

Predicting High School Dropout Risk from Academic Engagement, Attendance, Self-Efficacy, and Family Support Using TabNet

Daniel. Farrugia¹, Camille. Brossard^{2*}, Rodrigo. Villagra³

¹ Department of Psychology, University of Malta, Msida, Malta

² Department of Cognitive Psychology, Sorbonne University, Paris, France

³ Department of Psychology, National University of Asunción, San Lorenzo, Paraguay

* Corresponding author email address: camille.brossard@sorbonne-universite.fr

Article Info

Article type:

Original Research

How to cite this article:

Farrugia, D., Brossard, C., & Villagra, R. (2026). Predicting High School Dropout Risk from Academic Engagement, Attendance, Self-Efficacy, and Family Support Using TabNet. *Journal of Adolescent and Youth Psychological Studies*, 7(9), 1-20.

<http://dx.doi.org/10.61838/kman.jayps.5684>



© 2026 the authors. Published by KMAN Publication Inc. (KMANPUB), Ontario, Canada. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

ABSTRACT

Objective: This study aimed to develop and evaluate an interpretable TabNet model for predicting verified high school dropout among French upper-secondary students using academic engagement, attendance, academic self-efficacy, and perceived family support.

Methods and Materials: A prospective longitudinal predictive design was employed with 1,248 students aged 15–18 years recruited from 12 public high schools across four regions of France. Baseline data on behavioral, emotional, cognitive, and overall academic engagement, academic self-efficacy, and perceived family support were collected through validated French-language questionnaires, while attendance, excused and unexcused absences, lateness, and dropout status were obtained from school administrative records. Students were followed through the end of the academic year and subsequent enrollment verification. Data from nine schools were used for model development, including training and validation subsets, while three schools formed an independent test set. TabNet performance was compared with penalized logistic regression, random forest, and gradient-boosted trees using discrimination, calibration, and classification indices.

Findings: Students who dropped out had significantly lower overall, behavioral, emotional, and cognitive engagement, lower attendance, weaker academic self-efficacy, and lower perceived family support, alongside significantly higher excused absences, unexcused absences, and lateness, with all comparisons reaching $p < .001$. In the independent test set, TabNet achieved an AUROC of .934, an AUPRC of .612, sensitivity of .857, specificity of .919, balanced accuracy of .888, and an F1 score of .640. TabNet significantly outperformed penalized logistic regression and random forest in AUROC, while its advantage over gradient-boosted trees was not statistically significant. Unexcused absence, overall attendance, behavioral engagement, academic self-efficacy, and family support were the most influential predictors.

Conclusion: TabNet accurately predicted high school dropout by integrating behavioral, academic, and relational indicators and may support early, targeted, and interpretable school-based prevention when used alongside professional judgment.

Keywords: High school dropout; TabNet; academic engagement; school attendance; academic self-efficacy; family support.

1. Introduction

High school dropout remains one of the most consequential educational challenges because leaving school before completing an upper-secondary qualification can restrict access to higher education, stable employment, vocational mobility, and long-term social participation. Dropout is rarely the result of a single event. Instead, it is typically the endpoint of a cumulative process involving declining engagement, repeated absence, weak academic confidence, strained relationships, family adversity, socioeconomic disadvantage, and insufficient institutional response. Systematic evidence has shown that dropout behavior emerges from interacting personal, family, school, and community conditions rather than from one isolated risk factor (Banaag et al., 2024). Studies conducted across rural and urban contexts similarly demonstrate that the relative importance of these conditions may vary by setting, but economic hardship, weak school attachment, poor academic performance, inadequate parental support, and attendance difficulties consistently appear among the principal determinants of educational discontinuation (Gbaguidi & Adetou, 2024). Consequently, contemporary dropout research increasingly treats school leaving as a multidimensional developmental and educational process that should be detected before formal withdrawal occurs.

The consequences of dropout extend beyond the absence of a diploma. Educational disconnection can become intertwined with social exclusion, unstable housing, psychological distress, unemployment, and reduced access to protective institutions. Research on youth homelessness, for example, indicates that educational disengagement and housing instability can reinforce one another and that upstream school-based prevention may reduce the likelihood of later disconnection (Kull et al., 2026). Evidence concerning adverse childhood experiences also suggests that cumulative exposure to family conflict, neglect, instability, or trauma may undermine academic functioning, school participation, and educational attainment during adolescence (Zheng, 2024). These findings underscore the need to conceptualize dropout risk not as a fixed characteristic of particular students, but as a dynamic probability that develops through the interaction of educational experiences and broader social conditions. Early identification is therefore critical because preventive intervention is more feasible while a student remains connected to school than after complete withdrawal has occurred.

School attendance is among the most observable indicators in this developmental process. Absence, truancy, lateness, partial attendance, and repeated failure to participate in scheduled instruction can signal emerging disengagement well before a student is officially classified as a dropout. Attendance problems have increasingly been described as early-warning signals that should trigger coordinated school-based and health-informed responses rather than punitive reactions alone (Kearney & Childs, 2023). However, attendance is not a simple binary distinction between presence and absence. Contemporary scholarship emphasizes that school attendance problems vary in frequency, duration, severity, function, and associated impairment, requiring more refined measurement than traditional categorical labels (Kearney & González, 2022). The movement away from historically rigid attendance categories toward dimensional assessment reflects recognition that school refusal, truancy, school withdrawal, and exclusion may overlap and may be influenced by different combinations of emotional, family, behavioral, and institutional factors (Kearney, 2022).

The importance of precise attendance measurement has become even more apparent as researchers and schools attempt to compare findings across educational systems. Inconsistent definitions of absence, chronic absenteeism, school attendance problems, and functional impairment can reduce the comparability of predictive models and intervention studies. Recent efforts toward next-generation common measurement emphasize the need to integrate duration, severity, context, and functional consequences when assessing attendance and absenteeism (Kearney et al., 2026). Functional impairment guidelines have also been proposed to distinguish occasional or developmentally limited absence from patterns that meaningfully interfere with academic, social, and family functioning (Kearney & González, 2022). Such distinctions are important for dropout prediction because a small number of excused absences may carry different implications from persistent unexcused absence, recurrent lateness, or prolonged disengagement from instruction.

Empirical studies across diverse educational settings confirm that attendance problems are closely connected to school leaving. Research in Tanzania has identified truancy management, parental involvement, guidance services, and school-level monitoring as central components of dropout prevention (Kataga & Mwila, 2024). Qualitative findings from Germany further indicate that students experiencing attendance problems value individualized support, trusting

relationships, flexible reintegration, and coordinated assistance from school personnel and families (Enderle et al., 2024). Studies of dropout behavior among youth likewise show that recurrent absence and withdrawal from school routines often form part of an observable behavioral trajectory preceding formal dropout (Faisal et al., 2023). These findings suggest that attendance data are especially valuable for early-warning systems because they are routinely recorded, temporally sensitive, and capable of capturing changes in student behavior over time.

Nevertheless, attendance alone does not fully explain why some students with irregular attendance eventually leave school while others re-engage. Academic engagement provides a broader account of the degree to which students participate in, identify with, and invest in schooling. Engagement is commonly understood as comprising behavioral, emotional, and cognitive dimensions. Behavioral engagement refers to participation, effort, persistence, and adherence to classroom expectations; emotional engagement includes interest, belonging, enjoyment, and affective connection to school; and cognitive engagement involves strategic learning, self-regulation, and willingness to invest effort in understanding complex material. Research has shown that affective and cognitive engagement are directly relevant to decisions about whether to continue or leave school, especially among students exposed to social or educational disadvantage (Zhao-hui et al., 2025). School engagement may therefore function both as a predictor of dropout and as a mechanism through which students interpret their educational experiences.

The developmental importance of engagement is also evident in research on school satisfaction. Age-related declines in satisfaction appear to be partly explained by changes in school engagement, suggesting that students' connection to school may weaken as academic and social demands increase across adolescence (Bălătescu & Cernea-Radu, 2025). Among marginalized populations, engagement can shape educational trajectories and future opportunities by influencing persistence, achievement, and access to subsequent educational pathways (Moreira et al., 2025). Research involving youth living in poverty likewise indicates that student engagement is closely related to aspirations, agency, and future-oriented intentions, although structural disadvantage may constrain the translation of these attitudes into educational success (Khodor et al., 2024). Engagement is therefore not merely an individual motivational attribute; it is also responsive to the

opportunities, relationships, and institutional climates available to students.

School disengagement may have consequences extending beyond educational attainment. Longitudinal evidence has associated disengagement with psychological maladjustment and accelerated biological aging, while supportive parenting may buffer some of these adverse effects (Ong et al., 2022). Students who become progressively detached from school may experience reduced belonging, lower motivation, diminished academic identity, and increasing exposure to behaviors that further weaken educational participation. Perceived burnout, declining motivation, and negative expectations about schooling have also been identified among the reasons students consider leaving education (Yusof et al., 2023). The multidimensional nature of engagement therefore makes it a potentially powerful predictor of dropout, particularly when behavioral, emotional, and cognitive components are examined separately rather than collapsed into a single general score.

