures that affect learning, which can lead to lower academic performance and hinder their overall

educational outcomes. Anxiety can be especially relevant in mathematics because selection, progression, and high-stakes testing frequently link to achievement. High mathematics anxiety may reduce persistence and make it harder to convert available learning opportunities into strong performance for disadvantaged students.

Prior research has underscored the significance of support structures at both familial and educational levels. Teacher support, structured classroom environment, school safety, positive discipline, familial support, and established learning routines can assist disadvantaged students in maintaining academic engagement [3–5, 9, 10]. These findings endorse the perspective that academic resilience is not solely a psychological outcome. It is both relational and institutional. Disadvantaged students are more likely to succeed when they have access to supportive adults, safer learning environments, and school conditions that protect learning time.

However, empirical patterns are not consistent across countries. The same protective factor may not be equally important in every education system because resilience is shaped by national curricula, school stratification, assessment pressures, family resources, and broader social inequality. The Turkish case should therefore not be treated as a simple replication of cross-national averages. It is important to examine which conditions distinguish resilient disadvantaged students within Türkiye rather than assume that international patterns apply uniformly.

Türkiye-focused PISA research provides a more specific baseline. In PISA 2012, resilient students' positive school attitudes and sense of belonging were salient, while school-level differences were modest [19]. A multilevel analysis linked resilient Turkish students' mathematics achievement to applied mathematical tasks, school educational resources, and student-related school climate [20]. More recent PISA 2022 evidence for Türkiye reports that socioeconomic status, mathematics anxiety, and effort predict mathematics achievement for both genders, whereas teacher support was significant only for female students [21]. Together, these studies indicate psychological, resource, school-climate, and gender-contingent support processes, but leave the post-pandemic digital configuration examined here insufficiently understood.

2.3 Digital inequality, ICT use, and post-pandemic learning conditions

Digital factors have become increasingly important in discussions of educational inequality. Earlier debates often focused on whether students had access to devices or an internet connection. More recent research shows that digital inequality is more complex. It includes not only material access but also the quality of access, digital skills, confidence in using technology, and the ability to use digital tools for academic purposes [7, 22]. This broader understanding is important because two students may both have access to technology but differ greatly in how effectively they can use it for learning.

The post-pandemic context makes this issue even more important. PISA 2022 was conducted after a period of major educational disruption, during which digital technologies became central to learning continuity in many systems [7]. In Türkiye, the Ministry of National Education strengthened the Education Information Network (EBA) as a national distance-learning platform during COVID-19 school closures, alongside the broader Fırsatları Artırma ve Teknolojiyi İyileştirme Hareketi (FATİH; Movement to Increase Opportunities and Improve Technology) infrastructure for connectivity, hardware, digital content, and teacher use of technology [23]. These initiatives created a

common national digital infrastructure, but did not remove inequalities in household connectivity, device availability, quiet learning spaces, teacher mediation, or students' capacity to translate access into academic use. ICT indicators should therefore be read as markers of unequal uptake and educational use within a post-pandemic infrastructure, not as simple measures of programme exposure or device ownership.

For this research, Information and Communication Technology (ICT) variables are especially important because they do not represent a single dimension. ICT resources refer to the availability of digital tools and infrastructure. ICT self-efficacy reflects students' confidence in using digital technologies. Subject-related ICT use indicates the connection between technology and academic work. Inquiry-based ICT use reflects whether students use digital tools to search, explore, and learn. These distinctions matter because access alone may not produce academic benefit unless students also have the confidence and support needed to use technology productively.

Simultaneously, ICT should not be regarded as inherently advantageous. Digital use can support learning, but it can also reflect distraction, unequal home environments, or non-academic screen time. Therefore, the role of ICT in academic resilience should be examined empirically rather than assumed in advance. In the context of disadvantaged students, ICT-related variables may function either as protective resources or as indicators of broader inequality. This makes them theoretically important for understanding why some disadvantaged students achieve strong mathematics outcomes while others do not.

2.4 Machine learning and explainable artificial intelligence in educational resilience research

Many studies of academic resilience use descriptive comparisons, regression models, or logistic classification. These approaches are useful for testing specified hypotheses, but they may be limited when resilience depends on non-linear relationships and interactions among student, family, school, psychological, and digital variables.

Machine learning methods can help identify complex predictive structures in educational data, including relationships that are non-linear or conditional. Recent work illustrates this breadth through prediction of students' performance [24], analysis of faculty feedback [25], and context-aware recommendations in mobile learning [26]. Predictive performance alone is not sufficient for research on disadvantaged students, however: the educational value of a model depends on whether its patterns can be interpreted cautiously and connected to theory.

Explainable artificial intelligence (XAI) makes model behaviour more transparent [14], which is especially important when research concerns disadvantaged students [13, 27]. SHAP quantifies global and local feature contributions to a fitted prediction [17]; here it is used to inspect the structure of predictive profiles, not to estimate intervention effects.

SHapley Additive exPlanations (SHAP) is grounded in cooperative game theory and expresses how each feature contributes to a model output [17]. Its global summaries support comparison of influential features across PV-specific models, while dependence profiles show direction and heterogeneity. These quantities explain model behaviour and do not establish that changing a feature would change resilience.

2.5 Conceptual framework: a layered model of academic resilience

This research is guided by a layered view of academic resilience. The psychological layer includes mathematics anxiety, confidence, belonging, and related emotional conditions; the digital layer includes ICT resources, self-efficacy, subject-related use, and broader activity patterns; and the relational layer includes family, teacher, peer, and school conditions. The framework assumes that these layers may combine rather than operate independently and motivates the pre-specified interaction features [1, 18].

The psychological layer includes mathematics anxiety, confidence, belonging, and related emotional conditions. Anxiety is especially relevant because it can disrupt persistence, attention, and problem solving; lower anxiety may therefore mark a protective condition for disadvantaged students.

The digital layer includes ICT resources, self-efficacy, subject-related use, inquiry-oriented use, and broader activity patterns. It treats technology as a source of opportunity, capability, and inequality, rather than assuming that access or frequency is automatically beneficial.

The relational layer includes family and teacher support, school safety, bullying, classroom climate, and belonging. These indicators represent the social and organizational conditions in which disadvantaged students attempt to adapt academically.

The empirical analysis can therefore support or refine the framework. Strong global importance for digital and psychological variables would show their centrality in the fitted profile; weaker global importance for relational indicators would suggest a more conditional or context-dependent role rather than their educational irrelevance. The $ICTRES \times ANXMAT$ and $ICTENQ \times TEACHSUP$ interactions make these cross-layer possibilities explicit.

3 Data and methods

3.1 Data source and analytic sample

This research used the Programme for International Student Assessment (PISA) 2022 student questionnaire database for Türkiye. The analysis used student-level achievement, background, school-identifier, questionnaire, and survey-weight data. PISA combines nationally representative achievement data with contextual information on students, families, schools, and learning environments [28, 29]. The full Turkish student sample comprised 7250 students after filtering the international database to Türkiye.

The PISA final student-weight variable (W_FSTUWT) was retained throughout the analysis to incorporate unequal student selection weights. The school-identifier variable ($CNTSCHID$) was retained because students are nested within schools. Random student-level splitting could place students from the same school in both training and test sets and produce optimistic performance estimates; all train-test procedures therefore respected school-level grouping [30].

The analysis focused on disadvantaged students in Türkiye. Students were initially classified according to their position in the national distribution of socioeconomic status. The final analytic sample consisted of students who fell into the bottom quartile of the weighted Turkish socioeconomic distribution.

3.2 Outcome definition: academic resilience

Academic resilience was defined using a country-specific and weight-aware approach. Socioeconomic disadvantage was measured using the PISA Index of Economic, Social and Cultural Status (ESCS), which is the standard PISA index for capturing students' socioeconomic background [28, 29]. Students were classified as disadvantaged if their ESCS score was at or below the weighted first quartile of the Turkish ESCS distribution. The weighted first-quartile (Q1) ESCS threshold was -2.0638 .

