International Journal of Management and Commerce Innovations ISSN 2348-7585 (Online)
Vol. 14, Issue 1, pp: (749-773), Month: April 2026 - September 2026, Available at: www.researchpublish.com

Predicting Youth Violence and Institutional Safety Outcomes through the Integration of Social Behavioral Community and Administrative Risk Indicators

Godbless Amokwaw

Department of Public Administration, Florida State University, Tallahassee, Florida, USA[1]

DOI: https://doi.org/10.5281/zenodo.21994115

Published Date: 18-August-2026

Abstract: **Youth violence and institutional safety incidents emerge from complex interactions among individual behaviour, social exposure, community conditions, prior institutional contact, and administrative records. Conventional prediction models often analyse these domains independently, limiting their capacity to represent nonlinear dependencies, temporal patterns, and cumulative risk pathways. This study proposes a novel *Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA)* for predicting youth violence and institutional safety outcomes through integrated multi-domain risk modelling. The proposed algorithm combines structured social indicators, behavioural histories, community-level vulnerability measures, and administrative risk variables within a weighted feature-fusion architecture incorporating gradient-based feature learning, attention-guided risk weighting, interaction modelling, and calibrated probabilistic classification. The system is designed to generate both individual violence-risk probabilities and institutional safety-risk scores while retaining interpretable contributions from major predictor groups. Comparative experiments evaluate SBCA-RFA against Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and a conventional Multilayer Perceptron. Performance is assessed using accuracy, precision, recall, F1-score, area under the Receiver Operating Characteristic curve, area under the Precision-Recall curve, Brier score, calibration error, false-negative rate, and balanced accuracy. Receiver Operating Characteristic curves, Precision-Recall curves, calibration plots, confusion matrices, feature-importance graphs, risk-distribution plots, and algorithm-comparison charts are used to examine predictive discrimination, reliability, interpretability, and operational usefulness. Ablation experiments further quantify the contribution of social, behavioural, community, and administrative indicator groups to predictive performance. Explainability is incorporated using Shapley Additive Explanations to identify the variables and cross-domain interactions most strongly associated with predicted violence and safety outcomes. The proposed framework is intended to demonstrate that integrated multi-source risk representation can provide stronger discrimination and more reliable early-warning capability than single-domain and conventional machine-learning approaches. The resulting system provides a technically rigorous foundation for data-informed prevention, targeted youth support, institutional resource allocation, and proactive safety management while emphasizing transparent and responsible risk assessment.**

Keywords: **Youth Violence; Institutional Safety; Social Indicators; Behavioral Risk; Administrative Risk.**

1. INTRODUCTION

1.1 Background of the Study

Youth violence has become an increasingly data-intensive public-safety problem because observable precursors are distributed across school participation, peer relations, family conditions, community exposure, behavioral histories, and institutional records rather than concentrated in a single diagnostic variable. Educational systems already generate longitudinal information on attendance, disciplinary episodes, engagement, academic progression, digital participation, and behavioral change, creating opportunities for earlier identification of adverse trajectories. Onwuzurike and Kpogli (2025) show that machine-learning approaches can transform heterogeneous educational interaction data into predictive representations of engagement and behavioral outcomes, particularly when nonlinear relationships and temporal variation are present. For youth-safety research, this principle is directly relevant because declining engagement, recurrent absenteeism, disciplinary escalation, peer conflict, and sudden behavioral shifts may function as interacting signals rather than isolated causes. A technically credible violence-prediction framework therefore requires a data-fusion architecture capable of learning relationships among individual, social, community, and administrative variables while preserving enough interpretability to support defensible interventions rather than opaque labeling. A multidimensional approach is also consistent with developmental evidence that antisocial behavior reflects interacting risk and protective processes across ecological domains. Gubbels et al. (2024), in a meta-analysis of 305 studies covering 1,850 potentially protective factors, identified meaningful protective effects across personal, family, peer, school, and social domains, reinforcing the need to represent youth outcomes as products of multiple systems. The present study extends this logic computationally through the proposed Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA). Instead of treating school misconduct, community disadvantage, prior institutional contact, family instability, or peer exposure as independent predictors, SBCA-RFA is designed to model cross-domain interactions, adaptive domain weights, and nonlinear risk accumulation. The framework will generate calibrated probabilities for youth violence and institutional safety outcomes and compare performance with Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron models. Evaluation will emphasize discrimination, calibration, false-negative control, and interpretability so that predictive performance can be examined alongside practical safety relevance. This framing enables results to identify which combinations of indicators improve prediction and which merely add complexity without practical safety value.

