Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting School Happiness Among Primary School Students Using Explainable Machine Learning: Evidence from a Nationally Representative Sample in Türkiye

Cüneyt Akar

akarcuneyt@gmail.com

Usak University

Taha İlter Akar

University of Erlangen-Nuremberg

Research Article

Keywords: school happiness, peer bullying, digital media use, teacher intervention, family support, machine learning

Posted Date: April 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9206375/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Predicting School Happiness Among Primary School Students Using Explainable Machine Learning: Evidence from a Nationally Representative Sample in Türkiye

Abstract

This study examines the factors predicting school happiness among primary school students in Türkiye using an explainable machine learning approach. Data were collected from a nationally representative sample of 4,134 primary school students. School happiness was modeled as a continuous outcome variable, and sociodemographic, familial, behavioral, and school-contextual factors were analyzed using supervised regression techniques. Linear models were compared with tree-based ensemble models to capture nonlinear relationships and complex interactions among predictors. Results indicated that tree-based ensemble models outperformed linear models in out-of-sample prediction. The Gradient Boosting Regressor achieved the highest predictive performance, explaining approximately 23% of the variance in school happiness. Model diagnostics and calibration analyses supported the generalizability of the findings. Explainability analyses revealed that peer bullying was the strongest negative predictor of school happiness, whereas reading frequency and teacher intervention emerged as protective factors. Excessive mobile device use was associated with lower predicted happiness levels, particularly at higher usage durations. Overall, the findings demonstrate that school happiness can be reliably predicted using explainable machine learning methods. While the results are predictive rather than causal, they highlight bullying prevention, teacher support, and balanced digital media use as critical leverage points for school-based interventions and educational policy.

Keywords: school happiness; peer bullying; digital media use; teacher intervention; family support; machine learning

Introduction

Children's happiness has become an increasingly central focus in educational and psychological research (Ben-Arieh & Frønes, 2011; Diener et al., 2017; The Children's Society, 2024; UNICEF Innocenti, 2025). This growing interest reflects an expanding recognition that subjective well-being is not only associated with individual health and academic achievement but is also closely linked to the sustainability of social well-being more broadly (Diener et al., 2017). The primary school period is particularly critical in this respect, as it represents a developmental stage in which self-concept is formed, social relationships are established, and the core components of happiness begin to take shape (Holder & Coleman, 2009).

International organizations have similarly emphasized happiness as a fundamental educational outcome. The OECD (2019) identifies students' life satisfaction as a key indicator for evaluating educational systems, while UNESCO's Happy Schools framework positions happiness as a central goal of education, highlighting social connectedness, safety, pedagogical support, and participation as essential dimensions of quality education (UNESCO, 2016, 2024). Within this framework, school happiness emerges as a construct that extends beyond individual development to encompass broader concerns related to social cohesion and sustainability.

Empirical research across diverse contexts supports this perspective. Studies conducted with children aged 7–14 in Italy have shown that self-esteem, self-concept, and loneliness are among the primary determinants of happiness (Baiocco et al., 2019). In Finland, research involving 737 students demonstrated a strong association between school-based happiness and overall life satisfaction, with this relationship being reinforced by social relationships and leisure activities (Uusitalo-Malmivaara, 2012). At the European level, school satisfaction among 10-year-old children has been identified as a robust predictor of overall childhood happiness (Gómez-Baya et al., 2021). Similarly, experimental evidence from Canada indicates that social–emotional

learning programs enhance cognitive and emotional development and, in turn, support school happiness (Schonert-Reichl et al., 2015).

School happiness can be conceptualized as a multidimensional form of subjective well-being that encompasses positive affect, life satisfaction, and context-specific domains such as school belonging and perceived social support (Huebner, 1991; Huebner et al., 2006). Within this framework, well-being is shaped not only by internal emotional states but also by children's relational and environmental experiences. Prior research suggests that girls tend to report higher levels of happiness than boys (Çankaya & Meydan, 2018), that maternal education contributes positively to children's social–emotional adjustment (Göcen, 2015), and that low socioeconomic status may increase children's vulnerability to diminished well-being (Dilmaç & Özkan, 2019). In addition to demographic and socioeconomic factors, daily behavioral patterns play a meaningful role in shaping children's happiness. Reading habits have been positively associated with well-being, whereas excessive screen time has been linked to loneliness and psychological difficulties (Burhan & Moradzadeh, 2020; Twenge & Campbell, 2018).

Among the factors negatively associated with school happiness, peer bullying stands out as one of the most influential. A national study conducted in Thailand reported that nearly half of students experienced bullying, with a substantial proportion of these students reporting low levels of happiness (Aunampai et al., 2022). Similar findings have been documented in Türkiye, where bullying has been shown to be a strong predictor of reduced school happiness (Akar & Ay, 2025). Bullying is typically characterized by intentionality, repetition, and power imbalance, and may take physical, verbal, or relational forms (Olweus, 1993; Bradshaw et al., 2015). With increasing digitalization, cyberbullying has further intensified the risks to children's well-being, adding complexity to the social environments in which children develop (Hinduja & Patchin, 2014; Kowalski et al., 2014; Gini et al., 2007; Juvonen & Gross, 2008).

Consequently, bullying is best understood as an ecological phenomenon shaped by interactions among individual, familial, and school-level factors (Arseneault et al., 2010; Espelage, 2014; Hong et al., 2016).

This complex network of relationships often involves nonlinear patterns and multivariate interactions that may not be adequately captured by traditional statistical approaches. In this context, machine learning offers a methodological advantage by enabling the modeling of complex interactions, ranking predictor importance, and improving predictive accuracy without restrictive linearity assumptions (Goldstein et al., 2017; Hilbert, 2021; Weller et al., 2021). Despite these advantages, nationally representative studies at the primary school level in Türkiye that simultaneously examine school happiness alongside both traditional and digital forms of bullying using machine learning approaches remain scarce. Moreover, official well-being statistics, such as those reported by the Turkish Statistical Institute (TÜİK, 2024), largely focus on adult populations and provide limited insight into children's subjective well-being.

Addressing this gap, the present study analyzes data from 4,134 primary school students sampled from Türkiye's seven geographical regions to examine the factors predicting school happiness using machine learning techniques. Predictor variables were categorized into four domains:

(a) **Sociodemographic factors** (gender, age, grade level, nationality, number of siblings, birth order, household size, and place of residence);

(b) **Family factors** (parental education, employment status, cohabitation status, household income, and reporting bullying to family members);

(c) **Behavioral factors** (reading frequency, television viewing time, and mobile device/tablet use); and

(d) **School-contextual factors** (teacher intervention and a composite bullying score encompassing physical, verbal, relational, and cyber components).

Accordingly, the primary aim of this study is to identify the relative importance of sociodemographic, familial, behavioral, and school-contextual predictors of school happiness among primary school students in Türkiye using machine learning algorithms. By adopting a predictive—rather than causal—framework, the study seeks to provide a data-driven basis for prioritizing key psychosocial and behavioral factors relevant to school-based interventions and educational policy.

Research Questions

Guided by machine learning analyses, the study addresses the following research question:

Which sociodemographic, family-related, behavioral, and school-contextual variables demonstrate the strongest predictive power for school happiness among primary school students in Türkiye?

Method

Research Design

This study employed a descriptive, correlational research design to examine the factors predicting school happiness among primary school students. Given the multidimensional nature of school happiness and the need to assess the relative importance of demographic, familial, behavioral, and school-contextual predictors, machine learning approaches were adopted to accommodate complex, multivariate data structures.

Traditional regression techniques are often constrained by linearity assumptions and limited capacity to model higher-order interactions (Goldstein et al., 2017; Hilbert, 2021). To address these limitations, the present study utilized supervised learning algorithms, which offer methodological advantages in handling heterogeneous predictors, modeling nonlinear

relationships, and producing robust predictive performance (Weller et al., 2021). Accordingly, machine learning was used not as a replacement for theory-driven inquiry, but as a complementary analytical framework for identifying salient predictors within a high-dimensional feature space.