Academic achievement was measured separately with each of the ten PISA mathematics plausible values (PVs; PV1MATH-PV10MATH). Following OECD guidance, the PVs were not averaged at the student level. Instead, for each PV, we calculated a weighted Turkish mathematics 75th-percentile (Q75) threshold, formed a PV-specific resilience label, and re-ran the complete school-grouped modelling pipeline. The PV-specific Q75 thresholds ranged from 514.378 to 516.335, and the number of resilient students ranged from 204 to 230. Performance metrics and SHAP quantities were then summarized across the ten parallel analyses. This design preserves measurement uncertainty in the outcome definition while retaining a transparent, country-specific threshold rule [29].

Thus, a student was classified as resilient if two conditions were concurrently satisfied: (i) the student belonged to the bottom quartile of the national ESCS distribution; and (ii) the student reached the top quartile of the national mathematics performance distribution.

The common disadvantaged subgroup comprised 1,885 students from 179 schools. Across the ten PV-specific labels, the mean weighted prevalence of academic resilience was 0.117; the corresponding unweighted resilient count varied from 204 to 230. This confirms that resilience is a minority outcome and justifies metrics designed for imbalanced classification.

The choice of national, quartile-based thresholds is theoretically and methodologically motivated. Academic resilience is relative to a risk context [1, 18], so disadvantaged students' achievement is evaluated against Türkiye's opportunity structure. An international achievement threshold would identify very few Turkish disadvantaged students because Türkiye's mean mathematics performance is below the OECD average [8], thereby mixing national performance differences with within-system resilience. Weighted quartile cutoffs are transparent and established in PISA-based resilience research [3–5]. Computing them with `W_FSTUWT` incorporates the final student weight into threshold estimation and better aligns the cutoffs with the weighted Turkish distribution; it does not implement full design-based variance estimation.

Because the outcome depends on threshold choices, changing the socioeconomic cutoff, achievement criterion, or reference distribution could alter subgroup composition, resilience prevalence, class imbalance, model performance, and SHAP rankings. The present national-quartile definition is therefore a policy-relevant operationalization rather than an absolute student category; alternative definitions require direct sensitivity analysis.

3.3 Predictor selection and theoretical domains

The predictor set was specified a priori from the conceptual framework rather than selected automatically. It included psychosocial and school-climate indicators (ANXMAT, BELONG, BULLIED, FEELSAFE); family and teacher support

(FAMSUP, FAMSUPSL, TEACHSUP); ICT resources and access (ICTRES, ICTAVHOM, ICTAVSCH, ICTHOME, ICTSCH, ICTQUAL); and digital competence and use (ICTEFFIC, ICTENQ, ICTFEED, ICTINFO, ICTOUT, ICTSUBJ, ICTWKDY, ICTWKEND, ICTREG, ICTDISTR). Gender (ST004D01T) was included as a descriptive demographic feature.

Four theory-guided interaction features were included: $ICTRES \times ANXMAT$, $ICTENQ \times TEACHSUP$, $ICTWKDY \times BELONG$, and $ICTEFFIC \times FAMSUP$. They were pre-specified to represent digital-psychological and digital-relational combinations rather than selected automatically. The explicit set was deliberately limited to four theory-mandated pairings instead of an exhaustive search that could overfit the small, rare-outcome sample; the tree models could still capture additional interactions through recursive splits (Table 1).

3.4 Missing data and preprocessing

This research did not use complete-case deletion because it would reduce the already imbalanced disadvantaged sample and could introduce additional bias when missingness is associated with student background, school context, or digital access. The observed data cannot identify whether questionnaire responses were missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR); in this setting it may plausibly be related to observed background and access conditions and,

Table 1 Predictor variables and interaction features by theoretical domain

Variable	Definition	Theoretical domain
ANXMAT	Mathematics anxiety	Psychosocial and school climate
BELONG	Sense of belonging	Psychosocial and school climate
BULLIED	Exposure to bullying	Psychosocial and school climate
FEELSAFE	Feeling safe at school	Psychosocial and school climate
FAMSUP	Family support	Family and teacher support
FAMSUPSL	Family support for self-regulated learning	Family and teacher support
TEACHSUP	Teacher support	Family and teacher support
ICTRES	ICT resources	ICT resources and access
ICTAVHOM	ICT availability at home	ICT resources and access
ICTAVSCH	ICT availability at school	ICT resources and access
ICTHOME	ICT use at home	ICT resources and access
ICTSCH	ICT use at school	ICT resources and access
ICTQUAL	Quality of ICT resources	ICT resources and access
ICTEFFIC	ICT self-efficacy	ICT use and digital competence
ICTENQ	Inquiry-based ICT use	ICT use and digital competence
ICTFEED	ICT use for feedback	ICT use and digital competence
ICTINFO	ICT use for information	ICT use and digital competence
ICTOUT	ICT use outside the classroom	ICT use and digital competence
ICTSUBJ	Subject-related ICT use	ICT use and digital competence
ICTWKDY	Weekday ICT use	ICT use and digital competence
ICTWKEND	Weekend ICT use	ICT use and digital competence
ICTREG	ICT use regulation	ICT use and digital competence
ICTDISTR	ICT-related distraction	ICT use and digital competence
ST004D01T	Student gender (1 = female, 2 = male)	Demographic
$ICTRES \times ANXMAT$	ICT resources $\times$ mathematics anxiety	Interaction (digital $\times$ psychological)
$ICTENQ \times TEACHSUP$	Inquiry-based ICT use $\times$ teacher support	Interaction (digital $\times$ relational)
$ICTWKDY \times BELONG$	Weekday ICT use $\times$ sense of belonging	Interaction (digital $\times$ psychological)
$ICTEFFIC \times FAMSUP$	ICT self-efficacy $\times$ family support	Interaction (digital $\times$ relational)

for some items, to unobserved response processes. Missingness indicators were retained because absence of a response may itself contain information about access or response patterns [31]. Median/mode single imputation was selected to maintain a common fold-specific representation across all model families; however, it does not propagate imputation uncertainty as multiple imputation would, and this is treated as a limitation.

To ensure a fair comparison, all five model families received the same fold-specific missing-data representation: numerical variables were median-imputed and accompanied by missingness indicators, while the categorical gender variable was imputed with its training-partition mode. CatBoost's native missing-value mechanism was therefore not used as a benchmarking advantage in the primary analysis. The shared representation makes differences in performance less dependent on an algorithm-specific treatment of missingness.

All data-dependent preprocessing was fit exclusively on the relevant training partition. Within the inner grouped development split, imputation values, missingness indicators, gender encoding, and SVM scaling parameters were learned from the inner training subset and then applied unchanged to validation data. After the configuration and threshold were selected, the same operations were refit on the outer training subset and applied unchanged to the final holdout test set. The pre-specified interaction features were deterministic row-wise products and involved no estimated parameters; all transformations that learn from the data remained training-only. Numerical variables were standardized only for the SVM after the common training-only imputation step.

For CatBoost, the shared imputed feature matrix was retained without scaling, and gender was passed as a categorical variable. ExtraTrees and XGBoost used the same imputed numerical representation and one-hot gender encoding; the SVM additionally used standardized numerical variables. Thus, model-specific operations were limited to those required by the algorithm rather than to different information about missing values. Several ICT measures capture related aspects of access, competence, and use; their SHAP attributions may therefore share explanatory contribution and are interpreted as a domain-level profile rather than as isolated causal importance.