1.2 Youth Violence as a Multidimensional Risk Phenomenon

Youth violence is best understood as a multidimensional developmental outcome produced by interacting individual, relational, institutional, and environmental influences. Violent behavior may be associated with prior aggression, impulsivity, substance use, school disengagement, victimization, family instability, peer delinquency, neighborhood disorder, socioeconomic stress, and previous disciplinary or justice-system contact, but the strength and direction of these relationships vary across individuals and contexts. Bauer et al. (2024) demonstrated this complexity in a longitudinal Brazilian birth cohort, where adverse childhood experiences and clustered social exposures were associated with later gang involvement and broader crime outcomes. Such findings challenge single-factor explanations and support models that represent cumulative and interacting exposures. For predictive purposes, a youth with moderate values on several risk dimensions may present a different safety profile from another youth with one severe indicator, even when traditional additive scores classify both similarly. Capturing these interaction structures requires models capable of learning thresholds, conditional dependencies, and nonlinear combinations rather than relying exclusively on fixed weights. The multidimensional character of youth risk also has a measurement implication: predictors arise from sources with different scales, frequencies, and meanings. Kpogli et al. (2024) illustrate how artificial intelligence can integrate heterogeneous indicators within data-driven K-12 systems to support individualized inference and adaptive decision-making, an approach transferable to safety-oriented analytics when safeguards are added for fairness and interpretability. In the proposed study, social indicators will represent peer, family, and relational conditions; behavioral indicators will capture aggression, disengagement, substance-related behavior, and disciplinary patterns; community indicators will characterize neighborhood disadvantage, violence exposure, and environmental instability; and administrative indicators will include attendance, suspensions, referrals, prior interventions, and related institutional records. SBCA-RFA will encode these domains separately before performing attention-guided fusion, allowing the model to estimate both within-domain importance and cross-domain interaction effects. This architecture is intended to distinguish chronic background vulnerability from dynamic escalation signals and to support risk stratification at individual and institutional levels. Graphical analysis of feature distributions, SHAP values, interaction effects, and risk-score trajectories will make the multidimensional structure visible rather than reducing violence risk to a single unexplained probability.

1.3 Institutional Safety Outcomes and Early-Warning Requirements

Institutional safety outcomes encompass more than the occurrence of a completed violent act. In schools, residential programs, juvenile facilities, and other youth-serving institutions, outcomes may include threats, physical aggression, weapon-related incidents, serious disciplinary events, emergency interventions, repeated conflict, staff or peer injury, and escalation requiring restrictive responses. Effective safety management therefore depends on identifying risk early to enable proportionate prevention. Imbiriba et al. (2023) demonstrated the feasibility of near-term aggression forecasting in psychiatric inpatient youths by combining wearable physiological signals with machine-learning classifiers; their best overall model predicted aggressive behavior three minutes before onset with a mean area under the receiver operating characteristic curve of 0.80. Although the data source and setting differ from school and community systems, the study establishes an important methodological principle: safety outcomes can be approached as forecastable events when temporally informative signals are captured and analyzed. Early-warning capability should consequently be evaluated not only by overall accuracy but also by sensitivity, false-negative rate, lead time, and calibration.

The design philosophy is comparable to predictive assurance systems in other safety-critical domains. Adewale (2026b) describes smart quality-assurance environments in which real-time monitoring, predictive analytics, and evidence-driven control replace delayed, inspection-only responses. Applied cautiously to youth institutions, the analogous principle is that safety systems should move from retrospective incident counting toward continuous, evidence-based risk estimation. The proposed SBCA-RFA operationalizes this transition using routinely available social, behavioral, community, and administrative indicators rather than relying on invasive sensing. Its output will include an individual violence-risk probability and an institutional safety-risk score that can be calibrated into interpretable risk bands. Comparative graphs will examine Receiver Operating Characteristic curves, Precision-Recall curves, confusion matrices, calibration plots, and error distributions across SBCA-RFA and benchmark algorithms. Attention will focus on false negatives because missed high-risk cases may carry greater safety consequences than some false alerts. The intended role of the model is decision support: predictions should trigger review, supportive assessment, or preventive resource allocation, not automatic punitive action, thereby linking early warning to responsible institutional response.