Participants

The study sample consisted of 4,134 primary school students selected from Türkiye's seven geographical regions. A stratified probability sampling strategy was employed, with each geographical region treated as a distinct stratum to ensure regional representativeness. Participants ranged in age from 7 to 11 years.

Türkiye is characterized by substantial regional diversity in terms of population density, socioeconomic conditions, and educational infrastructure. Ensuring national representativeness while maintaining feasibility in large-scale data collection poses both logistical and methodological challenges. To address these challenges, the sampling process prioritized not only numerical adequacy but also qualitative representativeness across regions.

Within each geographical region, primary schools located in provinces and districts identified as having moderate socioeconomic profiles were selected in collaboration with local educational authorities. This selection process considered field accessibility, institutional cooperation, and the sustainability of the data collection process. To enhance statistical reliability and support subgroup-level comparisons, a minimum of 300 student responses was collected from each region.

The determination of sample size was informed by methodological literature emphasizing the importance of adequate subgroup sizes in multivariate and predictive analyses (Memon et al., 2020; Althubaiti, 2022; Bujang et al., 2021). In addition, recent critiques of rigid sample size heuristics highlight the necessity of contextualized decision-making rather than reliance on

universal “rules of thumb” (Aguinis & Harden, 2024). Guided by these considerations, the final sample structure was designed to balance representativeness, analytical power, and practical feasibility.

Measures

School Bullying Scale

Students’ experiences of bullying were assessed using the *School Violence/Bullying Scale* originally developed by Çınkır and Karaman-Kepenekçi (2003), adapted for primary school students by Sarıgöl and Akar (2025) with the inclusion of a digital bullying dimension. The scale consists of 16 items across four subdimensions: physical bullying (4 items), verbal bullying (5 items), relational/emotional bullying (3 items), and digital bullying (4 items).

Items are rated on a 5-point Likert-type scale ranging from 1 (*Never*) to 5 (*Every day*), with higher scores indicating more frequent exposure to bullying. Sample items include: physical bullying (“Have you ever been hit or kicked at school?”), verbal bullying (“Have you been insulted or verbally harassed at school?”), relational bullying (“Have you ever been excluded from games or groups at school?”), and digital bullying (“Have hurtful things been said to you in WhatsApp groups or on social media platforms?”).

Confirmatory factor analysis (CFA) supported the structural validity of the scale (CMIN/DF = 1.87; RMSEA = .08; GFI = .85; AGFI = .80; CFI = .90; IFI = .91; RMR = .08), with all indices indicating acceptable model fit. Internal consistency was high, with a Cronbach’s alpha coefficient of .89, demonstrating strong reliability for use with primary school populations.

School Happiness Scale

School-based happiness was measured using the *School Happiness Scale* developed by Gündoğan and Akar (2019). The original version of the scale comprises nine items across two dimensions—happiness and unhappiness—and employs a 3-point Likert-type response format.

Both exploratory and confirmatory factor analyses supported the construct validity of the scale (CFI = 0.989; RMSEA = 0.026), and internal consistency reliability was reported as α = .76.

To enhance sensitivity for younger age groups, the scale was subsequently adapted by Akar and Ay (2025) to a 4-point Likert-type format ranging from 1 (*Never*) to 4 (*Always*). The present study employed this adapted version. In the current sample, the scale explained 66.86% of the total variance, and internal consistency reliability was satisfactory (Cronbach's α = .81).

Negatively worded items were reverse-coded prior to analysis. Total scores range from 9 to 36, with higher scores indicating higher levels of school happiness. Example items include "I look forward to going to school," "School is a happy place," and the reverse-coded item "I want school to end as soon as possible."

Personal Information Form

A *Personal Information Form* developed by the researchers was used to collect data on participants' individual, familial, behavioral, and school-contextual characteristics. The form was designed to systematically capture independent variables relevant to students' school happiness and to facilitate structured categorization within the predictive modeling framework.

The form included the following variable domains:

1. **Demographic variables:** gender, age, grade level, nationality, number of siblings, birth order, household size, and type of residence (urban/rural).
2. **Family-related variables:** maternal and paternal education levels, parental employment status, parental cohabitation status, household income level, and whether bullying experiences were reported to family members.
3. **Behavioral variables:** reading frequency, daily television viewing time, and daily mobile device/tablet use.

4. **Teacher-related factor:** perceived level of teacher intervention in response to bullying.

Data Analysis

Data from a nationally representative sample of 4,134 primary school students were used to model school happiness as a continuous outcome variable. The primary aim was to identify the relative importance of sociodemographic, familial, behavioral, and school-contextual predictors. Supervised machine learning approaches were employed to capture nonlinear relationships and complex interactions beyond the assumptions of linear models.

Prior to analysis, the dataset underwent quality screening to identify insufficient effort responses and non-recoverable data entry errors (Meade & Craig, 2012; Van den Broeck et al., 2005). After cleaning, all 4,134 observations were retained. Missing data were negligible; therefore, complete-case analyses were conducted.

School happiness was computed as the total score of a nine-item self-report scale with adequate internal consistency (Cronbach's $\alpha = .80$). Score distributions were approximately normal and free of floor or ceiling effects; thus, raw scores were used. To prevent data leakage, outcome-defining items and high-cardinality identifiers were excluded from the feature set.

Three model families were compared: linear models (Linear Regression, Ridge, Lasso, Elastic Net), tree-based ensemble models (Random Forest, Gradient Boosting, HistGradientBoosting), and instance-based methods (k-Nearest Neighbors, Support Vector Regression). All models were implemented within a leakage-safe pipeline, with preprocessing steps learned exclusively from training data (Pedregosa et al., 2011).

Model performance was evaluated using 10-fold cross-validation. Predictive accuracy was assessed using R^2, with MAE and RMSE reported as complementary metrics. Tree-based ensemble models consistently outperformed linear models, with the Gradient Boosting

Regressor achieving the highest performance and explaining approximately 23% of the variance in school happiness. This model was therefore selected for interpretation.

Results

Predictive Performance Across Models and Validation Strategies

Model comparison

Under the primary reporting standard of 10-fold cross-validation, the Gradient Boosting Regressor (GBR) demonstrated the strongest out-of-sample predictive performance. The GBR achieved the highest explained variance ($R^2 = 0.234$) while producing the lowest error estimates (MAE = 3.70; RMSE = 4.72). As illustrated in Figure 2 and detailed in Table 1, these results indicate a clear performance advantage of tree-based ensemble models over linear baseline models.

The Random Forest model yielded performance estimates closely comparable to those of the GBR ($R^2 = 0.232$; MAE = 3.69); however, the GBR exhibited a small yet consistent advantage in overall predictive stability. In contrast, the predictive capacity of linear models was notably limited. Both Linear Regression and RidgeCV models reached a performance plateau at approximately $R^2 \approx 0.199$, whereas the Lasso model—characterized by stronger regularization—exhibited substantially weaker performance ($R^2 = 0.061$).

Relative to the Dummy Baseline model, which generates mean-based predictions, the predictive gains achieved by the GBR were both substantial and statistically meaningful. This superiority was confirmed by non-overlapping confidence intervals across all performance metrics (see Supplementary Table S3a).

Model calibration

Calibration analyses indicated that tree-based ensemble models exhibited acceptable levels of calibration. The calibration slope for the GBR was estimated at 1.13, while the Random Forest

model demonstrated an almost proportional calibration slope (0.98; see Supplementary Table S6b). These findings suggest that both models effectively captured variability in students' school happiness scores without exhibiting substantial overfitting.

Decile-based calibration analyses further supported these conclusions, revealing a high degree of correspondence between observed and predicted mean happiness scores for both ensemble models (see Supplementary Tables S7a–S7b).

Comparison of validation regimes

Adjusted repeated-measures t-tests with Benjamini–Hochberg false discovery rate (BH–FDR) correction ($\alpha = 0.05$) revealed no statistically significant differences in model performance across alternative validation regimes (all adjusted p-values ≥ 0.326; see Supplementary Table S4). This result provides statistical justification for retaining 10-fold K-Fold cross-validation as the primary reporting standard.