3.5 Mathematical formulation and model comparison strategy

For each plausible value m and student i , the disadvantaged cohort and binary resilience outcome were formalized as follows:

$$D = \{i: ESCS_i \leq Q_{0.25}^w(ESCS)\}, y_i^{(m)} = I\left[PV_i^{(m)} \geq Q_{0.75}^w(PV^{(m)})\right], \quad m = 1, \dots, 10 \quad (1)$$

Here Q^w denotes a W_FSTUWT -weighted quantile and $I[\cdot]$ is the indicator function. Within each grouped split, model parameters and the minority-class decision threshold were selected by

$$\hat{\theta}_m = \arg \min_{\theta} \sum_{i \in T_m} w_i L\left(y_i^{(m)}, f_{\theta}(x_i)\right), \quad \hat{t}_m = \arg \max_t F1_w^{(m)}(t; V_m) \quad (2)$$

Here, T and V denote the grouped training and validation sets for plausible value m , w is the final student weight, and L is the model-specific weighted classification loss. Each holdout metric M was aggregated across the ten plausible values as

$$\underline{M} = \frac{1}{10} \sum_{m=1}^{10} M_m \quad (3)$$

For CatBoost, TreeSHAP additively decomposed each PV-specific model's log-odds output as

$$g(f_m(x_i)) = \varphi_0^{(m)} + \sum_j \varphi_{ij}^{(m)} \quad (4)$$

The benchmark separated predictive ranking from model explanation. PR-AUC was primary for model comparison, and CatBoost was considered for PV-aggregated SHAP only after its minority-class performance was compared with the traditional, margin-based, and tree-based alternatives.

Five model families were benchmarked:

1. Weighted Logistic Regression
2. Radial Basis Function Support Vector Machine (RBF-SVM)
3. ExtraTrees
4. Extreme Gradient Boosting (XGBoost)
5. CatBoost

Weighted logistic regression was included as the traditional statistical benchmark and received the same training-only imputation, encoding, scaling, interaction features, student weights, and threshold tuning as the other models. The RBF-SVM provided a non-linear margin-based comparator [32]. ExtraTrees [33], XGBoost [34], and CatBoost [35] represented randomized and boosted tree ensembles capable of modelling non-linearities and interactions.

The model comparison was not designed to select the model with the highest raw accuracy alone. Because academic resilience was a minority outcome, the evaluation focused on whether each model could balance overall discrimination with meaningful identification of resilient students. This was especially important because a model could achieve high accuracy simply by favouring the non-resilient majority class. In imbalanced classification problems, accuracy alone can therefore be misleading [36].

3.6 Validation design and threshold tuning

All train-test procedures were group-aware. A grouped holdout split reserved 20% of the 179 schools for final testing; the `test_size` parameter therefore applied to schools rather than to individual students. In the primary seed-42 split, 36 schools contained 288 disadvantaged students, all of whom were retained after cohort definition. The PV-specific test sets consequently contained 39 to 45 resilient students and 243 to 249 non-resilient students. The apparent difference from 20% of 1,885 students reflects unequal school cluster sizes, not dropped records or post-split exclusions.

The remaining 80% of the data were further divided into grouped training and validation subsets. The validation subset was used to tune hyperparameters and the classification threshold. Precision-Recall Area Under the Curve (PR-AUC) was the primary criterion for selecting a configuration within each model family because the resilient class was rare; balanced accuracy and resilient-class F1 were used as secondary

tie-breakers. The final holdout test set remained untouched until the selected configuration was evaluated.

Threshold tuning was carried out only on the validation subset. To quantify sensitivity to the school-grouped outer split, we also repeated the complete outer-split, inner-validation, tuning, and holdout-evaluation procedure across three independently seeded school-grouped splits and all ten plausible values (30 PV-specific external evaluations per model). We report cluster-bootstrap 95% confidence intervals (CIs) over the three repeat-level PV means. These intervals quantify split and plausible-value variability; they are not design-based PISA variance estimates.

The main evaluation metrics were the following:

- accuracy
- balanced accuracy
- precision for the resilient class
- recall for the resilient class
- F1 score for the resilient class
- Receiver Operating Characteristic Area Under the Curve (ROC-AUC)
- Precision-Recall Area Under the Curve (PR-AUC)

Balanced accuracy and PR-AUC were treated as especially important. Balanced accuracy is more informative than raw accuracy when classes are imbalanced because it gives equal weight to class-specific sensitivity [37]. PR-AUC is also important in rare-outcome settings because it focuses directly on the trade-off between precision and recall for the positive class [38].

3.7 Survey weights in model training and evaluation

The final student weight W_{FSTUWT} was used to make the analysis more consistent with the PISA sampling design. Weighted thresholds were used to define both socioeconomic disadvantage and high mathematics achievement. Student weights were also incorporated into model training where the algorithm allowed sample weights. In addition, holdout evaluation metrics were computed in a weight-aware manner so that model performance reflected the weighted structure of the Turkish PISA sample.

This survey-aware approach incorporates unequal selection weights, but it does not reproduce the full PISA complex-sample variance procedure. PISA 2022 uses Fay's variant of balanced repeated replication (Fay-BRR) with 80 replicate weights, together with stratification and school-level clustering, for design-based sampling variances [29]. This study uses W_{FSTUWT} and school-grouped validation but does not apply the 80 Fay-BRR weights or model stratification. The repeated-split intervals therefore quantify split and plausible-value variability and supplement rather than replace design-based uncertainty.

3.8 Final interpretation strategy

The interpretation strategy used two explicit stages. First, PR-AUC ranked minority-class performance, with balanced accuracy and resilient-class F1 providing complementary evidence. Second, CatBoost was retained as the explanatory anchor because it narrowly led the primary PR-AUC comparison (0.330 versus 0.328 for ExtraTrees), supports exact TreeSHAP, and produced stable PV-aggregated feature rankings. The

repeated grouped analysis did not establish unique predictive superiority, so all SHAP findings remain CatBoost-specific descriptive explanations.

4 Results

4.1 Weighted sample profile and prevalence

The initial Turkish sample included 7250 students. After the weighted ESCS threshold was applied, the disadvantaged subgroup comprised 1885 students. Across the ten PV-specific labels, the unweighted resilient count ranged from 204 to 230 (mean = 217), and the mean weighted prevalence was 11.7%.

4.2 Class imbalance and evaluation logic

The PV-specific labels confirmed a rare outcome in every analysis. The weighted resilience prevalence ranged from 0.110 to 0.123 across the ten plausible values (mean = 0.117). Therefore, raw accuracy was not treated as the primary basis for model comparison; PR-AUC, balanced accuracy, resilient-class precision, recall, and F1 were reported for each PV and summarized across the ten holdout evaluations.

4.3 Model comparison

Table 2 reports PV-aware weighted holdout performance. For each plausible value, hyperparameters and the decision threshold were selected using only the grouped validation set; the reported values summarize the ten final holdout evaluations as means and standard deviations (SDs). No model was selected on the basis of a single holdout result.

In the primary PV-aware holdout, CatBoost had mean PR-AUC of 0.330 (SD = 0.051) and resilient-class F1 of 0.301 (SD = 0.075), while RBF-SVM had PR-AUC of 0.304 (SD = 0.027). The weighted logistic baseline had PR-AUC of 0.289 (SD = 0.039), balanced accuracy of 0.615 (SD = 0.021), and resilient-class recall of 0.563 (SD = 0.180), illustrating the threshold-dependent precision-recall trade-off. The repeated results in Table 3 provide the broader ranking context.