Academic self-efficacy represents another central factor in educational persistence. It refers to students' beliefs about their ability to organize and execute the actions required to meet academic demands. Students with strong academic self-efficacy are more likely to persist with difficult tasks, use effective learning strategies, recover from setbacks, and interpret challenges as manageable. In contrast, students with low self-efficacy may avoid demanding work, expect failure, and disengage when they encounter academic difficulty. Evidence indicates that academic self-efficacy mediates the association between insufficient self-control and school dropout, demonstrating that confidence in academic capability can translate broader self-regulatory weaknesses into withdrawal intentions or behavior (Kurtyilmaz & Başak, 2023). Research on self-regulated learning similarly shows that self-efficacy and motivation are associated with intention to drop out and that supportive peer relationships may alter these relationships (Morelli et al., 2022).

Self-efficacy is especially relevant because it can distinguish students who face similar objective challenges but respond differently. Two students may have comparable grades or attendance records, yet the student who believes that academic difficulties can be overcome may be more likely to seek help and persist. Conversely, repeated failure, disciplinary conflict, or negative feedback may erode confidence and accelerate disengagement. Teachers' interpretations of behavioral and disciplinary difficulties

may also affect how students are supported, particularly in special education settings where attributions can influence intervention and inclusion (Mokano & Letuma, 2025). Therefore, incorporating self-efficacy into predictive models may capture a psychologically meaningful dimension of dropout risk that is not fully represented by administrative records.

Family support is another major protective factor. Families may assist students through emotional encouragement, academic monitoring, practical help, communication with school personnel, and reinforcement of long-term educational goals. The absence of such support may leave students less able to manage stress, academic failure, peer conflict, or attendance difficulties. Research on young people who left high school and later re-entered education highlights the importance of attachment, stable relationships, and supportive adults in both school completion and re-enrollment (Ramsdal & Wynn, 2022b). Related work similarly emphasizes that relational security and meaningful adult support can influence whether disconnected youth regain trust in education and return to school (Ramsdal & Wynn, 2022a). Parental involvement is not always straightforward, however, because parents may simultaneously manage employment, caregiving, financial pressure, and their own educational responsibilities, limiting the support they can provide (Damianidou, 2023).

Family support must also be understood within broader structural contexts. Studies in rural Malawi indicate that girls' dropout may reflect constrained autonomy, poverty, gendered expectations, early marriage, household labor, and limited educational opportunity, demonstrating that motivation operates within social and cultural conditions (Banda et al., 2024). Research in Romania has likewise shown that dropout risk in satellite towns is related to spatial inequality, transport difficulties, family resources, and uneven access to educational services (Petre et al., 2024). In remote Australian communities, retention is shaped by the cultural relevance of schooling, mobility, local opportunity, relationships with educators, and the capacity of schools to respond to community circumstances (Guenther et al., 2024). These findings caution against treating family support as an isolated interpersonal variable; its protective function may depend on whether schools and communities provide realistic pathways for continued participation.

Extracurricular participation may also strengthen engagement and belonging, particularly for students whose connection to formal classroom learning is fragile. Evidence from secondary schools in Tanzania suggests that

extracurricular activities can promote confidence, social integration, skill development, and attachment to school, although participation may be restricted by inadequate resources, gender norms, or competing responsibilities (Clement & Mwila, 2023). Short-term motivational interventions have similarly been explored for students already considered at elevated risk. Experiences of demotivated boys participating in a learning camp indicate that intensive support may strengthen motivation and perceived competence, although brief programs cannot by themselves eliminate the structural and relational causes of dropout (Ramsdal & Wynn, 2024). More comprehensive interventions combining educational and social support may therefore be required. A recent mixed-methods intervention in vocational high schools, for example, demonstrated the potential value of integrating social work with culturally responsive educational practices for students at risk of leaving school (Huang & Hai, 2026).

The diversity of dropout pathways makes accurate prediction difficult. Traditional statistical approaches often assume linear and additive relationships, although risk may instead emerge from nonlinear interactions. Low engagement may be especially consequential when combined with high absence, weak self-efficacy, and limited family support, whereas strong support in one domain may compensate for vulnerability in another. Machine-learning methods are increasingly used to identify such patterns. A machine-learning study of out-of-school adolescents in South Korea demonstrated that algorithms can distinguish disconnected youth from enrolled peers using multidimensional social, psychological, and behavioral characteristics (Han et al., 2024). Research in Turkish vocational and technical high schools has also shown that early dropout detection can be strengthened by predictive modeling applied to educational and student-level data (Korkmaz, 2025). These studies support the use of machine learning for screening, while also highlighting the need for transparent interpretation and context-specific validation.

Predictive analytics can help schools move from reactive documentation to proactive prevention. Sophisticated data strategies may identify combinations of attendance and engagement indicators that conventional monitoring overlooks, but their educational value depends on whether outputs can be translated into targeted policy and practical intervention (Kearney et al., 2023). This translation requires models that not only discriminate between students who do and do not drop out but also provide interpretable information regarding the factors contributing to individual

predictions. It also requires careful linkage between K–12 and later educational systems so that prevention addresses continuity, transition, and re-entry rather than focusing only on a single institutional stage (Kearney et al., 2024). Predictive tools should therefore support professional judgment, not replace it, and should direct students toward assistance rather than punishment or exclusion.

TabNet is particularly relevant to this objective because it is designed for structured tabular data and uses sequential attention to select salient predictors at different stages of decision-making. Unlike fully opaque neural networks, TabNet can generate feature-importance information that helps explain whether a prediction is driven primarily by attendance, engagement, self-efficacy, family support, or their interaction. This interpretability is important in educational settings, where predictions affect real students and must be understandable to school leaders, counselors, teachers, and families. The approach is also well suited to datasets combining questionnaire measures with routinely collected administrative indicators. Such integration may improve predictive accuracy by connecting observable behavior with psychological and relational processes.

Although previous studies have identified many correlates of dropout, several gaps remain. Some investigations focus mainly on demographic disadvantage, while others emphasize attendance, motivation, or family conditions separately. Studies of early detection often rely on grades and administrative records but do not include perceived academic competence or family support. Conversely, psychosocial studies may explain dropout intentions without testing predictive performance against verified longitudinal dropout outcomes. Existing models are also frequently developed in a single school or region, limiting evidence of generalizability. Finally, few studies combine multidimensional engagement, detailed attendance indicators, self-efficacy, and family support within an interpretable deep-learning framework and compare its performance with conventional machine-learning and statistical models.

These gaps are especially important in the French context, where students progress through general, technological, and vocational upper-secondary pathways that differ in curriculum, institutional expectations, and post-school opportunities. A model intended for early identification must perform across these pathways and should be evaluated using students from schools not involved in model development. It should also assess both discrimination and calibration, because a model may rank students accurately

while producing unreliable absolute probabilities. Furthermore, subgroup evaluation is needed to determine whether predictive performance differs across gender, grade, educational pathway, or socioeconomic position. Without such evaluation, apparently strong overall performance may conceal weaknesses for particular student groups.

Accordingly, this study aimed to develop and evaluate an interpretable TabNet model for predicting verified high school dropout among students in France using academic engagement, attendance, academic self-efficacy, and perceived family support.

2. Methods and Materials

2.1. Study Design and Participants

This study employed a prospective longitudinal predictive design to develop and evaluate a TabNet-based model for identifying high school students at risk of dropping out. The study population consisted of students enrolled in French upper-secondary schools during the 2024–2025 academic year. A multistage sampling procedure was used to recruit participants from 12 public lycées located in the Île-de-France, Hauts-de-France, Auvergne-Rhône-Alpes, and Nouvelle-Aquitaine regions. These regions were selected to provide variation in geographical location, urbanization, socioeconomic conditions, and school characteristics. Within each participating school, students were recruited from seconde, première, and terminale classes. To be eligible for inclusion, students were required to be between 15 and 18 years of age, enrolled as full-time students at the beginning of the academic year, able to read and understand French, and willing to permit the linkage of their questionnaire responses with school attendance and enrollment records. Students participating in short-term exchange programs, students who had already formally initiated a transfer to another educational institution, and students whose administrative records could not be reliably linked to their questionnaire data were excluded.

A total of 1,326 students were invited to participate in the study. Of these, 1,279 students provided informed consent or assent accompanied by parental consent when required. Thirty-one students were subsequently excluded because more than 30% of their questionnaire responses were missing or their survey data could not be matched with school administrative records. The final analytical sample therefore consisted of exactly 1,248 students. The participants included 639 girls, 594 boys, and 15 students who identified as nonbinary or preferred not to report their

gender. Their ages ranged from 15 to 18 years, with a mean age of 16.42 years and a standard deviation of 0.91. Regarding educational level, 432 participants were enrolled in seconde, 414 were enrolled in première, and 402 were enrolled in terminale. Students were enrolled in general, technological, and vocational educational pathways, thereby reflecting the principal forms of upper-secondary education available within the French educational system.

Baseline questionnaire data were collected between September and October 2024. Attendance data were obtained continuously from school administrative systems throughout the 2024–2025 academic year. Dropout status was determined through administrative follow-up conducted at the end of the academic year and again on October 31, 2025, to establish whether each student had remained in education, completed upper-secondary schooling, transferred to another recognized educational institution, or left formal education without completing the relevant qualification. A student was classified as having dropped out when the student discontinued attendance before completing the expected qualification and was not enrolled in another recognized general, technological, vocational, or apprenticeship program by the follow-up date. Students whose transfer to another educational institution was administratively verified were classified as continuing students rather than dropouts. This operational definition was used to distinguish genuine educational disengagement from ordinary school mobility.

