Machine Learning for Predicting Adolescent Health Risk Behaviors

Tresna Dey

Abstract

In the sprawling urban landscapes of Dhaka and the quiet village paths of Jibannagar, adolescence represents a pivotal period of growth and transformation. For nearly 36 million young Bangladeshis—over one-fifth of the nation's population—this decade, encompassing ages 10 to 19, is both a gateway to opportunity and a window of heightened vulnerability. Recent research from early 2026 underscores the stakes: one in six adolescents engages in at least one significant risk-taking behavior, ranging from tobacco use and alcohol consumption to thrill-seeking activities and physical inactivity.

These behaviors, if left unaddressed, can develop into lifelong patterns that compromise health, social development, and economic productivity.

As Bangladesh advances toward its "Smart Bangladesh 2041" vision, the approach to adolescent health is shifting from reactive management to proactive prediction. Rather than waiting for risk behaviors to manifest, Machine Learning (ML) and Predictive Analytics are being harnessed to create a "Digital Sentinel" capable of identifying patterns of vulnerability before they escalate. This data-driven approach allows health authorities, schools, and community organizations to intervene early, offering support precisely where it is most needed and tailoring interventions to the behavioral realities of young people across diverse socio-economic and geographic contexts.

The adolescent risk landscape in Bangladesh is shaped by a complex interplay of geography, gender, socio-economic status, and peer influence. Tobacco use remains particularly persistent among male adolescents, even as overall adult smoking rates decline. School-based surveys conducted between 2025 and 2026 indicate that prevalence among students aged 16 to 19 can be as high as 41.2% in certain regions, highlighting a continuing public health challenge. Gender differences are pronounced: risk-taking behaviors, including drug use and thrill-seeking activities, are significantly more prevalent among boys, particularly within lower-middle-class communities. In contrast, psychological distress, social withdrawal, and indirect forms of risk-taking such as exposure to online hazards or unhealthy dietary behaviors are more

commonly observed among adolescent girls.

Peer dynamics and social pressures further exacerbate vulnerability. Approximately 43% of adolescents report experiencing social or verbal bullying, and this exposure correlates strongly with tobacco and alcohol use, as well as engagement in unsafe behaviors. Mental health also emerges as a critical driver of risk: around 6% of adolescents suffer from major depressive disorder, and those who report feelings of sadness, hopelessness, or social isolation are disproportionately represented among thrill-seekers and substance users. These patterns reveal that risk behaviors are rarely isolated phenomena; they are embedded within broader social and psychological contexts.

Machine Learning provides a powerful framework to decode these complex risk patterns. By leveraging large survey-based datasets, such as the Global School-Based Student Health Survey (GSHS), algorithms are trained to detect clusters of vulnerability that might otherwise remain hidden. Predictive models—including Random Forest and XGBoost—have demonstrated robust accuracy, with Area Under the Curve (AUC) metrics often exceeding 85% when predicting substance use and other health risk behaviors. These models allow researchers and policymakers to quantify the likelihood that an adolescent will engage in one or more high-risk behaviors, enabling precision-targeted interventions.

Feature analysis using techniques such as SHapley Additive exPlanations (SHAP) has provided additional insights into the drivers of adolescent risk. Peer influence consistently emerges as the strongest predictor of substance abuse, followed by lack of parental engagement and lower monthly family income. These findings underscore the importance of addressing the social environment and family dynamics in any intervention strategy. Models are also being refined to account for gender-specific manifestations of risk. For instance, boys are more likely to exhibit overt physical risk-taking, whereas girls may display more subtle signals of psychological distress or indirect risk behaviors. Incorporating these distinctions enhances the predictive accuracy of models and informs tailored intervention design.

A particularly critical insight from ML-driven analysis is the identification of co-occurring risk behaviors. Adolescents who smoke or consume alcohol often exhibit additional risk factors, such as physical inactivity and poor dietary habits, forming distinct “High-Risk Clusters.” Recognizing these clusters allows for comprehensive interventions that address multiple behavioral risks simultaneously, rather than focusing narrowly on a single issue. This holistic approach is essential for effective prevention and health promotion in the adolescent population.

The translation of predictive insights into actionable policy and intervention frameworks is a central goal of this research. The National Strategy for Adolescent Health 2017–2030 provides a policy backdrop against which ML-informed interventions can be deployed. In schools, AI-powered dashboards can assist counselors and health educators in identifying at-risk student segments, enabling targeted awareness

programs on the harms of tobacco, alcohol, and unsafe behaviors. Predictive models also support parental engagement by generating personalized insights that guide “parenting nudges,” encouraging conversations about health and risk with adolescents at critical developmental stages.

Digital literacy has emerged as both a mediator and an intervention point. With adolescents often spending three or more hours daily on smartphones and social media platforms, these tools can influence behaviors both positively and negatively. ML models highlight the potential of digital literacy training to function as a preventive measure, empowering adolescents to critically evaluate online information, resist harmful peer influence, and make informed health decisions. By integrating digital literacy with behavioral monitoring, authorities can intervene before patterns of risky behavior become entrenched.

Community-level interventions are equally important. Insights derived from ML can inform the deployment of local health workers, peer mentoring programs, and school-based support systems, ensuring that high-risk adolescents receive timely guidance and support. These interventions are particularly critical in low-resource areas where formal health infrastructure is limited and where socio-economic pressures may compound behavioral risks.

Ethical considerations remain central to the use of AI and ML in adolescent health. Predictive models must be applied in ways that respect privacy, avoid stigmatization, and maintain the trust of both adolescents and their families. Strategies emphasize human oversight and ethical decision-making: algorithms serve as tools for early identification, not as determinants of punitive action. The objective is to support adolescents, empower families, and guide health educators rather than label or constrain individuals based on predictive outputs.

In conclusion, adolescence in Bangladesh represents both a demographic opportunity and a critical period of vulnerability. Approximately one-sixth of youth engage in behaviors that place their health and future prospects at risk, yet predictive modeling offers the ability to intervene proactively. By integrating Machine Learning with public health policy, schools, and community programs, Bangladesh is transitioning from a reactive approach to adolescent risk toward a predictive, preventive framework. This paradigm shift enables interventions that are timely, precise, and contextually informed, addressing both behavioral patterns and social determinants.

As we navigate 2026, the “Digital Sentinel” created through AI and ML becomes a silent guardian of adolescent health, supporting parents, educators, and policymakers in maintaining the wellbeing of our youth. In doing so, we maximize the potential of Bangladesh’s demographic dividend, ensuring that the nation’s 50 million young citizens grow into a healthy, productive, and resilient population. By harnessing predictive analytics, we are redefining adolescent health strategy—not merely responding to risk, but anticipating it and guiding young people safely through one of the most formative stages of life.