Complementary linear model findings

Complementary ordinary least squares (OLS) estimates presented in the supplementary materials (see Supplementary Table S13) were consistent with the patterns observed in the ensemble-based analyses. Specifically, exposure to peer bullying and mobile phone use emerged as negative predictors of school happiness, whereas teacher intervention and reading frequency were identified as positive predictors.

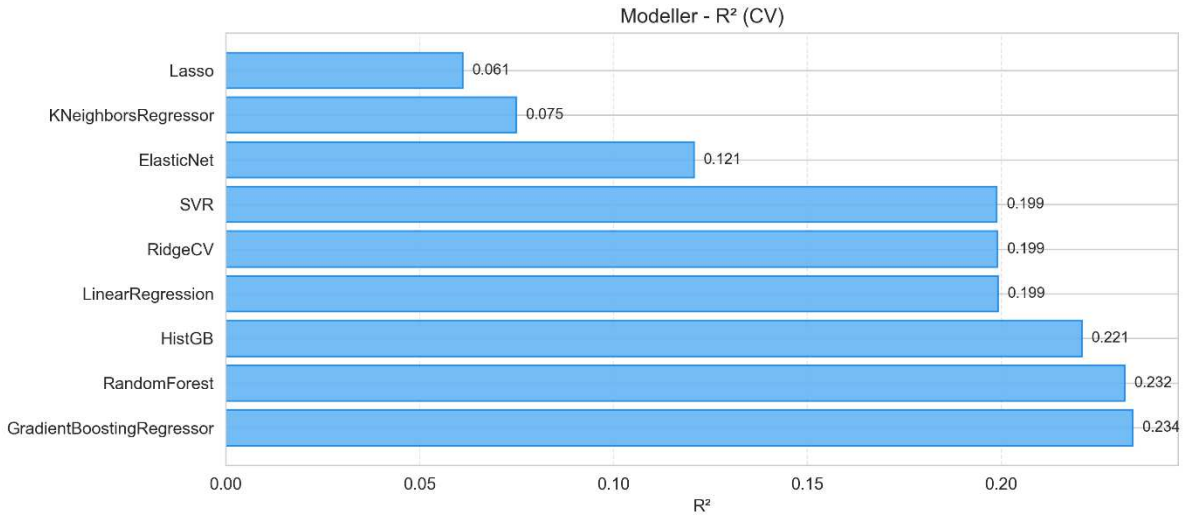


Figure 1

Distribution of Cross-Validated R² Scores Across Models

Table 1

Cross-Validated Model Performance (10-Fold K-Fold with Unified Leakage-Safe Preprocessing)

model	**R2_CV**	**MAE_CV**	**RMSE_CV**	**R2_train**	**MAE_train**	**RMSE_train**	**TrainingTime**
GradientBoostingRegressor	0.233983	3.70049	4.719097	0.3322	3.473809	4.414135	0.253869
RandomForest	0.2318	3.69399	4.72572	0.8962	1.350795	1.740051	0.799382
HistGB	0.2208	3.71925	4.758605	0.5548	2.831234	3.604194	0.272422
LinearRegression	0.1992	3.79193	4.823769	0.2099	3.771785	4.801402	0.006133
RidgeCV	0.1990	3.7948	4.824514	0.2098	3.773315	4.801602	0.008756
SVR	0.1989	3.76416	4.825898	0.2868	3.4468	4.561751	0.291988
ElasticNet	0.1207	4.03014	5.056126	0.124	4.025483	5.054082	0.005693
KNeighborsRegressor	0.0749	4.05577	5.185018	0.3927	3.277455	4.209372	0.007822
Lasso	0.0610	4.17668	5.224954	0.0639	4.174622	5.226094	0.006302

Explainability Analysis

To enhance the interpretability of model predictions, the Gradient Boosting Regressor was selected as the primary explanatory model, while the Random Forest model was examined as a

complementary reference. Feature importance profiles derived from both models showed substantial overlap, indicating a stable and consistent hierarchy of predictors.

Across both models, peer bullying emerged as the strongest and most consistently negative predictor of school happiness. This was followed by gender, monthly reading frequency, and daily mobile device use. Higher reading frequency and stronger teacher intervention were associated with higher predicted levels of school happiness, whereas prolonged mobile device use was associated with lower happiness scores.

The regional variable exhibited a context-sensitive, moderate effect on school happiness predictions. Parental education displayed a non-linear pattern, with students whose parents had moderate levels of education reporting relatively higher happiness, while the contribution of very high parental education levels appeared more limited. In contrast, household income, household size, and sibling order demonstrated weak and unstable contributions across models.

Overall, the explainability analyses indicate that behavioral and psychosocial factors play a more prominent role in predicting school happiness than relatively stable sociodemographic characteristics. In particular, negative social experiences and daily interaction patterns emerged as dominant contributors to students' predicted levels of school happiness.

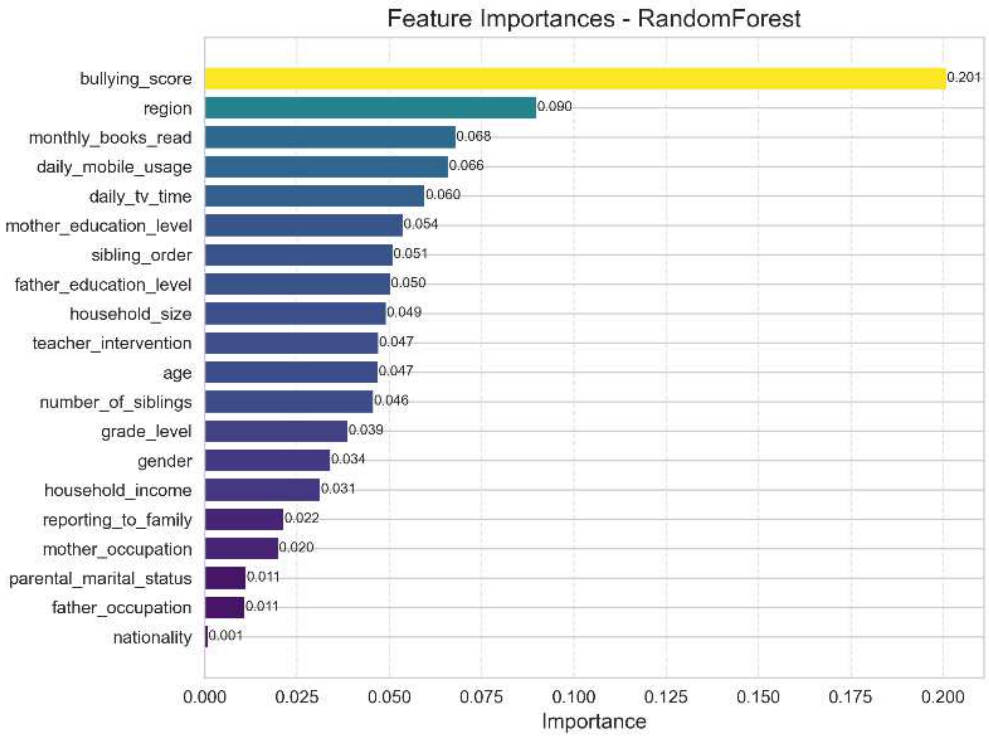


Figure 2

Raw feature importance (Random Forest).