2.2. Measures

Academic engagement was measured using a French-language 18-item School Engagement Scale designed to assess behavioral, emotional, and cognitive engagement in school. The behavioral engagement dimension evaluated students' participation in classroom activities, persistence in completing schoolwork, adherence to school expectations, and effort invested in academic tasks. The emotional engagement dimension assessed students' feelings of interest, belonging, enjoyment, and emotional connection to their school and learning environment. The cognitive engagement dimension measured the use of self-regulated learning strategies, willingness to invest mental effort, perceived value of learning, and attempts to understand complex academic material. Each dimension was represented by six items. Students responded to all items on a five-point scale ranging from 1, indicating strongly disagree, to 5, indicating strongly agree. After reverse-

scored items had been recoded, item scores were summed and divided by the number of completed items to produce an overall academic engagement score ranging from 1 to 5. Higher scores represented stronger engagement with school and academic activities. Scores for behavioral, emotional, and cognitive engagement were also retained as separate model features to allow the algorithm to identify potentially different contributions of the three engagement dimensions. The French wording of the items was reviewed by educational psychologists and secondary-school teachers, and cognitive interviews with a small group of students were conducted before the main data collection to confirm clarity and age appropriateness. The overall scale demonstrated satisfactory internal consistency in the study sample, with a Cronbach's alpha coefficient of .89.

Attendance was assessed using objective administrative information extracted from the electronic attendance systems of the participating schools. The records included the number of scheduled instructional sessions, the number of sessions attended, the number of excused absences, the number of unexcused absences, and the number of recorded late arrivals. Because the total number of scheduled sessions differed according to grade level, educational pathway, and school timetable, raw absence counts were converted into proportional indicators. The primary attendance measure was calculated by dividing the number of attended instructional sessions by the total number of scheduled instructional sessions and multiplying the resulting value by 100. Higher percentages therefore indicated more consistent attendance. Separate rates were also calculated for excused absences and unexcused absences because these forms of nonattendance may reflect different educational and personal processes. The number of late arrivals was converted into a rate per 100 scheduled school days. Attendance information was aggregated over the first two academic terms so that only information available before the final dropout determination was used as a predictor. Administrative entries that had been corrected or withdrawn by authorized school personnel were excluded from the final dataset.

Academic self-efficacy was measured using a French-language 10-item Academic Self-Efficacy Scale. The instrument assessed students' confidence in their ability to understand instructional material, organize their academic work, complete difficult assignments, prepare effectively for examinations, overcome learning difficulties, maintain concentration, and achieve educational goals. Responses were recorded on a four-point scale ranging from 1,

indicating not at all true of me, to 4, indicating completely true of me. Item scores were summed to produce a total score ranging from 10 to 40, with higher scores indicating greater confidence in the ability to manage academic demands successfully. The score was treated as a continuous predictor in the primary model. Item-level missingness was handled by calculating the participant's mean response across completed items when at least eight of the ten items had been answered. A total score was considered missing when fewer than eight items were completed. The scale demonstrated good internal consistency in the present sample, with a Cronbach's alpha coefficient of .87.

Perceived family support was assessed using the four-item family subscale of the French version of the Multidimensional Scale of Perceived Social Support. The items evaluated the extent to which students believed that their family was available to provide emotional assistance, practical help, encouragement, and opportunities to discuss personal or academic problems. Students responded on a seven-point scale ranging from 1, indicating very strongly disagree, to 7, indicating very strongly agree. The mean of the four item scores was calculated, resulting in a family-support score ranging from 1 to 7. Higher scores represented stronger perceived support from family members. A scale score was calculated when at least three items had been completed. The family-support subscale showed strong internal consistency in the study sample, with a Cronbach's alpha coefficient of .91.

A demographic questionnaire was used to collect information regarding age, gender, grade level, educational pathway, primary language spoken at home, parental educational level, household composition, and eligibility for means-tested school assistance. These variables were used to describe the sample, examine the representativeness of the development and test datasets, and conduct supplementary subgroup analyses. They were not included as primary predictors in the principal TabNet model because the study was specifically designed to determine whether dropout risk could be predicted from academic engagement, attendance, self-efficacy, and family support. This restriction also reduced the possibility that the model would rely predominantly on sensitive demographic or socioeconomic characteristics. The binary outcome variable was dropout status, coded as 1 for confirmed dropout and 0 for continued enrollment, verified transfer, or successful completion of upper-secondary education.

Questionnaires were administered during regular school hours in classrooms or supervised computer rooms. Students

completed the measures electronically using a secure survey platform, and the average completion time was approximately 20 minutes. Research assistants provided standardized instructions and were available to clarify procedural questions without influencing participants' responses. Questionnaire data were linked to attendance and enrollment records using pseudonymous identification codes generated by designated school administrators. The researchers did not receive students' names, home addresses, or other direct identifying information.

2.3. Data Analysis

Data analysis was conducted using Python with the PyTorch deep-learning framework and the PyTorch TabNet implementation. Before model development, the data were examined for impossible values, duplicate records, inconsistent identifiers, missingness patterns, and extreme observations. Continuous variables were summarized using means, standard deviations, medians, and ranges, while categorical variables were summarized using frequencies and percentages. Differences between students with complete and incomplete data were examined to identify possible systematic patterns of missingness. Missing continuous values were imputed using the median calculated exclusively from the training dataset, while missing categorical values were assigned a separate unknown category. Binary indicators were created to identify observations for which a predictor had been imputed. All preprocessing parameters were estimated from the training data and subsequently applied without modification to the validation and test datasets to prevent information leakage.

To provide an independent evaluation of model generalizability across school settings, the participants were divided according to school membership rather than through unrestricted individual-level randomization. Data from nine schools, comprising 936 students, were assigned to the model-development dataset, while data from three separate schools, comprising 312 students, were retained as an independent test dataset. The development dataset was then divided into a training set of 749 students and a validation set of 187 students using stratified random sampling based on dropout status, grade level, and educational pathway. The training dataset was used to estimate the model parameters, the validation dataset was used for hyperparameter selection and early stopping, and the independent test dataset was used only once for the final evaluation. No student from a test school was included in the training or validation data.

TabNet was selected because it is specifically designed for structured tabular data and employs sequential attention mechanisms to identify the features that are most informative at different stages of prediction. The model received the overall and dimension-specific academic engagement scores, attendance percentage, excused-absence rate, unexcused-absence rate, lateness rate, academic self-efficacy score, and family-support score as input features. The final architecture used decision and attention dimensions of 32, five sequential decision steps, a relaxation parameter of 1.5, two shared gated linear-unit layers, two independent gated linear-unit layers, and a sparsity regularization coefficient of 0.0001. These parameters were selected through Bayesian hyperparameter optimization conducted on the training and validation datasets. The search procedure considered decision and attention dimensions between 8 and 64, three to eight decision steps, relaxation parameters between 1.0 and 2.0, learning rates between 0.001 and 0.05, and sparsity coefficients between 0 and 0.001.

The model was trained using the Adam optimizer with an initial learning rate of 0.02, a batch size of 256, and a virtual batch size of 64. Binary cross-entropy was used as the loss function. Because dropout was expected to occur less frequently than continued enrollment, class weights inversely proportional to the class frequencies in the training dataset were incorporated into the loss function. This procedure increased the penalty associated with incorrectly classifying students who subsequently dropped out and reduced the likelihood that the model would achieve apparently high accuracy by predicting the majority class. Training was permitted to continue for a maximum of 200 epochs, with early stopping applied when the validation area under the receiver operating characteristic curve did not improve for 25 consecutive epochs. The model parameters associated with the highest validation performance were retained. A fixed random seed of 42 was used for data partitioning and model initialization to improve reproducibility.

The primary measure of predictive discrimination was the area under the receiver operating characteristic curve. Because dropout represented the less frequent and educationally important outcome, the area under the precision–recall curve was also treated as a central performance measure. Additional indices included accuracy, balanced accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and the F1 score. The probability threshold used to classify students as being

at risk was selected using the validation dataset by maximizing Youden’s index while ensuring that sensitivity remained sufficiently high for an early-warning application. The selected threshold was then applied unchanged to the independent test dataset. Ninety-five percent confidence intervals for performance measures were calculated using 2,000 stratified bootstrap samples.

Model calibration was evaluated to determine whether predicted probabilities corresponded with the observed frequency of dropout. Calibration was examined using the Brier score, calibration intercept, calibration slope, and graphical comparison of predicted and observed probabilities across probability groups. Decision-curve analysis was conducted to examine the potential net benefit of using the model across a range of risk thresholds relevant to school-based screening. This analysis was used to determine whether applying the model would provide greater practical value than referring all students or no students for additional assessment.