1.4 Social, Behavioral, Community, and Administrative Risk Indicators

Accurate youth-risk prediction depends heavily on how relevant indicators are defined, synchronized, and integrated. Social indicators may include peer-network characteristics, family supervision, social support, victimization, and interpersonal conflict; behavioral indicators may include aggression, impulsivity, substance-related conduct, disengagement, rule violations, and abrupt changes in participation; community indicators may include neighborhood violence, deprivation, instability, and access to protective resources; while administrative indicators may include attendance, suspensions, referrals, prior interventions, incident histories, academic disruption, and justice-system contact. Kanbar et al. (2024) developed an Automated Risk Assessment approach for school violence that used machine learning to assess adolescent violence and aggression risk while explicitly examining potential bias associated with demographic and socioeconomic characteristics. That work illustrates both the predictive value of structured behavioral information and the methodological requirement to test whether risk outputs reflect meaningful safety signals rather than protected or socially patterned attributes. Indicator integration must therefore be accompanied by fairness auditing and transparent feature attribution.

The technical challenge resembles integration problems in other information-intensive systems, where fragmented records reduce analytical reliability unless data are harmonized across sources. Nwokocha et al. (2021) emphasize system integration and interoperable information flow in health-information environments, providing a useful architectural precedent for combining records generated by different operational components. In this study, the four indicator domains will undergo cleaning, missing-value treatment, categorical encoding, scaling where appropriate, temporal alignment, and leakage control. Domain-specific feature encoders will construct latent representations before SBCA-RFA applies adaptive weighting and cross-domain interaction learning. This design prevents a numerically dominant administrative feature group from automatically overwhelming less frequent but important social or community variables. SHAP-based explanations will quantify global and case-level contributions, while ablation experiments will remove each domain in turn to measure its incremental predictive value. Correlation heat maps, interaction plots, feature-importance graphs, and subgroup performance charts will be used to determine whether the integrated model gains accuracy through complementary information. The resulting architecture treats administrative records as one component of a broader ecological signal rather than as a proxy for inherent dangerousness, which is essential for technically and ethically defensible risk assessment.

1.5 Limitations of Conventional Youth Violence Risk Assessment

Conventional youth violence risk assessment commonly depends on structured checklists, professional judgment, historical risk scores, or relatively simple statistical models. These approaches can be useful for organizing evidence, but they may struggle when predictors interact nonlinearly, change over time, contain missing information, or differ substantially across institutional contexts. Parmigiani et al. (2022), in a systematic review of machine-learning approaches to violence-risk prediction, reported a general tendency for machine-learning methods to outperform structured violence-risk instruments in some settings, while also identifying substantial heterogeneity, limited comparability, and concerns about generalizability. The limitation is therefore not simply that traditional instruments are "inaccurate"; rather, fixed scoring structures may compress complex developmental pathways into additive totals and may not exploit interaction patterns among social, behavioral, community, and administrative variables. Moreover, performance reported as a single accuracy or area-under-the-curve value can conceal poor calibration, class imbalance, or clinically important false negatives. A superior framework must consequently be evaluated across several complementary metrics and under strict separation of training, validation, and test data.

A second limitation concerns evidence fragmentation and traceability. Adewale (2026a) argues, in the context of digital assurance, that fragmented evidence trails and weak provenance can undermine the reliability of automated verification. The same systems principle applies to youth-risk analytics: if attendance, disciplinary, community, intervention, and behavioral data are collected inconsistently or merged without provenance controls, sophisticated algorithms may produce precise-looking but unreliable predictions. The proposed SBCA-RFA therefore treats data quality, domain provenance, temporal ordering, and missingness as core modeling considerations rather than preprocessing details. It will be compared with Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron baselines using accuracy, precision, recall, F1-score, balanced accuracy, ROC-AUC, PR-AUC, Brier score, expected calibration error, and false-negative rate. Ablation analysis will test whether each risk domain adds measurable value, while SHAP explanations and subgroup testing will examine interpretability and fairness. This design directly addresses the principal weakness of conventional assessment: limited ability to integrate heterogeneous evidence while demonstrating why a prediction was produced and how reliably it generalizes.