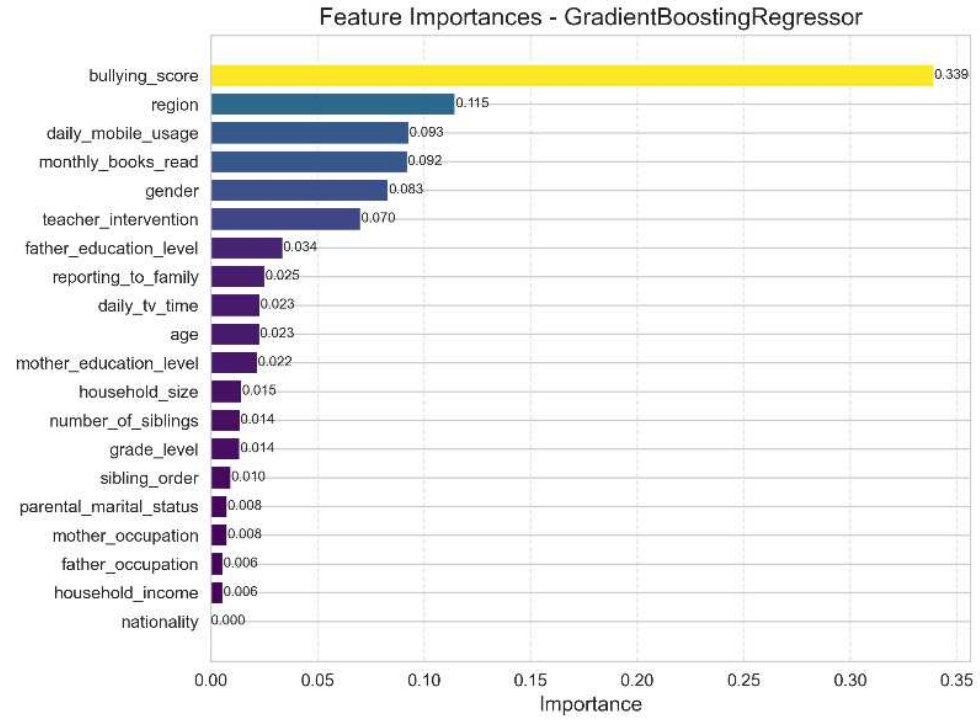


Figure 3

Raw feature importance (GradientBoostingRegressor).

Combined (Consensus) Feature Importance Index

To integrate the relative importance assigned to predictors across different models, feature importance estimates derived from the Gradient Boosting Regressor and the Random Forest models were normalized and combined to construct a model-agnostic consensus ranking. This combined index revealed a high degree of stability in the hierarchy of predictors across ensemble-based modeling approaches.

The resulting ranking exhibited a three-tier structure. Peer bullying emerged as the strongest negative determinant of school happiness, occupying the top tier of the consensus index. The second tier included monthly reading frequency, geographical region, daily mobile device use, gender, and teacher intervention, all of which demonstrated comparatively strong contributions to school happiness predictions. The third tier comprised parental education, age, and reporting bullying to family members, which showed moderate yet consistent effects across models.

In contrast, socioeconomic background indicators such as household income, household size, and parental occupation displayed limited contributions to school happiness predictions. Detailed importance values and stability analyses for all predictors are provided in the supplementary materials.

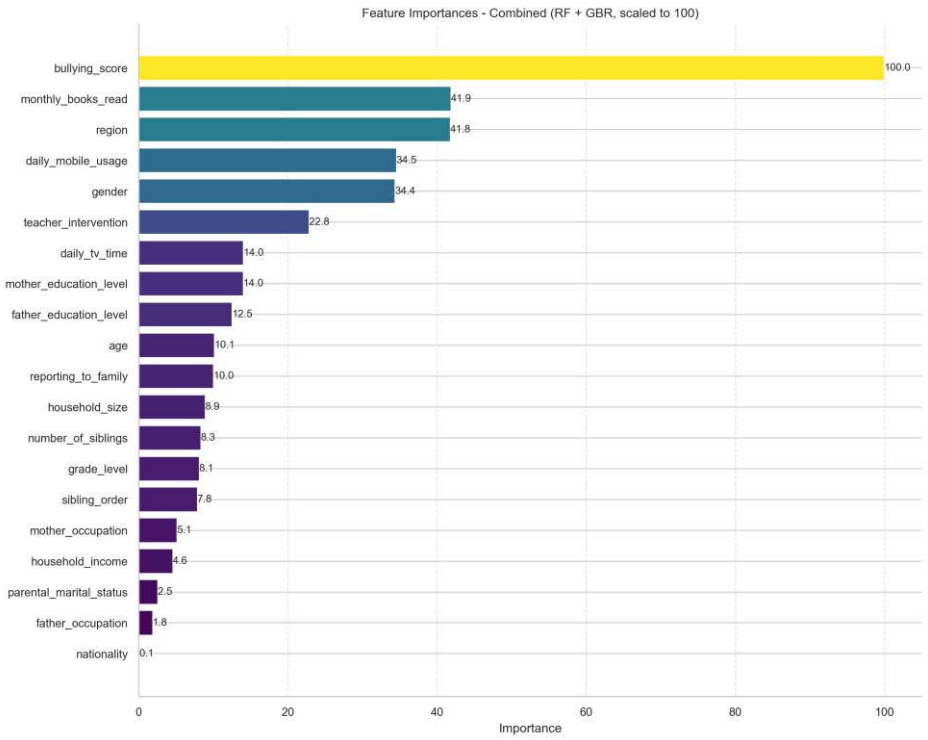


Figure 4

Combined normalized feature importance index across RF and

Interpretation of Model Findings

The modest differences observed between the Gradient Boosting Regressor and Random Forest models reflect their underlying algorithmic structures. Gradient Boosting is more sensitive to subtle nonlinear patterns, whereas Random Forest emphasizes broader behavioral variance. Combining feature importance estimates from both models into a consensus index balances these differences and yields a stable, interpretable hierarchy of predictors.

Substantively, peer bullying and digital screen use emerged as the strongest negative predictors of school happiness, while teacher support and reading habits functioned as protective factors associated with higher predicted happiness. Parental education and family communication contributed at a secondary level, whereas the explanatory power of relatively static sociodemographic characteristics remained limited. This overall pattern suggests that school

happiness is shaped primarily by behavioral and psychosocial experiences rather than fixed background attributes.

To mitigate limitations associated with impurity-based feature importance measures, model interpretations were further examined using SHAP (SHapley Additive exPlanations). SHAP analyses provided model-agnostic validation by quantifying both the direction and magnitude of each predictor's contribution to model predictions.

Model-Agnostic Explainability via SHAP

SHAP summary and importance plots closely aligned with impurity-based rankings, indicating a highly consistent explanatory structure. Peer bullying emerged as the most influential negative contributor, with higher levels associated with marked decreases in predicted happiness. Male gender was associated with systematically lower predicted happiness levels. Reading frequency showed a positive association, whereas higher levels of daily mobile device use were linked to reduced happiness. Regional effects were moderate and context-sensitive, while teacher intervention consistently contributed positively, functioning as a stabilizing factor across the distribution.

Overall, SHAP-based findings confirm that behavioral and psychosocial indicators dominate the prediction of school happiness, while static sociodemographic variables play a more modest role. The convergence of results across multiple explainability approaches strengthens the robustness and interpretability of the identified predictive patterns.

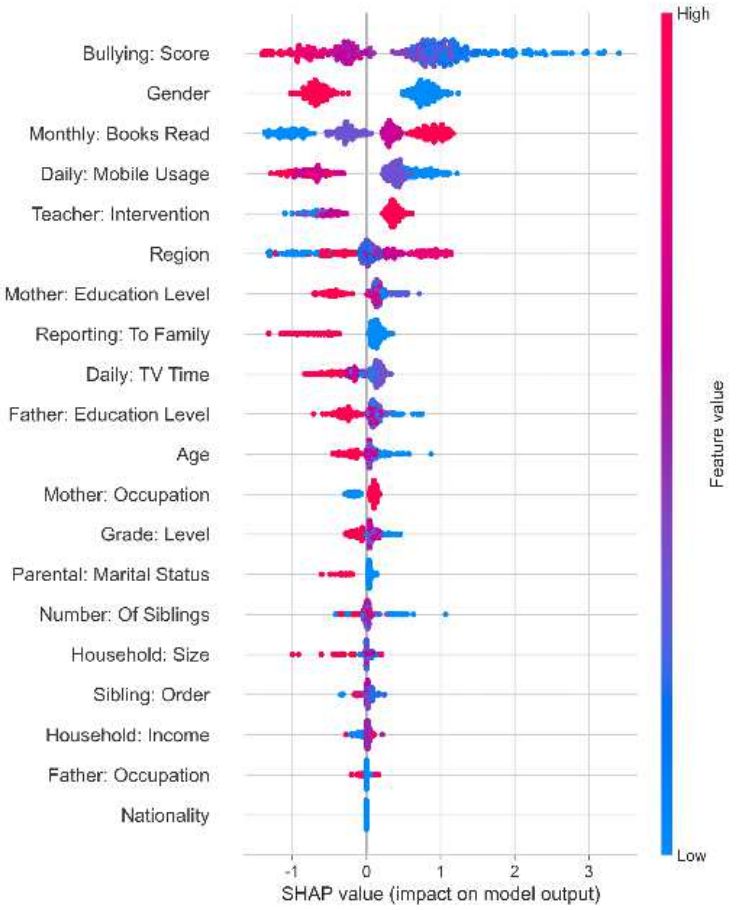


Figure 5

Global SHAP summary plot for the Gradient Boosting Regressor.