To establish whether TabNet provided an advantage over conventional statistical and machine-learning approaches, its performance was compared with penalized logistic regression, random forest, and gradient-boosted decision-tree models. All comparison models were developed using the same training, validation, and test partitions and the same predictor set. Hyperparameters for each comparison model were optimized using the validation data. Differences in the area under the receiver operating characteristic curve were evaluated using DeLong’s method, while differences in paired classification errors were examined using McNemar’s test. Statistical tests were two-sided, and a significance level of .05 was applied.

The interpretability of the TabNet model was examined using its sequential attention masks and aggregated feature-importance scores. Global importance estimates were used to identify which dimensions of engagement, attendance, self-efficacy, and family support contributed most strongly to predictions across the complete test sample. Local attention masks were inspected to determine how the model combined predictors when estimating risk for individual students. These explanations were treated as indicators of the model’s predictive reasoning rather than evidence of causal relationships. Sensitivity analyses were also conducted by repeating the modeling procedure without imputed cases, by using alternative probability thresholds, and by evaluating model performance separately across gender, grade level, educational pathway, and school region. The independent test dataset remained untouched during model selection, and

all final performance statistics were calculated only after the complete analytical procedure had been specified.

3. Findings and Results

The final sample consisted of 1,248 students enrolled in 12 public upper-secondary schools in France. The participants' ages ranged from 15 to 18 years, with a mean age of 16.42 years and a standard deviation of 0.91. The sample included 639 girls, representing 51.2% of the participants, 594 boys, representing 47.6%, and 15 students who identified as nonbinary or preferred not to report their gender, representing 1.2% of the sample. Regarding grade level, 432 students, or 34.6%, were enrolled in *seconde*, 414 students, or 33.2%, were enrolled in *première*, and 402 students, or 32.2%, were enrolled in *terminale*. A total of 694 students, representing 55.6% of the sample, were enrolled in the general educational pathway, 287 students, representing 23.0%, were enrolled in the technological pathway, and 267 students, representing 21.4%, were enrolled in the vocational pathway. French was the primary language spoken at home by 1,026 students, or 82.2%, whereas 222 students, or 17.8%, primarily spoke another language at home. In relation to household composition, 849 students, representing 68.0% of the sample, lived in two-parent households, 310 students, representing 24.8%, lived in single-parent households, and 89 students, representing 7.1%, lived with other relatives, foster carers, or legal guardians. At least one parent had completed tertiary education for 512 students, or 41.0%, while 497 students, or 39.8%, reported that the highest parental educational level was upper-secondary education, and 239 students, or 19.2%, reported a parental educational level below upper-secondary education. In addition, 456 students, representing 36.5% of the sample, were eligible for means-tested school assistance.

At the end of the follow-up period, 112 students had left formal education without completing their expected upper-secondary qualification and without verified enrollment in another recognized educational or apprenticeship program. The observed dropout rate was therefore 9.0%. The remaining 1,136 students, representing 91.0% of the sample, had remained enrolled, completed upper-secondary education, or transferred to another verified educational program. Of the continuing students, 1,049 remained enrolled in or completed their original educational pathway, while 87 had transferred to another recognized educational

institution or apprenticeship program. The dropout rate was 7.7% among girls, 10.3% among boys, and 13.3% among students who identified as nonbinary or did not report their gender. However, the association between gender and dropout status was not statistically significant, $\chi^2(2) = 2.90$, $p = .234$. Dropout occurred among 10.9% of students in *seconde*, 9.4% of students in *première*, and 6.5% of students in *terminale*. Although this pattern suggested that dropout was more common during the earlier stage of upper-secondary education, the overall association between grade level and dropout status did not reach statistical significance, $\chi^2(2) = 5.11$, $p = .078$.

Dropout status differed significantly across educational pathways, $\chi^2(2) = 23.94$, $p < .001$. The dropout rate was 6.1% in the general pathway, 9.4% in the technological pathway, and 16.1% in the vocational pathway. Students receiving means-tested school assistance also had a higher dropout rate than students who did not receive such assistance. Specifically, 66 of the 456 students receiving assistance, or 14.5%, dropped out, compared with 46 of the 792 students who did not receive assistance, or 5.8%. This difference was statistically significant, $\chi^2(1) = 25.55$, $p < .001$. These demographic variables were examined to characterize the sample and assess potential subgroup differences, but they were not entered as primary predictors in the principal TabNet model.

The model-development dataset contained 936 students from nine schools, including 84 students who dropped out and 852 students who continued their education. Within this dataset, 749 students, including 67 students who dropped out, were assigned to the training sample, while 187 students, including 17 students who dropped out, were assigned to the validation sample. The independent test dataset consisted of 312 students from three schools that were not represented in the model-development data. This test dataset included 28 students who dropped out and 284 continuing students. The dropout prevalence was therefore approximately 9.0% in each dataset, indicating that the school-level partitioning procedure preserved the overall outcome distribution. No outcome data were missing. Predictor-level missingness ranged from 0.4% for attendance information to 4.3% for perceived family support. The median percentage of missing predictor values per participant was 1.1%, and no participant in the final analytical sample had more than 30% missing questionnaire information.

Table 1

Comparison of Academic Engagement, Attendance, Academic Self-Efficacy, and Family Support According to Dropout Status

Predictor	Continuing students, n = 1,136, M (SD)	Students who dropped out, n = 112, M (SD)	Welch's t	Cohen's d	p
Overall academic engagement	3.72 (0.54)	2.84 (0.61)	14.71	1.61	< .001
Behavioral engagement	3.78 (0.58)	2.70 (0.64)	17.18	1.84	< .001
Emotional engagement	3.69 (0.63)	2.98 (0.70)	10.33	1.12	< .001
Cognitive engagement	3.68 (0.57)	2.85 (0.65)	13.03	1.44	< .001
Attendance rate (%)	94.10 (4.30)	79.60 (8.70)	17.43	2.99	< .001
Excused absence rate (%)	3.70 (3.00)	7.80 (5.10)	-8.37	-1.26	< .001
Unexcused absence rate (%)	2.20 (2.50)	12.60 (7.40)	-14.79	-3.20	< .001
Late arrivals per 100 school days	3.80 (4.10)	12.90 (8.30)	-11.47	-1.96	< .001
Academic self-efficacy	30.80 (4.60)	22.90 (5.30)	15.22	1.69	< .001
Perceived family support	5.61 (1.02)	4.12 (1.31)	11.69	1.42	< .001

As shown in Table 1, all primary predictor variables differed significantly between students who continued their education and students who dropped out. Continuing students reported substantially greater overall academic engagement than students who subsequently dropped out, with mean scores of 3.72 and 2.84, respectively, $t = 14.71$, $p < .001$. The corresponding effect size was large, $d = 1.61$. Among the engagement dimensions, the largest difference was observed for behavioral engagement. Continuing students had a mean behavioral engagement score of 3.78, whereas students who dropped out had a mean score of 2.70, $t = 17.18$, $p < .001$, $d = 1.84$. This finding indicated that participation in classroom activities, persistence with schoolwork, and compliance with academic expectations were particularly strongly differentiated between the two outcome groups. Cognitive engagement was also substantially higher among continuing students, $t = 13.03$, $p < .001$, $d = 1.44$, while the difference in emotional engagement was slightly smaller but remained large and statistically significant, $t = 10.33$, $p < .001$, $d = 1.12$.

Attendance-related indicators produced the largest group differences. Continuing students attended an average of 94.10% of scheduled instructional sessions, compared with

79.60% among students who dropped out, $t = 17.43$, $p < .001$. The effect size for this difference was exceptionally large, $d = 2.99$. Students who dropped out also had substantially higher unexcused absence rates, with a mean of 12.60%, compared with 2.20% among continuing students, $t = -14.79$, $p < .001$, $d = -3.20$. The magnitude of this difference exceeded that observed for any psychosocial predictor. Excused absences were also more frequent among students who dropped out, although the difference was smaller than the difference in unexcused absences, $t = -8.37$, $p < .001$, $d = -1.26$. In addition, students who dropped out recorded an average of 12.90 late arrivals per 100 school days, compared with 3.80 among continuing students, $t = -11.47$, $p < .001$, $d = -1.96$. Academic self-efficacy and family support also clearly distinguished the groups. Continuing students reported higher academic self-efficacy, $t = 15.22$, $p < .001$, $d = 1.69$, and greater perceived family support, $t = 11.69$, $p < .001$, $d = 1.42$. Collectively, these findings demonstrated that students who subsequently dropped out were characterized by lower engagement, weaker confidence in their academic abilities, less perceived family support, lower overall attendance, and more frequent unexcused absences and late arrivals.

Table 2

Predictive Performance of TabNet and Comparison Models in the Independent Test Dataset

Model	Classification threshold	AUROC (95% CI)	AUPRC (95% CI)	Accuracy	Balanced accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1 score	Brier score
TabNet	0.31	.934 (.891–.968)	.612 (.459–.744)	.913	.888	.857	.919	.511	.985	.640	.061
Gradient-boosted trees	0.29	.918 (.868–.957)	.561 (.405–.701)	.904	.867	.821	.912	.479	.981	.605	.066

Random forest	0.27	.901 (.842–.946)	.503 (.356–.649)	.897	.847	.786	.908	.458	.977	.579	.070
Penalized logistic regression	0.24	.886 (.824–.938)	.462 (.318–.612)	.885	.824	.750	.898	.420	.973	.538	.075

Table 2 presents the predictive performance of TabNet and the three comparison models in the independent school-level test dataset. TabNet demonstrated the strongest overall discrimination, with an AUROC of .934 and a 95% confidence interval ranging from .891 to .968. This value indicated that, when presented with one student who dropped out and one student who continued, the model assigned a higher estimated risk to the student who dropped out in approximately 93.4% of comparisons. The model also obtained an AUPRC of .612, which was substantially higher than the no-skill precision–recall value of .090 determined by the prevalence of dropout in the test dataset. The AUPRC finding was particularly important because it demonstrated that TabNet retained meaningful predictive precision despite the relatively low frequency of dropout.