Interpreted against the base rate, these results have substantive meaning. A random classifier would therefore be expected to have a weighted resilience prevalence of around 0.117, with a PR-AUC close to this value. The observed PR-AUC of around 0.29–0.33, therefore, indicates a nearly three-fold increase over chance but still leaves a large portion of the variation in resilience unexplained. This ceiling means that the profile of psychological, digital, family and school conditions, based on the questionnaire,

Table 2 PV-aware weighted holdout performance by model family, mean (SD)

Model	Accuracy	Balanced accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC
Logistic Regression	0.653 (0.109)	0.615 (0.021)	0.223 (0.046)	0.563 (0.180)	0.302 (0.018)	0.691 (0.014)	0.289 (0.039)
RBF-SVM	0.817 (0.039)	0.589 (0.032)	0.322 (0.073)	0.277 (0.111)	0.279 (0.058)	0.694 (0.028)	0.304 (0.027)
ExtraTrees	0.847 (0.028)	0.571 (0.030)	0.397 (0.091)	0.194 (0.094)	0.242 (0.075)	0.712 (0.018)	0.328 (0.035)
XGBoost	0.835 (0.027)	0.574 (0.036)	0.336 (0.084)	0.217 (0.100)	0.250 (0.079)	0.695 (0.027)	0.288 (0.030)
CatBoost	0.837 (0.040)	0.596 (0.038)	0.383 (0.114)	0.267 (0.101)	0.301 (0.075)	0.712 (0.027)	0.330 (0.051)

Table 3 Repeated school-grouped PV-aware holdout performance, mean (95% cluster-bootstrap CI)

Model	PR-AUC	Balanced accuracy	Resilient-class F1	Resilient-class recall
ExtraTrees	0.332 [0.318, 0.348]	0.661 [0.597, 0.696]	0.348 [0.296, 0.383]	0.441 [0.265, 0.573]
Logistic Regression	0.302 [0.241, 0.340]	0.655 [0.590, 0.707]	0.309 [0.276, 0.356]	0.517 [0.323, 0.685]
RBF-SVM	0.302 [0.253, 0.332]	0.642 [0.610, 0.664]	0.314 [0.278, 0.359]	0.430 [0.321, 0.488]
CatBoost	0.296 [0.267, 0.316]	0.636 [0.587, 0.668]	0.310 [0.280, 0.341]	0.393 [0.240, 0.476]
XGBoost	0.292 [0.249, 0.321]	0.631 [0.611, 0.657]	0.306 [0.288, 0.319]	0.384 [0.316, 0.462]

differentiates resilient from non-resilient disadvantaged students to a statistically significant but limited extent, in line with the process-oriented view of resilience presented in the conceptual framework section: resilience is considered context-dependent adaptation, not as a fixed and fully measurable student characteristic.

The difference between model families is also interpretable. The weighted logistic baseline had the best recall for the resilient class but low precision, whereas the tree-based models achieved higher precision with some loss in recall. In practical terms, the linear additive profile casts a wide net that identifies many potentially resilient students but mislabels many others. The non-linear models identify narrower, more internally consistent resilient profiles but miss students whose resilience does not fit those profiles. Non-linear models improve precision without significantly increasing overall discrimination, suggesting that conditional combinations of predictors sharpen, rather than radically restructure, the resilience profile recoverable from these data.

4.4 Selection of CatBoost for PV-aggregated SHAP analysis

CatBoost was selected as an explanatory anchor, not as an overall predictive winner. It narrowly led the primary PR-AUC comparison (0.330 versus 0.328 for ExtraTrees), supports exact TreeSHAP, and yielded a stable interpretive profile: each of the nine highest PV-aggregated features appeared among the top ten in at least 90% of the PV-specific CatBoost models. In the repeated analysis, ExtraTrees had the highest mean PR-AUC and several model intervals overlapped. The CatBoost explanations are therefore used to interpret one competitive and PV-stable model, not to claim that it is uniquely optimal.

4.5 Primary CatBoost performance and repeated model comparison

Across three independently seeded outer splits and ten plausible values, ExtraTrees had the highest mean PR-AUC (0.332, 95% CI 0.318 to 0.348). Weighted logistic regression obtained 0.302 (95% CI 0.241 to 0.340), RBF-SVM 0.302 (95% CI 0.253 to 0.332), CatBoost 0.296 (95% CI 0.267 to 0.316), and XGBoost 0.292 (95% CI 0.249 to 0.321). The overlapping intervals preclude a categorical ranking among the latter four models.

4.6 PV-aware performance distribution

Figure 1 displays the distribution of PR-AUC across the ten PV-specific primary holdout evaluations. These values illustrate plausible-value sensitivity for one school-grouped split. The repeated split results in Table 3 provide the broader uncertainty assessment and show that no categorical model-superiority claim should be inferred from the primary split alone.

The primary and repeated evaluations, taken together, support a robustness reading, rather than a ranking reading. The five families of models all converged to a similar and modest level of discrimination across seeds, splits and plausible values and the

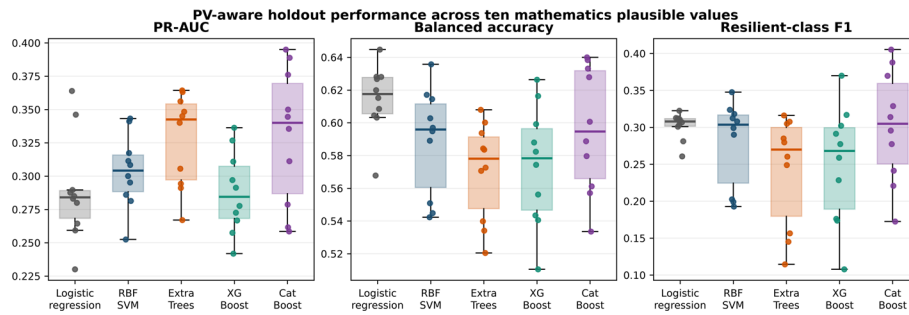


Fig. 1 PV-aware PR-AUC across ten mathematics plausible values

Table 4 PV-aggregated SHAP feature importance

Rank	Feature	Mean SHAP	Mean rank	Top-10 frequency (%)
1	ICT resources	0.456	1.2	100
2	Weekday ICT use	0.377	2.3	100
3	ICT resources × mathematics anxiety	0.299	5.3	90
4	Digital self-efficacy	0.296	5.8	90
5	Inquiry-based ICT use	0.288	5.8	90
6	Subject-related ICT use	0.277	6.1	100
7	Gender	0.260	7.6	90
8	Mathematics anxiety	0.247	7.6	90
9	Family support for school learning	0.221	8.9	90
10	Digital self-efficacy × family support	0.210	9.5	50

confidence intervals of four out of five families overlapped significantly. The convergence suggests that the substantive findings of the study do not depend on the choice of learner: the predictive signal is in the variables and their combinations, not in any particular algorithm. The variation across plausible values also indicates that some of the observed performance differences are a function of measurement uncertainty in the achievement construct itself. This further supports the decision to interpret only patterns that are consistent across the ten parallel analyses.

4.7 Global SHAP importance

The PV-aggregated SHAP results are shown in Table 4; Figs. 2, 3, 4 and 5. Mean absolute SHAP values were computed separately for each PV-specific CatBoost model and then averaged; average ranks and top-ten frequencies indicate stability across the ten analyses. The leading features span digital resources, digital use, digital self-efficacy, mathematics anxiety, family support, gender, and the ICTRES × ANXMAT interaction.