Each point represents an individual observation's contribution to the model prediction, with color indicating the corresponding feature value (red = higher, blue = lower). The position along the x-axis reflects the SHAP value, capturing both the direction and magnitude of each feature's effect on predicted school happiness. Features are ordered by mean absolute SHAP values, with higher-ranked variables exerting stronger global influence.

Positive SHAP values indicate higher predicted happiness, whereas negative values indicate lower predicted happiness. Peer bullying emerged as the most influential negative contributor, followed by gender, monthly reading frequency, daily mobile device use, and teacher

intervention. This pattern closely mirrors impurity-based feature importance rankings, providing convergent evidence for the robustness of the identified predictor hierarchy.

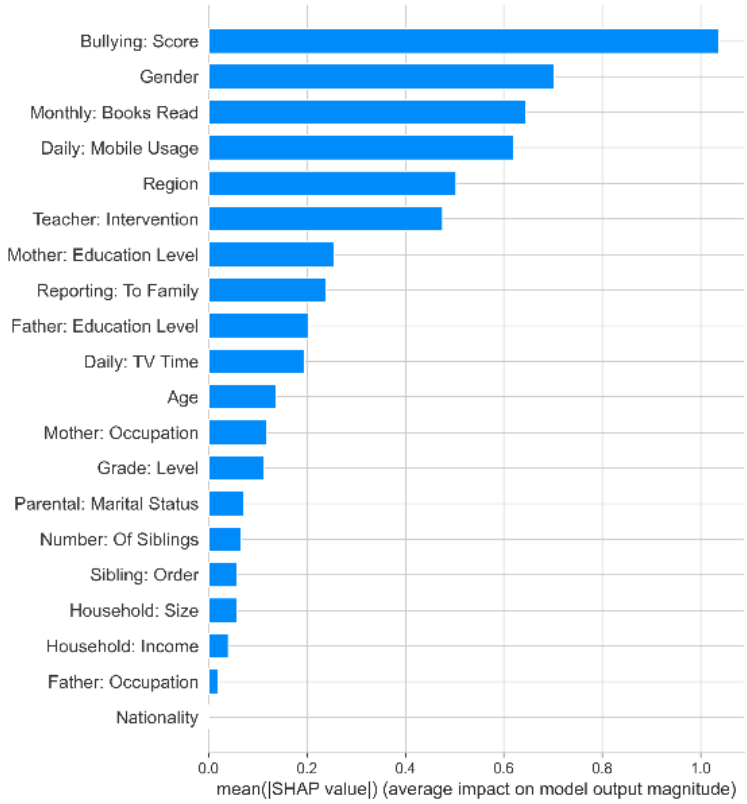


Figure 6

Mean absolute SHAP (|SHAP|) bar plot summarizing global feature importa

The bars display the mean absolute SHAP values across the full sample, summarizing each predictor's overall contribution to the model output regardless of effect direction. Larger bars indicate stronger global influence on predicted school happiness.

Consistent with the beeswarm visualization, peer bullying emerged as the most influential predictor, followed by gender, monthly reading frequency, daily mobile device use, and teacher intervention. The close alignment between this ranking and impurity-based feature importance measures confirms the stability of the findings across explainability approaches.

To further examine functional form and directionality, SHAP dependence plots were generated for the most influential predictors (Figures 7 and 8), revealing distinct behavioral and developmental patterns associated with systematic changes in predicted school happiness.

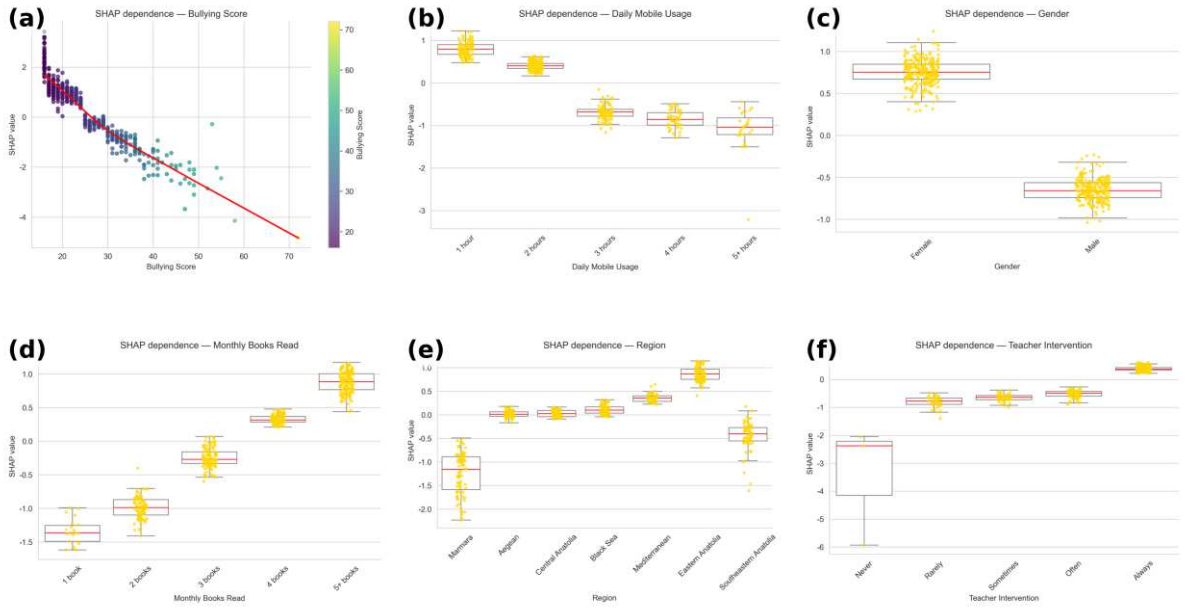


Figure 7

SHAP dependence plots for the top six predictors of student happiness in the Gradient Boosting Regressor model

Interpretation of SHAP Dependence Plots

(a) Peer Bullying Score exhibited a pronounced threshold effect. For students with bullying scores above approximately 50, SHAP values consistently dropped below −2, indicating sharp declines in predicted school happiness. This pattern confirms peer bullying as a dominant behavioral risk factor with a strong nonlinear impact on school happiness predictions.

(b) Gender displayed a clear categorical separation. Male students clustered around SHAP values of approximately −1.0, whereas female students were concentrated around positive SHAP values (approximately +0.6). This pattern indicates systematically lower predicted levels of school happiness among male students.

(c) Monthly reading frequency demonstrated an almost linear positive gradient. Students who reported reading one book per month showed SHAP values around −1.2, while those reading four or more books per month exhibited SHAP values approaching +1.0. This pattern suggests cumulative benefits of reading frequency for school happiness predictions.

(d) Daily mobile device use revealed a nonlinear and negative pattern. Moderate use (1–2 hours per day) was associated with largely neutral SHAP values, whereas usage exceeding three hours per day shifted SHAP values below −1.5. This finding indicates that prolonged screen exposure functions as a behavioral risk factor for reduced school happiness.