At the probability threshold of .31 selected in the validation dataset, TabNet correctly identified 24 of the 28 students who dropped out, producing a sensitivity of .857. It correctly classified 261 of the 284 continuing students, corresponding to a specificity of .919. The model therefore produced four false-negative classifications and 23 false-positive classifications. Overall, 285 of the 312 students were classified correctly, resulting in an accuracy of .913 and a balanced accuracy of .888. The positive predictive value was .511, indicating that approximately 51.1% of students classified as high risk subsequently met the administrative definition of dropout. The negative predictive value was .985, indicating that 98.5% of students classified as lower risk continued their education, completed their program, or transferred to another recognized educational setting. The F1 score of .640 reflected the model's comparatively favorable balance between sensitivity and positive predictive value.

The gradient-boosted tree model achieved the second-highest discrimination, with an AUROC of .918 and an AUPRC of .561. Its sensitivity of .821 indicated that it correctly identified 23 of the 28 dropout cases, while its specificity of .912 showed that it correctly classified 259 of the 284 continuing students. The random forest model produced an AUROC of .901 and an AUPRC of .503, while penalized logistic regression demonstrated the lowest discrimination, with an AUROC of .886 and an AUPRC of .462. The difference between the AUROCs of TabNet and penalized logistic regression was statistically significant according to DeLong's test, $p = .021$. The difference between TabNet and random forest was also statistically significant, $p = .048$. Although TabNet had a higher AUROC than the gradient-boosted tree model, the difference between these two models did not reach statistical significance, $p = .173$.

TabNet also produced the lowest Brier score, at .061, indicating that its predicted probabilities were more accurate than those generated by the comparison models. The Brier scores were .066 for gradient-boosted trees, .070 for random forest, and .075 for penalized logistic regression. Although positive predictive values were lower than sensitivity and specificity for all models, this pattern was expected because only 9.0% of students in the test dataset dropped out. The high negative predictive value of TabNet suggested that the model was especially effective in identifying students who were unlikely to drop out. At the same time, the positive predictive value indicated that model-generated high-risk classifications should be treated as signals for additional professional assessment rather than as definitive determinations that a student would leave school.

Table 3

Global Feature Importance and Permutation-Based Contribution of Predictors in the Final TabNet Model

Rank	Predictor	Normalized TabNet importance	Decrease in test AUROC after permutation	Pattern associated with greater predicted dropout risk
1	Unexcused absence rate	.216	.081	Higher unexcused absence
2	Overall attendance rate	.188	.070	Lower attendance
3	Behavioral engagement	.142	.047	Lower behavioral engagement
4	Academic self-efficacy	.115	.036	Lower academic self-efficacy
5	Perceived family support	.094	.029	Lower family support

6	Cognitive engagement	.079	.024	Lower cognitive engagement
7	Late arrivals	.072	.020	More frequent lateness
8	Overall academic engagement	.049	.014	Lower overall engagement
9	Emotional engagement	.027	.008	Lower emotional engagement
10	Excused absence rate	.018	.005	Higher excused absence

As reported in Table 3, attendance-related indicators were the most influential features in the final TabNet model. The unexcused absence rate had the highest normalized importance value, at .216, and permuting this feature reduced the test-set AUROC by .081. This was the largest decline associated with any individual predictor, indicating that unexcused absence information made the strongest unique contribution to distinguishing students who dropped out from those who continued their education. Overall attendance rate was the second-most influential predictor, with a normalized importance of .188 and an AUROC reduction of .070 following permutation. Examination of the model's local attention patterns indicated that predicted risk increased sharply when overall attendance fell below approximately 88% and increased further when the unexcused absence rate exceeded approximately 8%. These values were not treated as universal causal thresholds but represented regions in the observed data where the model assigned progressively greater dropout probabilities.

Behavioral engagement was the most influential questionnaire-based predictor, with a normalized importance of .142. Permuting behavioral engagement reduced the AUROC by .047, demonstrating that the variable contributed predictive information beyond that already provided by attendance records. Lower behavioral engagement was particularly influential when it occurred together with frequent unexcused absence or lateness. Academic self-efficacy ranked fourth, with a normalized importance value of .115 and an AUROC reduction of .036 after permutation. Students who expressed limited confidence in their capacity to organize schoolwork, complete difficult assignments, prepare for examinations,

and overcome academic challenges generally received higher estimated dropout probabilities.

Perceived family support accounted for .094 of the total normalized importance and produced a .029 reduction in AUROC after permutation. Its influence was most apparent among students with moderately low engagement or attendance, suggesting that family support helped the model distinguish between students whose emerging academic difficulties were more or less likely to develop into dropout. Cognitive engagement and lateness also made meaningful contributions, with normalized importance values of .079 and .072, respectively. Lower cognitive engagement was associated with elevated risk, particularly when students reported limited investment in understanding difficult learning material or limited use of self-regulated learning strategies. Frequent late arrival contributed less than unexcused absence but remained an informative behavioral sign of disengagement.

Overall academic engagement, emotional engagement, and excused absence rates had comparatively smaller unique importance values. The lower importance of the overall engagement score did not indicate that engagement was unimportant. Instead, much of its predictive information was represented more specifically by the behavioral and cognitive engagement subscales. Similarly, excused absence was associated with dropout in the bivariate analyses but had limited unique importance once unexcused absence, total attendance, and lateness were considered simultaneously. These results demonstrated that TabNet did not rely on one isolated variable. Rather, the model integrated attendance behavior, school engagement, perceived academic competence, and family support when assigning individual risk probabilities.

Figure 1

Receiver Operating Characteristic and Precision–Recall Curves for TabNet and the Comparison Models in the Independent Test Dataset

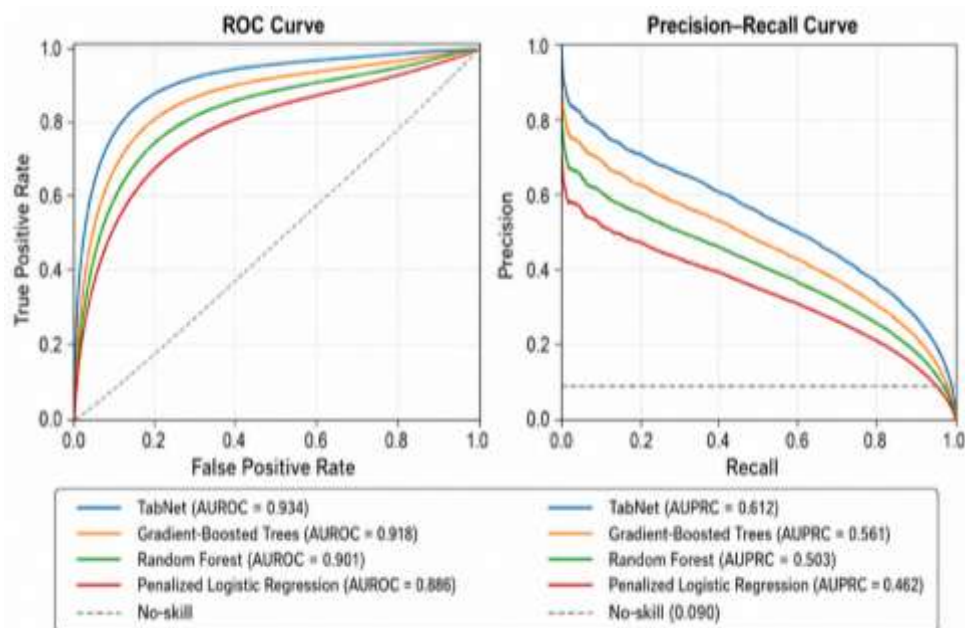


Figure 1 illustrates the receiver operating characteristic and precision–recall curves for TabNet, gradient-boosted trees, random forest, and penalized logistic regression. Across most classification thresholds, the TabNet receiver operating characteristic curve remained above the curves of the comparison models, reflecting its higher sensitivity at comparable false-positive rates. The advantage of TabNet was particularly apparent in the region corresponding to false-positive rates between .05 and .20, which represented the range most relevant to a school-based early-warning system seeking to identify a high proportion of students who might drop out without referring an impractically large proportion of the student body. At the selected probability threshold of .31, the model achieved a true-positive rate of .857 and a false-positive rate of .081.

The precision–recall curve provided a more demanding evaluation because it focused directly on the minority dropout class. The TabNet curve remained above the no-skill reference level of .090 across the complete recall range and generally remained above the curves of the three comparison models. At the selected threshold, recall was .857 and precision was .511. Thus, the model identified more than four-fifths of the students who dropped out, while slightly more than half of the students classified as high risk were confirmed as dropout cases during follow-up. Precision decreased as the threshold was lowered to achieve sensitivities above .90, demonstrating the practical trade-off between identifying a greater proportion of students at risk

and increasing the number of continuing students referred for additional assessment.