The ranking not only confirms the layered framework, but also refines it in three ways. The dominance of digital indicators cannot be explained solely by material access: the ranking incorporates digital self-efficacy and subject-related use in addition to ICT resources. So in the fitted model, what distinguishes resilient disadvantaged students is not just having technology but also the ability to convert it to academic activity. Second, the fact that mathematics anxiety and the ICTRES × ANXMAT interaction are among the strongest features indicates that the model captures a cross-layer configuration rather than a list of independent predictors: digital opportunity and emotional regulation jointly define the resilient profile, so that neither layer alone summarises it. Third, the rather low aggregate position of family support, and the absence of teacher

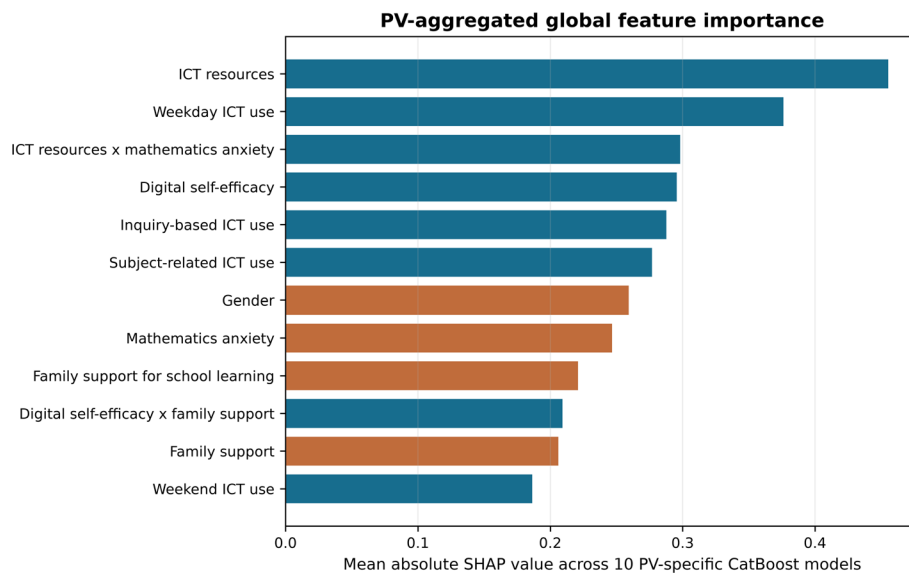


Fig. 2 PV-aggregated global feature importance

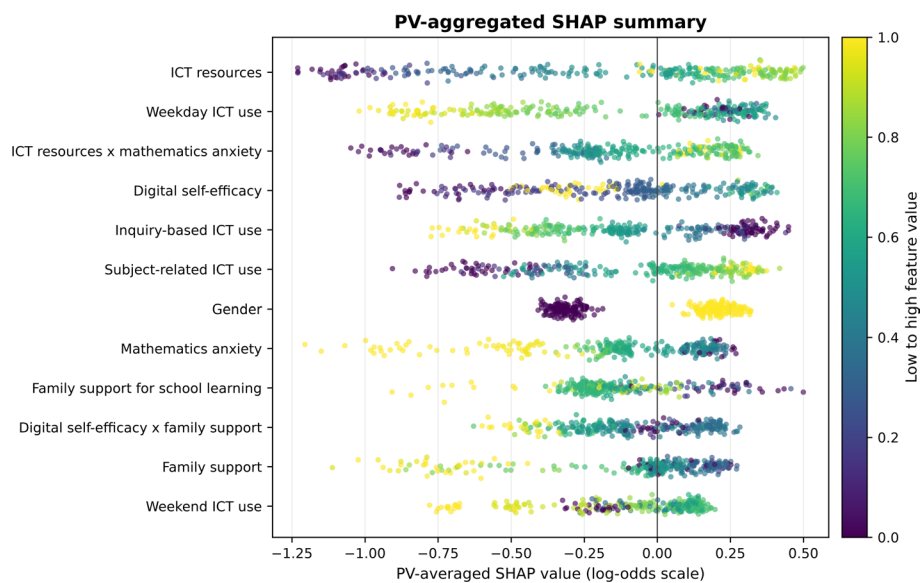


Fig. 3 PV-aggregated SHAP summary

support and school climate indicators at the top of the ranking, should be interpreted as bounded global prominence and not as irrelevance. Global SHAP averages contributions over the entire subgroup, which means that conditions that are important for specific students or school settings may have small global values. The pattern is therefore consistent with relational conditions operating conditionally, as the framework anticipated, rather than uniformly across all disadvantaged students.

4.8 Direction and shape of the strongest SHAP effects

The global SHAP summary plot is shown in Fig. 3, and dependence plots for ICT resources and mathematics anxiety are shown in Figs. 4 and 5.