(e) Geographical region exhibited moderate dispersion with context-sensitive effects. Students from western and southern regions tended to show relatively positive SHAP values ($\Delta \approx +0.3$ to +0.6), whereas those from eastern regions clustered more frequently in the negative range ($\Delta \approx -0.4$). This pattern suggests regional contextual differences in predicted school happiness.

(f) Teacher intervention emerged as a balancing and protective factor. Students reporting frequent teacher intervention showed SHAP values above +0.8, while those indicating "never" or "rarely" clustered below zero. This pattern suggests that effective teacher involvement may mitigate the negative effects of adverse school experiences.

Additional SHAP dependence plots for secondary predictors—including parental education, reporting bullying to family members, daily television viewing time, household income, and household size—are presented in Supplementary Figure S14 and Figure 9. Although these variables exhibited relatively smaller absolute SHAP magnitudes (generally mean $|\text{SHAP}| < 0.3$), their directional effects remained consistent and supported the behavioral–psychosocial hierarchy identified in the primary analyses.

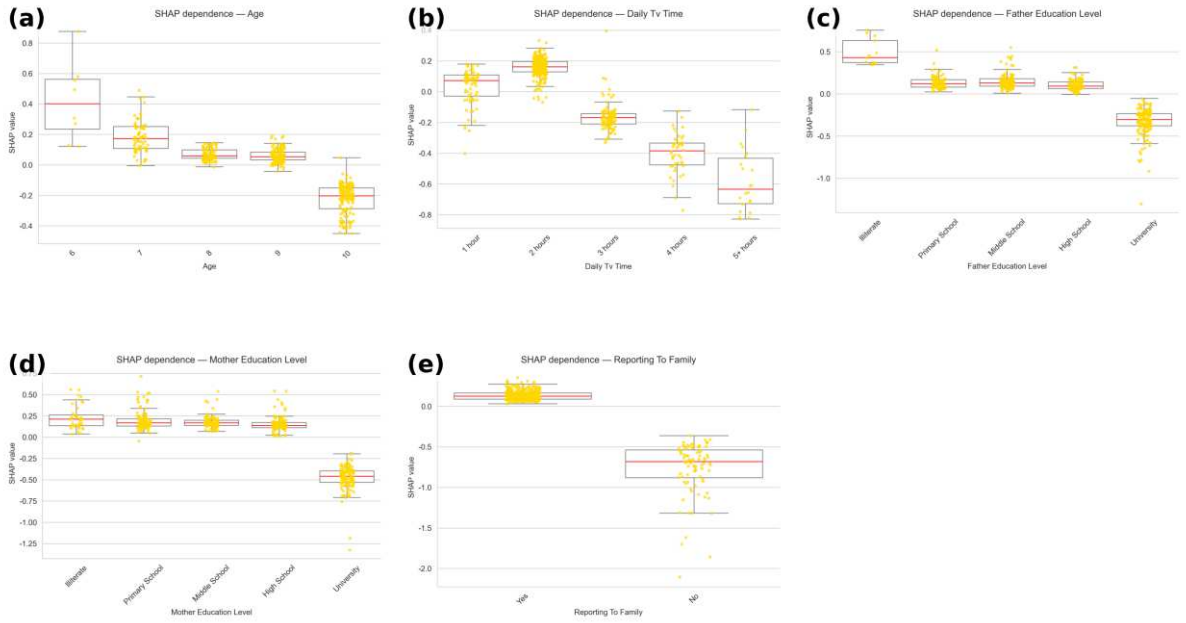


Figure 8

SHAP dependence plots for secondary predictors of student happiness in the Gradient Boosting Regressor model

This figure presents SHAP dependence relationships for five secondary predictors that exhibit moderate yet directionally consistent contributions to the prediction of school happiness: (a) age, (b) daily television viewing time, (c) paternal education level, (d) maternal education level, and (e) reporting experiences to family members. Each subplot reflects the marginal effect of the corresponding predictor on the model output (SHAP value) while holding other predictors constant. Yellow points represent individual observations, boxplots summarize the distributions, and red lines indicate median values.

(a) Age displayed a mild but monotonic decreasing trend. Younger students (ages 6–7) exhibited positive SHAP values (median ≈ +0.3–0.4), indicating higher predicted happiness, whereas older students (age 10) clustered around near-zero or negative SHAP values (median ≈ −0.2). This pattern suggests a developmental decline in predicted well-being during middle childhood.

(b) Daily television viewing time demonstrated a nonlinear negative pattern. Moderate viewing durations (1–2 hours per day) were associated with neutral or slightly positive SHAP values, whereas viewing times exceeding three hours per day shifted SHAP values markedly into the negative range (median ≈ −0.5 to −0.7). This finding is consistent with cumulative effects of prolonged screen exposure.

(c) Paternal education level exhibited an inverted U-shaped pattern. Both low education levels (illiterate) and the highest level (university degree) were associated with negative SHAP values, while middle school and high school levels showed neutral or slightly positive contributions. This pattern indicates a nonlinear moderating role of paternal education.

(d) Maternal education level showed a similar but attenuated pattern. University-level maternal education was associated with lower SHAP values (median ≈ −0.7), whereas primary and lower secondary education levels corresponded to slight positive SHAP shifts. This pattern may reflect differential parental expectations and interaction dynamics.

(e) Reporting experiences to family members revealed a clear categorical separation. Students who reported sharing their problems or experiences with family members clustered around SHAP ≈ 0 (baseline), whereas those who did not report such experiences showed SHAP values concentrated below −1.0. This pattern indicates that the absence of family communication is associated with a substantial decrease in predicted school happiness.

Overall, the absolute SHAP magnitudes of these secondary predictors (mean |SHAP| ≈ 0.30–0.36) were smaller than those of the dominant behavioral predictors (peer bullying, mobile device use, reading frequency, and teacher intervention). Nevertheless, the consistency of effect directions suggests that psychosocial and communicative factors contribute meaningfully—albeit at a moderate level—to subjective school happiness and reinforce the hierarchical explanatory structure of the model.

Discussion

Children's happiness has increasingly become a central focus of research and policy agendas in education and psychology (The Children's Society, 2024; UNICEF Innocenti, 2025). This emphasis reflects a growing consensus that subjective well-being is closely linked not only to individual health and academic achievement but also to the sustainability of societal well-being (Diener et al., 2017). Primary school years constitute a critical developmental period in which self-concept, social relationships, and the cognitive–affective foundations of happiness are formed (Holder & Coleman, 2009). The findings of the present study support this framework by showing that school happiness among primary school students is largely shaped by peer relationships, perceived safety, and everyday behavioral patterns.

International organizations increasingly conceptualize school happiness as a core educational outcome and a measurable policy domain. The OECD positions students' life satisfaction as a key indicator of educational quality (OECD, 2019), while UNESCO's Happy Schools framework highlights social connectedness, safety, pedagogical support, and participation as foundational dimensions of quality education (UNESCO, 2016, 2024). In this context, the prominence of peer bullying and teacher intervention in the present findings closely aligns with the core pillars of the "happy school" approach.

Peer Bullying and Safety: A Central Vulnerability in School Happiness

The most salient finding of this study is the strong negative association between peer bullying and school happiness. School happiness is widely conceptualized as a multidimensional form of subjective well-being that encompasses positive affect, life satisfaction, and context-specific domains such as school belonging and perceived social support (Huebner, 1991; Huebner et al., 2006; Veenhoven, 2019). Within this framework, bullying directly undermines core components of well-being by eroding students' sense of safety and belonging within the school environment. Characterized by intentionality, repetition, and power imbalance (Olweus, 1993;

Bradshaw et al., 2015), bullying compromises students' sense of security, trust in peers, and capacity to form positive social relationships.