Calibration analysis further indicated that the predicted probabilities generated by TabNet corresponded closely with the observed outcome rates. The calibration intercept was -0.03 , and the calibration slope was 0.97 . An intercept close to zero indicated minimal systematic overprediction or underprediction, while a slope close to one indicated limited overfitting when the model was applied to students from schools that had not contributed to model development. Students in the lowest predicted-risk group had an observed dropout rate of less than 1%, while students in the highest predicted-risk group had an observed dropout rate of approximately 48%. The expected calibration error was .028, and the maximum calibration error across probability groups was .071. These findings, together with the Brier score of .061, indicated that the model provided not only strong rank-order discrimination but also reasonably accurate estimates of absolute dropout probability.

Decision-curve analysis showed that using TabNet provided a greater estimated net benefit than either referring all students or referring no students across probability thresholds ranging from approximately .10 to .55. The practical advantage was most evident between thresholds of .20 and .40, within which the model retained high sensitivity while substantially reducing unnecessary referrals. At the selected threshold of .31, 47 of the 312 students in the independent test dataset were classified as high risk. These

classifications included 24 true-positive cases and 23 false-positive cases. Thus, 15.1% of the test sample would have been referred for further assessment, while 85.7% of the students who subsequently dropped out would have been identified. This result suggested that the model could support a targeted, resource-conscious screening process when combined with professional review and appropriate student-support procedures.

Sensitivity analyses supported the robustness of the main findings. When the model was re-estimated using complete cases only, the test-set AUROC was .928, compared with .934 in the primary analysis, indicating that median imputation and missingness indicators did not materially inflate model performance. When the high-risk classification threshold was lowered from .31 to .25, sensitivity increased from .857 to .893, while specificity decreased from .919 to .877. Conversely, increasing the threshold to .40 reduced sensitivity to .750 but increased specificity to .951. These results confirmed that the model's operating characteristics could be adjusted according to the relative importance assigned to missed dropout cases and unnecessary referrals.

Subgroup analyses did not reveal substantial deterioration in model discrimination. The AUROC was .928 among girls and .936 among boys. The small number of nonbinary students prevented a reliable separate estimate for that subgroup. Across grade levels, the AUROC was .921 for students in seconde, .939 for students in première, and .944 for students in terminale. The AUROC was .929 in the general pathway, .920 in the technological pathway, and .915 in the vocational pathway. Although performance was slightly lower in the vocational pathway, the confidence intervals overlapped substantially, and the observed difference was not statistically significant. Model sensitivity was .833 among students receiving means-tested assistance and .870 among students who did not receive such assistance. The difference in false-positive rates between these groups was 3.4 percentage points and was not statistically significant after correction for multiple subgroup comparisons. The findings therefore provided no evidence of major performance disparities across the examined demographic and educational groups, although the relatively small numbers of dropout cases within some subgroups required cautious interpretation.

Overall, the findings demonstrated that high school dropout could be predicted with strong discrimination by integrating academic engagement, objective attendance indicators, academic self-efficacy, and perceived family support within a TabNet architecture. Attendance and

unexcused absence were the strongest individual predictors, but behavioral engagement, self-efficacy, and family support contributed additional information that improved predictions beyond attendance records alone. TabNet outperformed penalized logistic regression and random forest and produced modestly stronger performance than gradient-boosted trees. The combination of high sensitivity, strong negative predictive value, acceptable calibration, and interpretable feature-attention patterns supported the potential use of the model as an early-warning aid. However, its predictions were probabilistic rather than deterministic and were intended to supplement, rather than replace, professional educational judgment.

4. Discussion

The present study developed and evaluated a TabNet model for predicting verified high school dropout among French upper-secondary students using academic engagement, attendance, academic self-efficacy, and perceived family support. The findings demonstrated that students who dropped out differed substantially from continuing students across all four predictor domains. They reported lower behavioral, emotional, cognitive, and overall academic engagement, attended a smaller proportion of scheduled instructional sessions, accumulated more excused and unexcused absences, arrived late more frequently, expressed weaker confidence in their academic capabilities, and perceived less support from their families. The predictive analysis further showed that TabNet achieved strong discrimination in an independent dataset drawn from schools not used for model development. Its area under the receiver operating characteristic curve was .934, and its area under the precision-recall curve was .612, exceeding the performance of penalized logistic regression, random forest, and gradient-boosted trees. Unexcused absence, overall attendance, behavioral engagement, academic self-efficacy, and family support emerged as the most influential predictors. Taken together, the results support the view that high school dropout is best understood as the outcome of interacting behavioral, psychological, and relational processes rather than as the consequence of a single isolated risk factor.

The markedly lower academic engagement observed among students who dropped out is consistent with the conceptualization of dropout as a gradual process of detachment from school. The largest engagement-related difference concerned behavioral engagement, suggesting

that reduced classroom participation, weak persistence, incomplete academic work, and declining adherence to school routines may be especially visible signs of emerging dropout risk. This pattern is consistent with evidence showing that both affective and cognitive engagement influence students' decisions to continue or leave education (Zhao-hui et al., 2025). It also supports findings that school engagement shapes educational trajectories and future opportunities among students exposed to social and cultural marginalization (Moreira et al., 2025). Engagement may function as a proximal indicator through which broader difficulties become expressed in daily school behavior. Students may initially experience low confidence, family stress, or dissatisfaction, but these experiences become educationally consequential when they reduce effort, participation, persistence, and identification with school.

The lower emotional and cognitive engagement scores among dropout students further indicate that educational withdrawal involved more than visible noncompliance. Students who left school reported weaker feelings of belonging and interest as well as less investment in understanding difficult academic material and regulating their own learning. Previous evidence suggests that engagement mediates age-related variations in school satisfaction, indicating that students may become less satisfied with school as their connection to its activities, relationships, and purposes weakens (Bălătescu & Cernea-Radu, 2025). Similarly, research on student motivation, burnout, and dropout reasons has identified declining motivation and negative emotional experiences as important components of withdrawal intentions (Yusof et al., 2023). The current results extend these findings by showing that separate engagement dimensions contributed to an independently evaluated predictive model. Behavioral engagement was the most influential engagement indicator, but cognitive and emotional engagement also added information, suggesting that early-warning systems should not rely exclusively on observable conduct.

The especially strong contribution of attendance indicators was one of the clearest findings of the study. Students who dropped out attended, on average, only 79.60% of scheduled sessions, compared with 94.10% among students who continued their education. Unexcused absence produced the largest standardized group difference and ranked first in the TabNet feature-importance analysis. These findings align closely with research describing absenteeism as an early-warning signal that frequently precedes more complete educational disconnection

(Kearney & Childs, 2023). Attendance problems can interrupt learning, weaken peer and teacher relationships, increase difficulty in catching up with academic work, and reinforce avoidance when returning to school becomes increasingly stressful. Repeated absence may therefore operate both as a marker of existing disengagement and as a mechanism that intensifies future dropout risk.

The distinction between unexcused and excused absence was also important. Although both were higher among students who dropped out, unexcused absence contributed much more strongly to prediction. This result supports arguments that school attendance problems should be measured dimensionally rather than treated as a single homogeneous category (Kearney, 2022). Absences differ in their causes, duration, severity, and functional consequences, and these distinctions may have different implications for intervention. The stronger influence of unexcused absence may reflect deliberate avoidance, weakened parental monitoring, conflict with school, or a reduced sense of obligation toward attendance. By contrast, excused absence may include illness and other circumstances that are less directly indicative of disengagement. The findings therefore reinforce calls for common, next-generation attendance measurement that captures the form and functional significance of nonattendance rather than relying only on total absence counts (Kearney et al., 2026).

Frequent lateness also contributed to dropout prediction, although less strongly than unexcused absence and overall attendance. Lateness may represent an earlier or less severe stage of attendance deterioration. Students who are repeatedly late remain partially connected to school, but their difficulty arriving on time may indicate sleep problems, transportation barriers, family responsibilities, low motivation, avoidance of particular classes, or conflict with institutional expectations. Behavioral studies of school dropout have similarly found that repeated deviations from normal school routines can form part of a progressive pattern that culminates in withdrawal (Faisal et al., 2023). Research on strategies for managing truancy and dropout also emphasizes the importance of systematic attendance monitoring and early school responses (Kataga & Mwila, 2024). Thus, lateness should not automatically be treated as a minor disciplinary matter, particularly when it co-occurs with low engagement or declining academic confidence.

The findings related to academic self-efficacy were also substantial. Students who continued their education reported considerably stronger confidence in their ability to organize

schoolwork, prepare for examinations, understand difficult content, and overcome academic challenges. Academic self-efficacy ranked fourth in the TabNet model, indicating that it provided predictive value beyond attendance and engagement. This finding is consistent with evidence that academic self-efficacy mediates the relationship between insufficient self-control and dropout (Kurtyilmaz & Başak, 2023). Students with weak self-efficacy may interpret academic setbacks as evidence that they are incapable of succeeding, which can reduce effort, help-seeking, and persistence. As disengagement and absence increase, missed instruction may further undermine competence, producing a reciprocal cycle in which low confidence contributes to avoidance and avoidance generates additional academic difficulty.