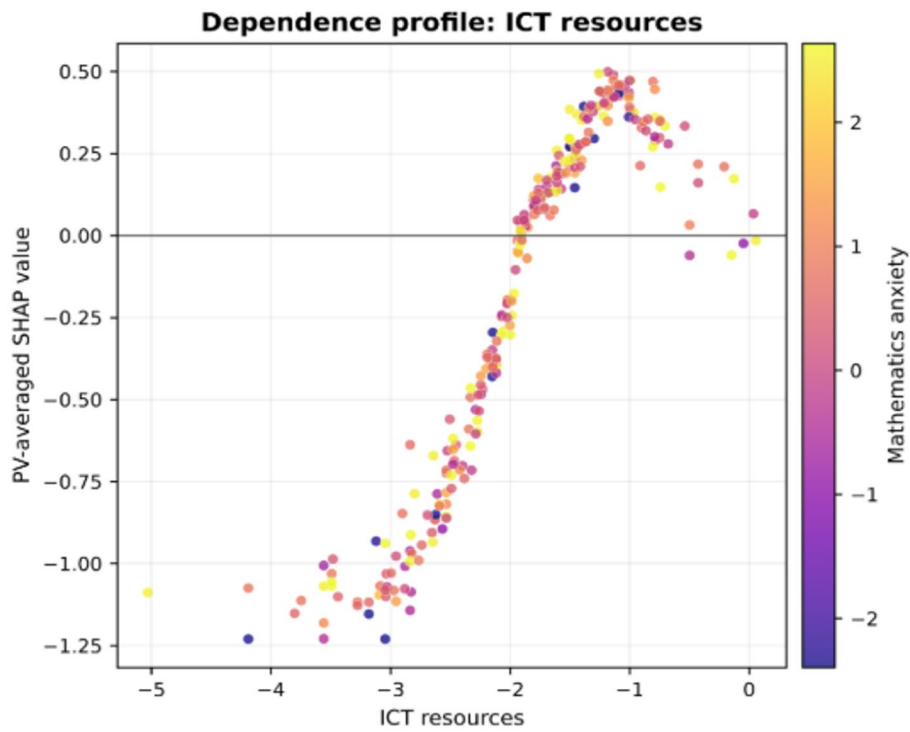


Fig. 4 PV-Averaged SHAP Dependence Profile for ICT Resources

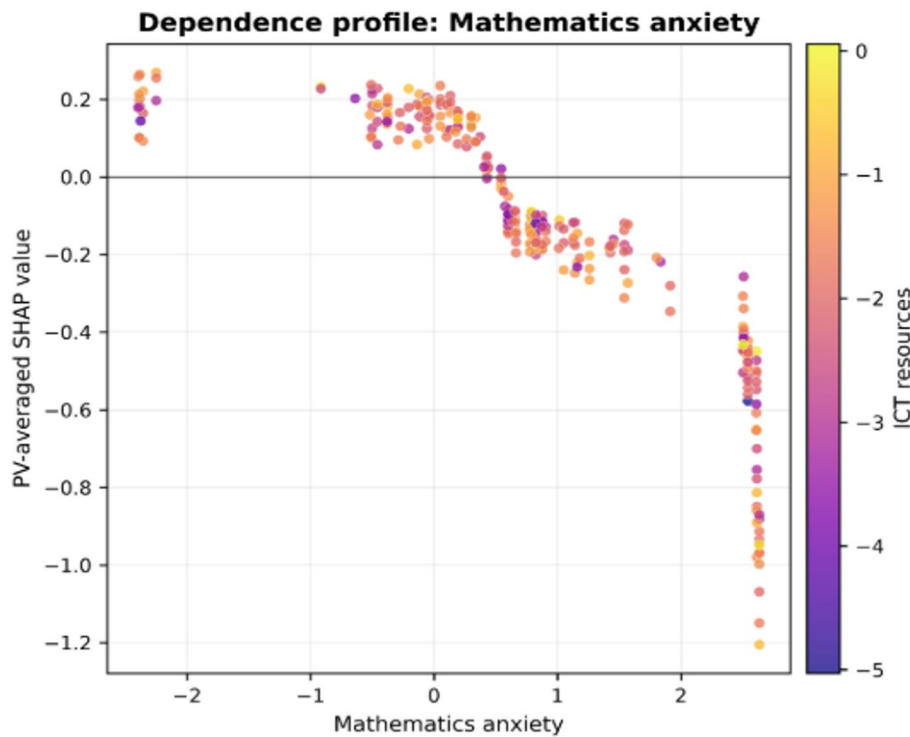


Fig. 5 PV-Averaged SHAP Dependence Profile for Mathematics Anxiety

The global SHAP summary shows that higher ICT resources, digital self-efficacy, and subject-related ICT use generally moved predictions toward resilience. In contrast, higher weekday ICT use (ICTWKDY) and inquiry-based ICT use (ICTENQ) appeared more often on the negative side of the SHAP axis. The model therefore distinguishes academically useful digital capacity from ICT exposure or frequency alone; it does not support a simple interpretation of ‘more ICT use is better.’

This divergence within the digital layer is one of the clearest interpretive results of the analysis. In the same models, indicators of sheer frequency or generic activity (weekday use, inquiry-oriented browsing) and indicators of digital learning capacity (resources, self-efficacy, subject-related use) contributed in opposite directions. The pattern is similar to the literature on the usage gap, where it is quality and academic orientation of the use of technology, rather than quantity, that differentiates educational outcomes. For disadvantaged students in a post-pandemic context, frequent unstructured use may signal the displacement of study time or weakly mediated home environments, while purposeful, subject-linked use signals that digital access has been converted into a learning opportunity. Thus, the difference between digital capacity and digital exposure was recovered from the data rather than imposed by variable selection.

The ICTRES dependence plot shows a clear non-linear pattern. Very low ICT resource values were associated with strongly negative SHAP values, meaning that they reduced the model’s predicted probability of resilience. As ICTRES increased, SHAP values became less negative and then positive. However, the relationship was not perfectly linear at the upper end of the scale. This suggests that ICT resources mattered strongly, but their contribution depended on the wider student profile.

The ANXMAT dependence plot shows the opposite pattern. Lower and near-average mathematics anxiety values were generally associated with neutral or positive SHAP contributions. As mathematics anxiety increased, SHAP values became increasingly negative. The highest anxiety values produced the strongest negative contributions to predicted resilience. This supports mathematics anxiety as one of the clearest model-based emotional markers.

The importance of ANXMAT becomes more meaningful when read together with ICTRES x ANXMAT. The model did not treat digital and emotional conditions as separate dimensions; it captured a pattern in which the contribution of resources was partly conditioned by mathematics anxiety.

5 Discussion

5.1 Refinement of the layered framework

The findings refine the layered framework in two ways. First, resources, self-efficacy, subject-related use, and ICT frequency do not form a single digital dimension; their SHAP prominence and directions differ. Second, ICTRES x ANXMAT shows that digital and psychological conditions jointly characterize the fitted resilience profile. Relational indicators remain visible through family support and teacher-related interactions but are more conditional than globally dominant.

Compared with prior Türkiye studies emphasizing school attitudes, educational resources, school climate, mathematics anxiety, and gender-contingent teacher support [19–21], the present findings add a post-pandemic digital-learning dimension and a digital-emotional interaction. This extension does not displace earlier protective

factors; it shows how their global prominence may change within the 2022 disadvantaged subgroup.

Resilience remains a rare outcome in this subgroup. The model detects a group-level signal, but survey measures do not cleanly separate resilient from non-resilient students. This boundary is central to the interpretation of both the model metrics and SHAP results.

5.2 Model performance and analytical scope

The model comparison indicates uncertainty rather than algorithmic dominance. CatBoost and ExtraTrees were nearly tied in the primary PR-AUC comparison, ExtraTrees led the repeated mean, and weighted logistic regression remained competitive on several threshold-dependent metrics. CatBoost's role is therefore interpretive: it anchors a stable TreeSHAP profile after, rather than instead of, predictive benchmarking.

The model's practical limits are equally important. Mean resilient-class recall is 0.267 (SD = 0.101), so many resilient students are missed. The model must therefore not be used for individual screening, labelling, resource allocation, or support decisions; its appropriate role is to characterize group-level predictive associations.

5.3 Digital learning capacity in Türkiye

In Türkiye's post-pandemic setting, digital variables may be influential because EBA and FATİH created common infrastructure while students remained unequal in their capacity to use that infrastructure for learning. The positive contributions of ICT resources, self-efficacy, and subject-related use therefore describe digital learning capacity: access that is usable, supported by confidence, and connected to academic tasks. They should not be read as effects of a particular platform or as evidence that device provision alone creates resilience [7, 23].

Digital frequency tells a different story. Higher weekday or inquiry-oriented ICT use was not uniformly favourable and may reflect unstructured use, distraction, compensatory searching, or other unmeasured conditions. Given the limited resilient-class recall and small positive class, these SHAP directions are exploratory and hypothesis-generating rather than evidence of a policy effect. They motivate research on the quality, purpose, and instructional guidance of digital engagement [39].

5.4 Mathematics anxiety and the digital-emotional interaction

Mathematics anxiety is a clear emotional marker in the model: high anxiety moves predictions away from resilience, with stronger negative contributions at the upper end of the scale. This aligns with resilience theory and evidence linking mathematics anxiety to lower achievement [2, 40].

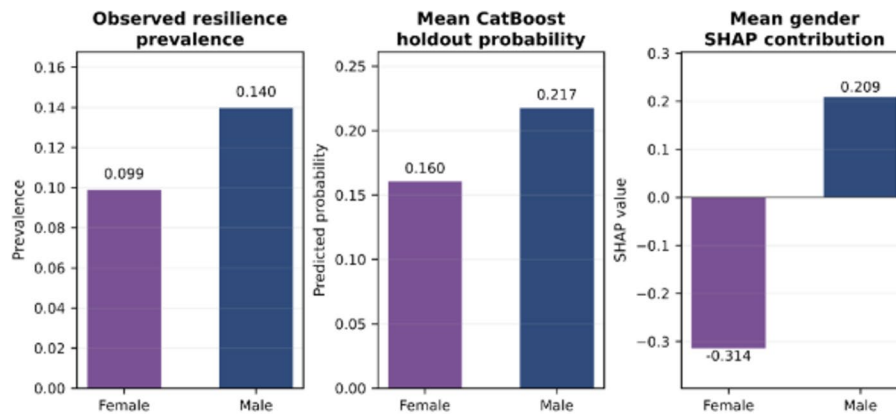
The ICTRES x ANXMAT interaction gives this finding additional educational meaning. Digital resources are not simply interchangeable with emotional readiness: resources may be less readily converted into mathematics learning when anxiety is high. This cross-layer pattern is what the model adds to a simple ranked list of protective factors.

5.5 Relational conditions and bounded global prominence

Family support is present in the resilience profile but its contribution is not uniformly positive. A plausible reading is that reported support can be compensatory: support may

Table 5 PV-aware gender patterns among disadvantaged students; values except n are mean (SD) across PVs

Gender	n	Weighted resilience prevalence	Mean holdout probability	Mean gender SHAP value
Female	1021	0.099 (0.006)	0.160 (0.038)	−0.314 (0.075)
Male	864	0.140 (0.007)	0.217 (0.042)	0.209 (0.050)

**Fig. 6** PV-aware gender patterns among disadvantaged students

increase when students experience academic difficulty. The result points to a relational process whose meaning depends on the wider profile, not to a harmful effect of family support.