The identification of bullying as the primary risk factor is consistent with international and national evidence (Aunampai et al., 2022; Akar & Ay, 2025). The expansion of cyberbullying in digitalized contexts further intensifies risks to children's well-being (Hinduja & Patchin, 2014; Kowalski et al., 2014; Juvonen & Gross, 2008). Accordingly, bullying should be understood as an ecological phenomenon shaped by interactions among individual, familial, and school-level factors (Arseneault et al., 2010; Espelage, 2014; Hong et al., 2016). From a school happiness perspective, bullying prevention represents a central leverage point for policy and practice.

Digital Device Use and Daily Life Patterns

The negative association between mobile device use and school happiness aligns with research linking excessive screen exposure to emotional and psychosocial difficulties, particularly during sensitive developmental periods (Twenge & Campbell, 2018; Burhan & Moradzadeh, 2020). Rather than a purely technological variable, digital device use appears to function as a broader indicator of daily life regulation intersecting with sleep, physical activity, and face-to-face social interaction. Moreover, prolonged digital use may indirectly increase exposure to online risks, including cyberbullying (Hinduja & Patchin, 2014; Kowalski et al., 2014), underscoring the need for balanced and developmentally sensitive guidance frameworks.

Reading Habits and Protective Cultural Practices

Reading frequency emerged as a protective factor for school happiness, highlighting the role of everyday cultural practices in supporting well-being. Beyond academic achievement, reading is associated with attention regulation, self-control, intrinsic motivation, and emotional functioning. Empirical evidence linking reading to higher happiness levels supports this

interpretation (Burhan & Moradzadeh, 2020), suggesting that interventions should complement risk reduction with the promotion of protective daily practices.

Teacher Intervention and the "Happy School" Climate

Teacher intervention was identified as a robust protective factor, consistent with school climate research and UNESCO's emphasis on pedagogical support and positive learning environments (UNESCO, 2016, 2024). Teachers' responses to bullying and classroom social dynamics foster students' perceptions of fairness, acceptance, and belonging. Experimental evidence indicates that social–emotional learning programs enhance both cognitive development and school happiness (Schonert-Reichl et al., 2015), highlighting teacher capacity building as a key implementation domain.

Demographic and Socioeconomic Context

Consistent with prior research, girls reported higher predicted happiness levels than boys (Çankaya & Meydan, 2018), while increasing age was associated with modest declines in school happiness (Goldbeck et al., 2007). Although family socioeconomic resources may indirectly shape well-being (Bradley & Corwyn, 2002; Eccles & Roeser, 2011; Evans & Cassells, 2014), the present findings suggest that psychosocial and behavioral factors play a more dominant role than relatively static background characteristics. This pattern underscores the importance of equity-oriented school policies that target relational and environmental conditions.

Methodological Contribution

School happiness reflects a multidimensional system shaped by interrelated individual, familial, and institutional contexts (Ben-Arieh & Frønes, 2011; UNICEF, 2021). In such contexts, machine learning—particularly tree-based ensemble models—offers clear advantages in capturing nonlinear relationships and ranking predictor importance (Goldstein et al., 2017;

Hilbert, 2021; Weller et al., 2021). The scarcity of nationally representative, primary school–level studies in Türkiye integrating traditional and digital bullying within an explainable machine learning framework further strengthens the contribution of this research. Given that official well-being statistics largely focus on adults (TÜİK, 2024), the present findings provide timely evidence on children's subjective well-being.

Conclusion

Using data from a nationally representative sample of 4,134 primary school students, this study examined the predictors of school happiness through supervised machine learning regression models. The findings demonstrate that school happiness is not a random or idiosyncratic outcome but can be predicted with moderate accuracy based on behavioral and psychosocial variables.

Comparative analyses revealed that tree-based ensemble models outperformed linear models in out-of-sample prediction. The Gradient Boosting Regressor achieved the highest predictive performance, explaining approximately 23% of the variance in school happiness ($R^2 = 0.234$; MAE = 3.70; RMSE = 4.72), with the Random Forest model yielding comparable results. Diagnostic and calibration analyses indicated that this predictive performance was not driven by overfitting and that model estimates were generalizable.

Across models, peer bullying emerged as the strongest negative predictor of school happiness, whereas reading frequency and teacher intervention functioned as protective factors. Excessive mobile device use—particularly at higher levels—was associated with reduced happiness, while demographic and socioeconomic variables contributed more modestly. An explained variance of approximately 23% represents a meaningful level of prediction for a complex subjective well-being construct.

Overall, the findings suggest that school happiness is shaped more by students' social experiences and daily behavioral patterns than by fixed background characteristics. From a policy perspective, bullying prevention, teacher capacity building, and balanced digital media use emerge as critical leverage points for school-based interventions. This study provides a strong empirical foundation for evidence-based educational policies aimed at enhancing school happiness in the Turkish context.

Implications and Recommendations

1. **Anti-bullying initiatives should be a central school priority.** Preventive, whole-school programs and safe reporting mechanisms should complement disciplinary approaches.
2. **Teachers' intervention capacity should be systematically strengthened.** Professional development focusing on classroom climate, peer dynamics, and bullying prevention is essential.
3. **Balanced digital media guidance should be promoted.** Schools should provide age-appropriate guidance for students and families on healthy digital habits.
4. **Reading-oriented school cultures should be supported.** Structured reading times, accessible libraries, and reading-based social activities can enhance well-being.
5. **Family–school communication should be reinforced.** Trust-based and regular communication channels can support students' emotional well-being.
6. **School happiness should be monitored as an educational outcome.** Well-being indicators should complement academic metrics in educational evaluation systems.

Limitations and Future Research

Despite its strengths, the study is subject to several limitations. The cross-sectional design precludes causal inference, and although multiple predictors were included, not all potential determinants could be examined simultaneously (Creswell & Plano Clark, 2011). Future research should employ longitudinal designs, multiple data sources (e.g., teacher and parent reports, observational measures), and comparative modeling approaches to further elucidate the determinants of school happiness.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2013) and national ethical guidelines. Ethical approval was obtained from the relevant university ethics committee (Approval No: …). Parental consent and children's informed assent were secured, participation was voluntary, and all data were anonymized. Particular care was taken when addressing bullying-related items, and students requiring psychosocial support were referred to school counseling services.

Conflict of Interest

The authors declare no conflict of interest. The study was conducted independently without external financial support.

References

Akar, C., & Ay, Ö. (2025). The impact of bullying on happiness at primary school: The role of sociodemographic and behavioral variables. *Participatory Educational Research, 12*(4), 229–250. https://doi.org/10.17275/per.25.58.12.4

Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and mental health problems: "Much ado about nothing"? *Psychological Medicine, 40*(5), 717–729. https://doi.org/10.1017/S0033291709991383

Aunampai, A., Techataweewan, W., & Suwanmonkha, S. (2022). Bullying victimization and happiness among Thai primary school students. *Children and Youth Services Review, 136,* 106393. https://doi.org/10.1016/j.childyouth.2022.106393

Baiocco, R., Laghi, F., Di Norcia, A., & Cacioppo, M. (2019). Happiness in children and early adolescents: The role of self-esteem, self-concept, and loneliness. *Journal of Happiness Studies, 20*(5), 1681–1699. https://doi.org/10.1007/s10902-018-0005-0

Ben-Arieh, A., & Frønes, I. (2011). Taxonomy for child well-being indicators: A framework for the analysis of the well-being of children. *Childhood, 18*(4), 460–476. https://doi.org/10.1177/0907568211398159

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371–399. https://doi.org/10.1146/annurev.psych.53.100901.135233

Bradshaw, C. P., Sawyer, A. L., & O'Brennan, L. M. (2015). A social disorganization perspective on bullying-related attitudes and behaviors: The influence of school context. *American Journal of Community Psychology, 56*(3), 359–372. DOI: 10.1007/s10464-009-9240-1

Burhan, R. & Moradzadeh, J. (2020). The role of media consumption in adolescent mental health. Journal of Adolescence, 85, 65-77. https://doi.org/10.1016/j.adolescence. 2020.09.006.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). SAGE Publications.

Çankaya, İ. (2011). İlköğretimde akran zorbalığı [Bullying among peers in primary school]. Uludağ Üniversitesi Eğitim Fakültesi Dergisi, 24(1), 81-92.