Related evidence from research on self-regulated learning indicates that self-efficacy and motivation are associated with dropout intentions, while supportive friendships may reduce the strength of this relationship (Morelli et al., 2022). The present findings suggest that similar protective and risk processes may operate during high school. Self-efficacy is important because it captures students' interpretation of their academic capacity rather than only their objective performance. Two students with similar grades may differ considerably in their beliefs about whether improvement is possible. A student who perceives difficulties as temporary and manageable may continue investing effort, whereas a student who expects failure may withdraw even before performance declines to a critical level. The contribution of self-efficacy to the TabNet model therefore demonstrates the value of combining psychological assessments with administrative school data.

Perceived family support also distinguished students who dropped out from those who continued and made a meaningful contribution to the predictive model. Students who perceived their families as emotionally available, encouraging, and willing to help with difficulties were less likely to leave school. This finding aligns with research emphasizing the importance of attachment and supportive adult relationships in both school completion and re-enrollment after dropout (Ramsdal & Wynn, 2022b). Young people who have already left school often describe reliable relationships and a sense of being valued as central to restoring educational participation (Ramsdal & Wynn, 2022a). Family support may protect against dropout by helping students manage stress, maintain routines, communicate with school personnel, and sustain long-term

goals when immediate educational experiences become discouraging.

The contribution of family support should not, however, be interpreted as evidence that families alone are responsible for preventing dropout. Parents and caregivers may face economic hardship, employment demands, health problems, housing instability, or their own educational responsibilities, all of which can limit their capacity to provide sustained academic assistance (Damianidou, 2023). The association between educational disconnection and youth homelessness further illustrates that school and family instability may reinforce one another (Kull et al., 2026). Structural conditions such as poverty, gendered expectations, transportation difficulties, and unequal access to educational resources may also constrain family support and students' capacity to remain enrolled (Banda et al., 2024; Petre et al., 2024). Consequently, lower perceived support should be treated as an indicator of a need for additional institutional assistance, not as a basis for blaming parents or students.

The higher dropout rate observed in the vocational pathway is consistent with evidence from studies of vocational and technical schools, where students may face accumulated academic difficulties, weaker school identification, restricted perceived opportunities, and greater socioeconomic vulnerability. Research in Turkey has shown that early dropout detection in vocational and technical high schools can be improved through the analysis of multiple educational indicators (Korkmaz, 2025). A mixed-methods intervention among vocational students in China also demonstrated that dropout prevention may benefit from combining educational assistance with social-work and culturally responsive support (Huang & Hai, 2026). The present results indicate that predictive systems in France should be evaluated carefully across educational pathways. Although the TabNet model performed well in the vocational subgroup, its discrimination was slightly lower than in the general and technological pathways, suggesting that additional pathway-specific variables may be needed.

The strong performance of TabNet supports the use of advanced machine-learning methods for dropout prediction. Its AUROC of .934 indicated excellent discrimination, while its AUPRC of .612 showed meaningful performance for the relatively uncommon dropout outcome. This is important because accuracy alone can be misleading when most students remain enrolled. A model that classifies nearly everyone as a continuing student may appear accurate while failing to identify those most in need of intervention. The precision-recall results demonstrated that TabNet identified

a large proportion of dropout cases while maintaining a positive predictive value substantially above the 9% baseline prevalence. This finding aligns with machine-learning research showing that multidimensional data can distinguish out-of-school adolescents from enrolled peers (Han et al., 2024).

TabNet performed significantly better than penalized logistic regression and random forest and modestly better than gradient-boosted trees. The difference from logistic regression suggests that dropout risk was not fully represented by linear, additive relationships. The model likely benefited from identifying interactions among variables, such as the particularly high risk associated with simultaneous unexcused absence, low behavioral engagement, low self-efficacy, and limited family support. This is consistent with systematic and comparative evidence indicating that dropout results from multiple interacting individual, family, school, and community factors (Banaag et al., 2024; Gbaguidi & Adetou, 2024). However, the nonsignificant difference between TabNet and gradient-boosted trees also indicates that well-tuned conventional ensemble methods remain strong alternatives. Model choice should therefore consider not only small differences in discrimination but also calibration, interpretability, computational demands, and ease of implementation.

The model's calibration and high negative predictive value strengthen its potential usefulness. A negative predictive value of .985 indicated that students classified as lower risk were very unlikely to drop out during follow-up. This feature may help schools allocate more intensive assessment resources to a smaller group of students. Nevertheless, the positive predictive value of .511 shows that approximately half of those classified as high risk did not drop out. These false-positive classifications should not be viewed solely as model errors. Some students may have been genuinely vulnerable but remained enrolled because of support received during the year, while others may have displayed patterns similar to dropout students without ultimately leaving. Predictive classifications should therefore initiate supportive assessment rather than automatic labeling. Translating predictive analytics into policy requires that schools connect risk estimates to ethical, targeted, and responsive interventions (Kearney et al., 2023).

The feature-importance findings enhance the interpretability of the model. TabNet primarily relied on unexcused absence, overall attendance, behavioral engagement, self-efficacy, and family support, rather than on opaque or sensitive demographic characteristics. This

pattern has practical value because these factors are potentially responsive to intervention. Attendance can be addressed through individualized planning and barrier reduction; engagement can be strengthened through supportive instruction and meaningful participation; self-efficacy can be improved through achievable goals and academic assistance; and family support can be supplemented through school-based mentoring or social services. Students experiencing attendance problems have themselves emphasized the importance of trusting relationships, individualized responses, and coordinated support (Enderle et al., 2024). Likewise, extracurricular participation and structured motivational programs may strengthen belonging and competence among students who are weakly connected to ordinary classroom activities (Clement & Mwila, 2023; Ramsdal & Wynn, 2024).

5. Conclusion

The subgroup analyses produced no evidence of major predictive disparities across gender, grade level, pathway, or receipt of means-tested assistance. This is encouraging because a model with high overall accuracy may still perform poorly for particular groups. However, equivalent statistical performance does not guarantee equitable consequences. Schools may differ in their capacity to respond to high-risk classifications, and students from disadvantaged backgrounds may be more likely to experience surveillance or punitive action. Predictive systems should therefore be embedded in a preventive framework that emphasizes inclusion, continuity, and support across educational stages (Kearney et al., 2024). The model should be used to identify unmet needs, not to restrict access to educational opportunities.

6. Limitations & Suggestions

Several limitations should be considered when interpreting the findings. First, although the study used a longitudinal design and an independent school-level test dataset, participants were recruited from 12 public schools in four French regions, and the findings may not generalize to private schools, overseas territories, alternative education settings, or all national school populations. Second, academic engagement, self-efficacy, and family support were measured through self-report and may have been affected by social desirability, response style, or temporary emotional states. Third, attendance records were more objective but may still have varied in accuracy across

schools because of differences in recording practices. Fourth, the study assessed predictors primarily at baseline and through aggregated attendance data, limiting examination of week-to-week changes in dropout risk. Fifth, the relatively small number of dropout cases restricted the precision of subgroup estimates, particularly for nonbinary students and smaller educational pathways. Finally, although TabNet identified predictive patterns and feature importance, the findings do not establish causal relationships. High-risk indicators may reflect underlying circumstances not included in the model, such as mental health difficulties, academic achievement, bullying, housing instability, teacher support, school climate, or financial pressure.

Future studies should validate the model in larger and more geographically diverse samples, including private schools, rural communities, overseas regions, apprenticeship programs, and alternative educational settings. Research should also use repeated measurements to determine whether changes in attendance, engagement, self-efficacy, and family support improve prediction beyond static baseline values. Additional variables such as grades, course failure, school transitions, disciplinary events, peer relationships, teacher support, mental health, transportation, housing stability, and digital learning participation could be tested to determine whether they improve performance and calibration. External validation should be conducted across academic years to assess sensitivity to policy, curriculum, and social changes. Future research should also compare TabNet with newer interpretable machine-learning approaches and evaluate fairness using sufficiently large subgroup samples. Most importantly, prospective intervention studies should test whether acting on model-generated risk classifications actually reduces dropout, rather than assuming that accurate prediction automatically produces better educational outcomes.

Schools should integrate predictive results into multidisciplinary early-warning procedures involving teachers, counselors, attendance personnel, psychologists, social workers, students, and families. High-risk classifications should lead to a confidential, individualized assessment of the reasons behind absence and disengagement rather than disciplinary action or stigmatizing labels. Because unexcused absence and attendance were the strongest predictors, schools should review attendance patterns frequently and respond before absence becomes chronic. Interventions should also address behavioral engagement, academic confidence, and family

support through tutoring, mentoring, flexible reintegration plans, goal setting, family communication, transportation assistance, mental health referral, and opportunities for meaningful participation in school life. Staff should be trained to interpret probabilities cautiously, document the support offered, and periodically review whether the model performs equitably across student groups. The final decision about support should remain with qualified professionals and should incorporate the student's own perspective, circumstances, strengths, and preferences.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

Authors' Contributions

All authors equally contributed to this article.