Teacher support and school-climate indicators are less globally prominent than might be expected. This does not show that they are unimportant. The PISA indicators are broad and self-reported, may have limited variation within the disadvantaged subgroup, and may affect resilience through classroom practices or interactions not captured by a global SHAP rank. The weaker ICTENQ x TEACHSUP feature is consistent with a conditional, rather than absent, teacher role.

The relational layer is therefore retained in the framework but interpreted as bounded and context-dependent in this dataset. The negative local contribution of bullying is consistent with evidence that peer safety remains relevant for disadvantaged students' achievement [41]. Future work with mathematics-specific teacher measures, school-level designs, and longitudinal data should test these mechanisms directly.

5.6 Gender-related resilience patterns

As summarized in Table 5; Fig. 6, gender was a stable, non-causal feature of the PV-aggregated CatBoost models (mean SHAP rank = 7.6; top-ten frequency = 90.0%). Within the disadvantaged sample, mean weighted resilience prevalence was 0.099 for female students and 0.140 for male students; mean holdout probabilities showed the same descriptive pattern (0.160 and 0.217). This finding does not contradict the broader PISA 2022 Türkiye pattern in which boys and girls had similar average mathematics performance: the country note reports a larger share of top mathematics performers among boys (7%) than girls (4%), so the difference is best read as a distributional feature near the high-achievement threshold [8].

The ablation analysis shows that gender contains incremental predictive information in this sample: removing it reduces mean PR-AUC from 0.330 to 0.308 and resilient-class

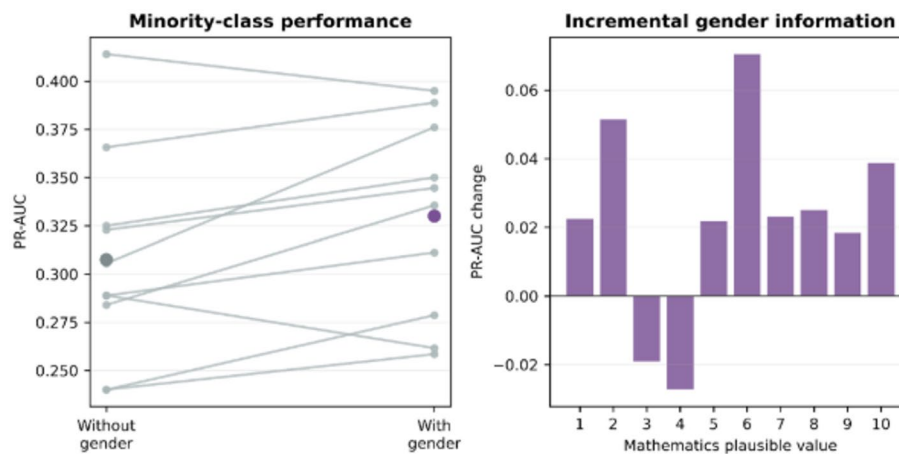


Fig. 7 Incremental gender information in PV-specific CatBoost models

F1 from 0.301 to 0.277. Figure 7 makes the PV-specific variation visible, so the result is not represented as exact or uniform across plausible values.

The gender pattern is descriptive and non-causal. It may proxy gendered experiences of mathematics learning, participation, family expectations, digital opportunity, or other unmeasured conditions. Together with the country-level distribution, it motivates investigation of high mathematics performance within disadvantaged groups, not gender-based labelling or claims about inherent capability.

5.7 Implications for educational policy

The findings support a layered equity response rather than an individual-resilience narrative. In Türkiye, digital equity should combine reliable access with digital self-efficacy, subject-related use, and teacher mediation; mathematics support should also address anxiety and threat-based learning conditions.

The model's distinction between digital learning capacity and ICT frequency cautions against device-only or screen-time policies. It also suggests that teacher and school conditions may be important enablers even when broad teacher and climate indicators are not globally dominant in the model.

Mathematics anxiety should be treated as part of educational equity. Supporting mathematics confidence, reducing threat-based learning environments, and strengthening formative support may be relevant to resilience-oriented policy, particularly where disadvantaged students have fewer compensatory learning resources outside school.

6 Conclusion

This study used a survey-aware, school-grouped, PV-specific machine-learning design to examine academic resilience among disadvantaged students in Türkiye. Repeating the workflow across ten mathematics plausible values preserved achievement-measurement uncertainty and showed that the main profile was not an artefact of a student-level PV average.

The main substantive finding is a PV-robust conditional profile: digital resources, self-efficacy, and subject-related ICT use are distinct from general ICT frequency; mathematics anxiety is a prominent negative feature; and $ICTRES \times ANXMAT$ links

digital opportunity with emotional conditions. Relational indicators appear more context-dependent than globally dominant.

The repeated grouped analysis and competitive weighted logistic baseline preclude a categorical model ranking. CatBoost provides the explanatory anchor for one competitive, PV-stable model rather than an operationally dominant classifier.

The cross-sectional design cannot establish causality, and SHAP explains fitted models rather than effects of changing students' circumstances. Single imputation does not propagate imputation uncertainty. The PV-aware and repeated grouped analyses do not replace design-based uncertainty from the 80 Fay-BRR weights and stratification. Models were developed and evaluated only in Türkiye, so generalizability to other countries or independent cohorts was not tested. Alternative cutoffs may change subgroup composition, class imbalance, performance, and SHAP rankings. Only a focused gender ablation was conducted; the unique incremental value of broader domains and explicit interaction features was not established. These limits preclude individual-level use of the predictions.

Future work should test alternative resilience definitions and external samples, implement full Fay-BRR uncertainty, conduct pre-specified domain and interaction ablations, use mathematics-specific and school-level support measures, and examine the digital-emotional and gendered patterns longitudinally.

Author contributions

The authors confirm their contribution to the paper as follows: study conception and design: M.E., E.B.K.D; data preparation: M.E.; analysis and interpretation of results: M.E., E.B.K.D., K.D.; draft manuscript preparation: E.B.K.D, K.D. All authors reviewed the results and approved the final version of the manuscript.

Funding

This research received no external funding.

Data availability

Publicly available datasets were analyzed in this study. The name of the data file can be found as "Student questionnaire data file": [<https://webfs.oecd.org/pisa2022/index.html>].

Code availability

Analyses were conducted in Python 3.12.13 using CatBoost 1.2.10, XGBoost 3.3.0, scikit-learn 1.8.0, pandas 3.0.3, NumPy 2.5.1, and pyreadstat 1.3.2.

Declarations

Ethics approval and consent to participate

This study analyzed publicly available, de-identified secondary data from the Programme for International Student Assessment (PISA) 2022. The authors did not recruit or contact participants and did not use personally identifiable data. Ethical approval was therefore not required for this secondary analysis.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 11 May 2026 / Accepted: 21 August 2026

Published online: 01 September 2026

References

1. Masten AS. Ordinary magic: resilience processes in development. *Am Psychol*. 2001;56(3):227–38. <https://doi.org/10.1037/0003-066X.56.3.227>.
2. Martin AJ, Marsh HW. Academic resilience and its psychological and educational correlates: a construct validity approach. *Psychol Sch*. 2006;43(3):267–81. <https://doi.org/10.1002/pits.20149>.
3. Organisation for Economic Co-operation and Development Against the odds: disadvantaged students who succeed in school. OECD Publishing; 2011. <https://doi.org/10.1787/9789264090873-en>.