Çankaya, Z. C., & Meydan, B. (2018). Ergenlik döneminde mutluluk ve umut [Happiness and hope in adolescence]. Elektronik Sosyal Bilimler Dergisi, 17(65), 207-222. https://doi.org/10.17755/esosder.316977.

Diener, E., Oishi, S., & Lucas, R. E. (2017). If, why, and when subjective well-being influences health, and future needed research. *Applied Psychology: Health and Well-Being, 9*(2), 133–167. https://doi.org/10.1111/aphw.12090

Dilmaç, B., & Özkan, C. (2019). Lise öğrencilerinde öznel mutluluk, suçluluk ve utancın yordayıcısı olarak siber zorbalık [Cyberbullying as a predictor of subjective happiness, guilt, and shame in high school students]. Türk Eğitim Bilimleri Dergisi, 17(1), 195-212.

Espelage, D. L. (2014). *Ecological Theory: Preventing Youth Bullying, Aggression, and Victimization*. Theory Into Practice, 53(4), 257–264. https://doi.org/10.1080/00405841.2014.947216

Eccles, J. S., & Roeser, R. W. (2011). Schools as developmental contexts during adolescence. *Journal of Research on Adolescence, 21*(1), 225–241. https://doi.org/10.1111/j.1532-7795.2010.00725.x

Evans, G. W., & Cassells, R. C. (2014). Childhood poverty, cumulative risk exposure, and mental health in emerging adults. *Clinical Psychological Science, 2*(3), 287–296. https://doi.org/10.1177/2167702613501496

Gini, G., Albiero, P., Benelli, B., & Altoè, G. (2007). Does empathy predict adolescents' bullying and defending behavior? *Aggressive Behavior, 33*(5), 467–476. https://doi.org/10.1002/ab.20204

Goldstein, B. A., Navar, A. M., & Carter, R. E. (2017). Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic

challenges. *European Heart Journal, 38*(23), 1805–1814. https://doi.org/10.1093/eurheartj/ehw302

Gómez-Baya, D., García-Moro, F. J., Muñoz-Silva, A., & Martín-Romero, N. (2021). School satisfaction and happiness in 10-year-old children from seven European countries. *Children, 8*(5), 370. https://doi.org/10.3390/children8050370

Goldbeck, L., Schmitz, T. G., Besier, T., Herschbach, P., & Henrich, G. (2007). Life satisfaction decreases during adolescence. *Quality of Life Research, 16*(6), 969–979. https://doi.org/10.1007/s11136-007-9205-5

Göcen, G. (2015). 11-12 Yaş grubundaki çocukların minnettarlıkları ve hayat memnuniyetlerine etki eden aile ile ilgili faktörler [Family-related factors affecting gratitude and life satisfaction in children of a specific age group]. Değerler Eğitimi Dergisi, 13(29), 83-116.

Gündoğan, A., & Akar, C. (2019). *İlkokul öğrencileri için okulda mutluluk ölçeği: Geçerlilik ve güvenirlik çalışması* [The happiness at school scale for elementary school students: A validity and reliability study]. *Türk Akademik Yayınlar Dergisi (Turkish Academic Publications Journal), 3*(1), 61-75.

Hilbert, S., Coors, S., Kraus, E., Bischl, B., Lindl, A., Frei, M., Wild, J., Krauss, S., Goretzko, D., & Stachl, C. (2021). Machine learning for the educational sciences. *Review of Education, 9*(3), e3310. https://doi.org/10.1002/rev3.3310

Hinduja, S., & Patchin, J. W. (2014). *Bullying beyond the schoolyard: Preventing and responding to cyberbullying* (2nd ed.). Corwin Press.

Holder, M. D., & Coleman, B. (2009). The contribution of social relationships to children's happiness. *Journal of Happiness Studies, 10*(3), 329–349. https://doi.org/10.1007/s10902-007-9083-0

Hong, J. S., Lee, J., Espelage, D. L., Hunter, S. C., Patton, D. U., & Rivers, T. (2016). Understanding the correlates of face-to-face and cyberbullying victimization among U.S. adolescents: A social-ecological analysis. *Violence and Victims, 31*(4), 638–653. https://doi.org/10.1891/0886-6708.VV-D-15-00014

Huebner, E. S. (1991). Initial development of the Student's Life Satisfaction Scale. *School Psychology International, 12*(3), 231–240. https://doi.org/10.1177/0143034391123010

Huebner, E. S., Seligson, J. L., Valois, R. F., & Suldo, S. M. (2006). A review of the Brief Multidimensional Students' Life Satisfaction Scale. *Social Indicators Research, 79*(3), 477–484. https://doi.org/10.1007/s11205-005-5395-9

Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? Bullying experiences in cyberspace. *Journal of School Health, 78*(9), 496–505. https://doi.org/10.1111/j.1746-1561.2008.00335.x

Kepenekçi, Y. K., & Çınkır, Ş. (2003). Öğrenciler arası zorbalık [Bullying among students]. *Kuram ve Uygulamada Eğitim Yönetimi, 34*, 236–253.

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin, 140*(4), 1073–1138. https://doi.org/10.1037/a0035618

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455. https://doi.org/10.1037/a0028085

OECD. (2019). *PISA 2018 results (Volume III): What school life means for students' lives*. https://doi.org/10.1787/acd78851-en

Olweus, D. (1993). *Bullying at school: What we know and what we can do.* Blackwell.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., VanderPlas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Sarıgöl, M., & Akar, C. (2025). İlkokul öğrencilerinin zorbalığa maruz kalma düzeyleri ve demografik/kişisel faktörlerle ilişkisi [Primary school students' exposure to bullying and its relationship with demographic/personal factors]. *Uluslararası Liderlik Eğitimi Dergisi*, *9*(1). 1-20.

Schonert-Reichl, K. A., Oberle, E., Lawlor, M. S., Abbott, D., Thomson, K., Oberlander, T. F., & Diamond, A. (2015). Enhancing cognitive and social–emotional development through a simple-to-administer mindfulness-based school program for elementary school children: A randomized controlled trial. *Developmental Psychology, 51*(1), 52–66. https://doi.org/10.1037/a0038454

The Children's Society. (2024). *The Good Childhood Report 2024.* The Children's Society. https://www.childrenssociety.org.uk/good-childhood

Türkiye İstatistik Kurumu. (2024). *Yaşam memnuniyeti araştırması, 2024.* https://data.tuik.gov.tr

Twenge, J. M., & Campbell, W. K. (2018). Associations between screen time and lower psychological well-being among children and adolescents: Evidence from a population-based study. *Preventive Medicine Reports, 12*, 271–283. https://doi.org/10.1016/j.pmedr.2018.10.003

UNESCO. (2016). *Happy schools! A framework for learner well-being.* https://unesdoc.unesco.org/

UNESCO. (2024). *Why the world needs happy schools: Global report on happiness in and for learning*. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000389119

UNICEF Innocenti. (2025). *Child well-being in an unpredictable world (Innocenti Report Card 19).* https://www.unicef.org/innocenti/reports/child-well-being-in-an-unpredictable-world

UNICEF. (2021). *The State of the World's Children 2021: On my mind—Promoting, protecting and caring for children's mental health*. https://www.unicef.org/reports/state-worlds-children-2021

Uusitalo-Malmivaara, L. (2012). Global and school-related happiness in Finnish children. *Journal of Happiness Studies, 13*(4), 601–619. https://doi.org/10.1007/s10902-011-9282-6

Van den Broeck, J., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLOS Medicine, 2*(10), e267. https://doi.org/10.1371/journal.pmed.0020267

Weller, D. L., Love, T. M. T., & Wiedmann, M. (2021). Interpretability versus accuracy: A comparison of machine learning models built using different algorithms, performance measures, and features to predict *E. coli* levels in agricultural water. *Frontiers in Artificial Intelligence, 4,* Article 628441. https://doi.org/10.3389/frai.2021.628441

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplementary.docx