References

- Băltătescu, S., & Cernea-Radu, A. E. (2025). Age-Related Variations in School Satisfaction: The Mediating Role of School Engagement. *The Hungarian Educational Research Journal*, 15(1), 67-87. <https://doi.org/10.1556/063.2024.00302>
- Banaag, R., Sumodevilla, J. L., & Potane, J. D. (2024). Factors Affecting Student Drop Out Behavior: a Systematic Review. *International Journal of Educational Management and Innovation*, 5(1), 53-70. <https://doi.org/10.12928/ijemi.v5i1.9396>

- Banda, L. O. L., Banda, J. T., & Banda, C. V. (2024). Unraveling Secondary School Girls' Dropout in Rural Malawi Through Self-Determination. <https://doi.org/10.21203/rs.3.rs-3768943/v1>
- Clement, J., & Mwila, P. M. (2023). Extracurricular Activities: Prospects and Challenges Among Female Students in Secondary Schools in Chanika Ward, Tanzania. *International Journal of Social Science Management and Economics Research*, 01(01), 14-30. <https://doi.org/10.61421/ijssmer.2023.1102>
- Damianidou, E. (2023). Parents as Students: Seeking for the Balance. *International Journal about Parents in Education*. <https://doi.org/10.54195/ijpe.14111>
- Enderle, C., Kreitz-Sandberg, S., Backlund, Å., Isaksson, J., Fredriksson, U., & Ricking, H. (2024). Secondary School Students' Perspectives on Supports for Overcoming School Attendance Problems: A Qualitative Case Study in Germany. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1405395>
- Faisal, N. N. M., Puteri Rohani Megat Abdul, R., Nor Faizah Kaeni Mokh, H., & Bakar, A. H. A. (2023). Understanding School Dropouts Phenomenon Among Youths Through Behavioural Trends. *Environment-Behaviour Proceedings Journal*, 8(23), 3-11. <https://doi.org/10.21834/ebpj.v8i23.4474>
- Gbaguidi, V. E., & Adetou, D. (2024). Factors Affecting School Dropout: Comparative Study of Rural and Urban Settings. *International Journal of Educational Management and Development Studies*, 5(2), 233-256. <https://doi.org/10.53378/353073>
- Guenther, J., Oliver, R., Ober, R., & Holmes, C. (2024). Remote School Retention in Australia. *Australian and International Journal of Rural Education*, 34(3), 73-91. <https://doi.org/10.47381/aijre.v34i3.743>
- Han, Y., Park, J., Song, J., & Kang, D. M. (2024). Distinguishing Characteristics of Out-of-school Adolescents in South Korea: A Machine Learning Approach. *International Journal of Social Welfare*, 34(2). <https://doi.org/10.1111/ijsw.12699>
- Huang, G., & Hai, L. (2026). A Synergistic Social Work-Ethnic Education Intervention for Reducing Dropout Risk Among Male Students in Central Guangxi Zhuang Vocational High Schools: A Mixed-Methods and Quasi-Experimental Study. *Behavioral Sciences*, 16(6), 1023. <https://doi.org/10.3390/bs16061023>
- Kataga, E., & Mwila, P. M. (2024). Strategies for Managing Truancy and Dropout Rates in Secondary Schools: A Study in Temeke Municipality, Tanzania. *Contemporary Research Analysis Journal*, 01(06). <https://doi.org/10.55677/craj/04-2024-vol01i6>
- Kearney, C. A. (2022). Functional Impairment Guidelines for School Attendance Problems in Youth: Recommendations for Caseness in the Modern Era. *Professional Psychology Research and Practice*, 53(3), 295-303. <https://doi.org/10.1037/pro0000453>
- Kearney, C. A., Benoit, L., González, C., Heyne, D., Melvin, G., O'Toole, C., Genitty, C. G., & Keppens, G. (2026). School Attendance and Absenteeism: Towards Next-generation Common Measurement. *Review of Education*, 14(1). <https://doi.org/10.1002/rev3.70146>
- Kearney, C. A., & Childs, J. (2023). Translating Sophisticated Data Analytic Strategies Regarding School Attendance and Absenteeism Into Targeted Educational Policy. *Improving Schools*, 26(1), 5-22. <https://doi.org/10.1177/13654802231174986>
- Kearney, C. A., Dupont, R., Fensken, M., & González, C. (2023). School Attendance Problems and Absenteeism as Early Warning Signals: Review and Implications for Health-Based Protocols and School-Based Practices. *Frontiers in Education*, 8. <https://doi.org/10.3389/feduc.2023.1253595>
- Kearney, C. A., Fensken, M., & Dupont, R. (2024). The K-16 Education Movement: Common Themes Across K-12 and Higher Education Systems to Inform Development and Evaluation. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1272297>
- Kearney, C. A., & González, C. (2022). Unlearning School Attendance and Its Problems: Moving From Historical Categories to Postmodern Dimensions. *Frontiers in Education*, 7. <https://doi.org/10.3389/feduc.2022.977672>
- Khodor, R. M., Valero, O., & Álvarez, I. (2024). Exploring Entrepreneurial Intention and Student Engagement of Youth Living in Poverty. *Behavioral Sciences*, 14(11), 995. <https://doi.org/10.3390/bs14110995>
- Korkmaz, Ö. (2025). Detection of Early School Drop Out in Vocational and Technical High Schools in Turkey. *Sage Open*, 15(3). <https://doi.org/10.1177/21582440251370443>
- Kull, M., Farrell, A. F., Moore, F. O., & Morton, M. H. (2026). Understanding the Links Between Youth Homelessness and Educational Disconnection: New Opportunities for Prevention in the Upstream Project. *Psychology in the Schools*, 63(6), 1033-1048. <https://doi.org/10.1002/pits.70145>
- Kurtylmaz, Y., & Başak, B. E. (2023). Mediating Role of Academic Self-Efficacy Between Insufficient Self-Control and School Dropout. *International Journal of Contemporary Educational Research*, 10(1), 157-170. <https://doi.org/10.33200/ijcer.1181114>
- Mokano, M. F., & Letuma, M. C. (2025). Disciplinary Issues in Special Education: Teachers' Experiences and Attribution. *African Journal of Disability*, 14. <https://doi.org/10.4102/ajod.v14i0.1655>
- Moreira, T., Rosário, P., Ferreira, M. d. A., Pereira, A., & Fuentes, S. (2025). School Engagement and Implications for Future Opportunities: Roma Students' Trajectories. *Frontiers in psychology*, 16. <https://doi.org/10.3389/fpsyg.2025.1535063>
- Morelli, M., Chirumbolo, A., Baiocco, R., & Cattelino, E. (2022). Self-Regulated Learning Self-Efficacy, Motivation, and Intention to Drop-Out: The Moderating Role of Friendships at University. *Current Psychology*, 42(18), 15589-15599. <https://doi.org/10.1007/s12144-022-02834-4>
- Ong, M. L., Klopach, E. T., Carter, S., Simons, R. L., & Beach, S. R. H. (2022). School Disengagement Predicts Accelerated Aging Among Black American Youth: Mediation by Psychological Maladjustment and Moderation by Supportive Parenting. *International journal of environmental research and public health*, 19(19), 12034. <https://doi.org/10.3390/ijerph191912034>
- Petre, F.-B., Teodorescu, C., & Cioclu, A. (2024). School Dropout in Satellite Towns Around Bucharest, Romania. *Social Sciences*, 13(6), 285. <https://doi.org/10.3390/socsci13060285>
- Ramsdal, G. H., & Wynn, R. (2022a). Attachment and School Completion: Understanding Young People Who Have Dropped Out of High School and Important Factors in Their Re-Enrollment. *International journal of environmental research and public health*, 19(7), 3938. <https://doi.org/10.3390/ijerph19073938>
- Ramsdal, G. H., & Wynn, R. (2022b). Attachment and School Completion: Understanding Young People Who Have Dropped Out of High School and Important Factors in Their Re-Enrollment. <https://doi.org/10.31219/osf.io/wsj8h>
- Ramsdal, G. H., & Wynn, R. (2024). Do Two Weeks in a Learning Camp After Ninth Grade Make a Difference? Experiences of Demotivated Boys With an Increased Risk of School Dropout. *Behavioral Sciences*, 14(3), 189. <https://doi.org/10.3390/bs14030189>

- Yusof, R., Harith, N. H. M., Lokman, A., Batau, M. F. A., Zain, Z. M., & Rahmat, N. H. (2023). A Study of Perception on Students' Motivation, Burnout and Reasons for Dropout. *International Journal of Academic Research in Business and Social Sciences*, 13(7). <https://doi.org/10.6007/ijarbss/v13-i7/17187>
- Zhao-hui, C., Chebet, Y. M., & Sampene, A. K. (2025). To Continue or to Quit: Investigating the Role of Affective and Cognitive Engagement on Female School Dropout. *Psychology in the Schools*, 62(9), 3481-3497. <https://doi.org/10.1002/pits.23558>
- Zheng, R. (2024). The Impact of Adverse Childhood Experiences on Educational Outcomes in Adolescents. *Journal of Education Humanities and Social Sciences*, 26, 634-640. <https://doi.org/10.54097/qg7vxm92>