4. Agasisti T, Avvisati F, Borgonovi F, Longobardi S. Academic resilience: what schools and countries do to help disadvantaged students succeed in PISA. OECD Education Working Papers No. 167. OECD Publishing; 2018. <https://doi.org/10.1787/e22490ac-en>.
5. Organisation for Economic Co-operation and Development. Equity in education: breaking down barriers to social mobility. OECD Publishing; 2018. <https://doi.org/10.1787/9789264073234-en>.
6. Organisation for Economic Co-operation and Development. PISA 2022 results. Volume I: the state of learning and equity in education. OECD Publishing; 2023. <https://doi.org/10.1787/53f23881-en>.
7. Organisation for Economic Co-operation and Development. PISA 2022 results. Volume II: learning during and from disruption. OECD Publishing; 2023. <https://doi.org/10.1787/a97db61c-en>.
8. Organisation for Economic Co-operation and Development. PISA 2022 results. Volume I and II country notes: Türkiye. OECD Publishing; 2023. Available from: https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/turkiye_d67e6c05-en.html
9. Organisation for Economic Co-operation and Development. PISA 2018 results. Volume II: where all students can succeed. OECD Publishing; 2019. <https://doi.org/10.1787/b5fd1b8f-en>.
10. Agasisti T, Avvisati F, Borgonovi F, Longobardi S. What school factors are associated with the success of socio-economically disadvantaged students? An empirical investigation using PISA data. *Soc Indic Res*. 2021;157:749–81. <https://doi.org/10.1007/s11205-021-02668-w>.
11. Cefai C. A transactional, whole-school approach to resilience. In: Ungar M, editor. *Multisystemic resilience: adaptation and transformation in contexts of change*. Oxford University Press; 2021. <https://doi.org/10.1093/oso/9780190095888.003.0013>.
12. Yang Hansen K, Thorsen C, Johansson S. Identifying school-level factors predicting academic resilience in Sweden: evidence from PISA studies. *Eur J Psychol Educ*. 2025;40:67. <https://doi.org/10.1007/s10212-025-00970-6>.
13. Khosravi H, Shum SB, Chen G, Conati C, Gašević D, Kay J, Knight S, Martinez-Maldonado R, Sadiq S, Tsai YS. Explainable artificial intelligence in education. *Comput Educ Artif Intell*. 2022;3:100074. <https://doi.org/10.1016/j.caeai.2022.100074>.
14. Arrieta AB, Diaz-Rodríguez N, Del Ser J, Bannett A, Tabik S, Barbado A, García S, Gil-López S, Molina D, Benjamins R, Chatila R, Herrera F. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inform Fusion*. 2020;58:82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>.
15. Huang Y, Zhou Y, Chen J, Wu D. Applying machine learning and SHAP method to identify key influences on middle-school students' mathematics literacy performance. *J Intell*. 2024;12(10):93. <https://doi.org/10.3390/jintelligence12100093>.
16. Cheung KC, Sit PS, Zheng JQ. A machine-learning model of academic resilience in the times of the COVID-19 pandemic: evidence drawn from 79 countries/economies in the PISA 2022 mathematics study. *Br J Educ Psychol*. 2024;94(4):1224–44. <https://doi.org/10.1111/bjep.12715>.
17. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *Advances in neural information processing systems*. 2017;30. <https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions>
18. Luthar SS, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. *Child Dev*. 2000;71(3):543–62. <https://doi.org/10.1111/1467-8624.00164>.
19. Yılmaz Findik L. What makes a difference for resilient students in Turkey? *Eur J Educ Res*. 2016;16(64). <https://dergipark.org.tr/en/pub/ejer/article/258629>
20. Ünsal Özberk E, Yılmaz Findik L, Özberk E. Investigation of the variables affecting the math achievement of resilient students at school and student level. *Educ Sci*. 2018;43(194):111–29. <https://doi.org/10.15390/EB.2018.7153>.
21. Çağlayan N, Gelbal S. Öğrencilerin matematik başarısını yordayan faktörler: PISA 2022 Türkiye örneği. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*. 2026;27(2):758–92. <https://doi.org/10.29299/kefad.1687058>.
22. van Deursen AJAM, van Dijk JAGM. Toward a multifaceted model of internet access for understanding digital divides: an empirical investigation. *Inform Soc*. 2015;31(5):379–91. <https://doi.org/10.1080/01972243.2015.1069770>.
23. Ministry of National Education. Türkiye'de COVID-19 salgını sürecinde Millî Eğitim Bakanlığı tarafından atılan politika adımları. 2020. <https://www.meb.gov.tr/turkiyede-covid-19-salgini-surecinde-milli-egitim-bakanligi-tarafindan-atilan-politika-adimlari/haber/20824/tr>
24. Deshpande SB, Tangod KK, Karekar NV, Upparmani PS. A comparative study of predictions of students' performances in offline and online modes using adaptive neuro-fuzzy inference system. *Int J Adv Comput Sci Appl*. 2022. <https://doi.org/10.14569/IJACSA.2022.0131036>.
25. Deshpande SB, Tangod KK, Srinivasiah SH, Alahmadi AA, Alwetaishi M, Michael GKO, Rajendran S. Elevating educational insights: sentiment analysis of faculty feedback using advanced machine learning models. *Adv Continuous Discrete Models*. 2025;2025:89. <https://doi.org/10.1186/s13662-025-03933-9>.
26. Deshpande SB, Michael GKO, Kadkol RJ, Karekar NV, Deshpande U, Lamani D. Optimising content recommendations in context-aware mobile learning platform through machine learning. *Appl Comput Intell Soft Comput*. 2025;2025:6982455. <https://doi.org/10.1155/acs/6982455>.
27. Fiok K, Farahani FV, Karwowski W, Ahram T. Explainable artificial intelligence for education and training. *J De/F Model Simul*. 2022;19(2):133–44. <https://doi.org/10.1177/15485129211028651>.
28. Organisation for Economic Co-operation and Development. PISA 2022 assessment and analytical framework. OECD Publishing; 2023. <https://doi.org/10.1787/dfe0bf9c-en>.
29. Organisation for Economic Co-operation and Development. PISA 2022 technical report. OECD Publishing; 2024. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/pisa-2022-technical-report_599753f0/01820d6d-en.pdf
30. Vabalas A, Gowen E, Poliakov E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS ONE*. 2019;14(11):e0224365. <https://doi.org/10.1371/journal.pone.0224365>.
31. Little RJA, Rubin DB. *Statistical analysis with missing data*. 3rd ed. Wiley; 2019. <https://doi.org/10.1002/9781119482260>.
32. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20:273–97. <https://doi.org/10.1007/BF00994018>.
33. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn*. 2006;63:3–42. <https://doi.org/10.1007/s10994-006-6226-1>.
34. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*; 2016. p. 785–94. <https://doi.org/10.1145/2939672.2939785>.

35. Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst*. 2018. <https://doi.org/10.48550/arXiv.1706.09516>.
36. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans Knowl Data Eng*. 2009;21(9):1263–84. <https://doi.org/10.1109/TKDE.2008.239>.
37. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. The balanced accuracy and its posterior distribution. In: 2010 20th international conference on pattern recognition; 2010. p. 3121–4. <https://doi.org/10.1109/ICPR.2010.764>.
38. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*. 2015;10(3):e0118432. <https://doi.org/10.1371/journal.pone.0118432>.
39. Volante L, Klinger DA. PISA, global reference societies, and policy borrowing: the promises and pitfalls of academic resilience. *Policy Futures Educ*. 2022;21(7):755–64. <https://doi.org/10.1177/14782103211069002>.
40. Barroso C, Ganley CM, McGraw AL, Geer EA, Hart SA, Daucourt MC. A meta-analysis of the relation between math anxiety and math achievement. *Psychol Bull*. 2021;147(2):134–68. <https://doi.org/10.1037/bul0000307>.
41. Siebecke DE. Does social well-being predict academic resilience and achievement? Analysis of Swedish PISA 2018 data. *Eur J Psychol Educ*. 2023;39(3):2105–28. <https://doi.org/10.1007/s10212-023-00762-w>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.