1.6 Machine Learning for Integrated Youth Risk Prediction

Machine learning provides a computational basis for integrated youth-risk prediction because it can represent high-dimensional interactions, nonlinear boundaries, heterogeneous variables, and temporal patterns that are difficult to capture with fixed additive scores. Suh et al. (2025) developed interpretable machine-learning models for predicting criminal offending among adolescents exhibiting moderate-to-severe antisocial behavior and found that gradient boosting produced the strongest performance across tested configurations. Their results demonstrate that machine learning can extract useful predictive structure from youth-level information while still supporting feature-attribution analysis. However, the present problem is broader than predicting a single justice outcome from a limited predictor set. Institutional safety decisions require consideration of social context, observable behavior, community exposure, and administrative history, together with calibrated probabilities and transparent explanations. This motivates a fusion model in which each data domain is represented separately before cross-domain relationships are learned, rather than concatenating all variables into an undifferentiated feature matrix.

The proposed SBCA-RFA draws conceptually from multiscale analytics, where signals at different resolutions are decomposed and recombined to preserve complementary information. James and Akujuobi (2026) show, in a Zero Trust security setting, how multiscale signal processing can improve anomaly detection by retaining patterns that would be obscured under a single-resolution representation. Translating that computational principle to youth-risk modeling, SBCA-RFA will learn domain-specific embeddings, calculate weights, model pairwise and higher-order interactions, and fuse the resulting representations through an attention-guided risk layer. The output layer will estimate calibrated probabilities for youth violence and institutional safety outcomes. Model development will use separated training, validation, and holdout test partitions, with hyperparameters selected without access to final-test labels. Performance will be benchmarked against Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron. Comparative bar charts, ROC and Precision-Recall curves, calibration plots, confusion matrices, SHAP summaries, and ablation graphs will show where performance gains arise. The central technical hypothesis is that structured cross-domain fusion will improve discrimination and calibration without sacrificing interpretability, particularly for cases whose risk is distributed across several moderate indicators rather than dominated by one extreme variable.

1.7 Problem Statement

Despite availability of youth data, institutions still lack an integrated, interpretable method for converting fragmented social, behavioral, community, and administrative information into reliable forecasts of violence and safety outcomes. Existing systems frequently emphasize one data family, such as disciplinary history, clinical assessment, academic engagement, or prior offending, and consequently risk overlooking interactions that may be decisive for escalation. Rodriguez et al. (2026)

demonstrated that machine-learning models using clinical and neuroanatomical information could predict future homicide among high-risk juvenile offenders with meaningful discrimination, but the authors also cautioned that further validation is required before such methods inform probation, parole, or sentencing decisions. This illustrates the central problem confronting predictive youth-safety research: strong classification performance in a narrow sample does not automatically translate into generalizable, ethically deployable institutional decision support. Models must be tested for calibration, subgroup stability, interpretability, and operational error costs in addition to discrimination.

At the same time, K-12 systems use data-informed artificial intelligence to identify learning risks and support adaptive decision-making. Onwuzurike and Kpogli (2022) argue that predictive analytics can convert educational data into actionable intelligence for early identification and strategic intervention, but a safety-focused implementation requires a more integration of ecological and administrative risk evidence. The problem addressed by this study is therefore the absence of a unified predictive framework that can fuse social, behavioral, community, and administrative indicators, learn nonlinear cross-domain relationships, produce calibrated violence and institutional safety probabilities, and explain the contribution of each predictor domain. To address this gap, the study develops SBCA-RFA and evaluates it against Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron models. The evaluation will combine ROC-AUC and PR-AUC with recall, F1-score, balanced accuracy, false-negative rate, Brier score, calibration error, ablation analysis, and SHAP-based interpretation. The research problem is consequently both predictive and operational: improving early-warning performance while ensuring that model outputs remain transparent enough to support preventive, proportionate, and reviewable institutional action.

1.8 Research Objectives and Research Questions

Research Objectives

1. To develop an integrated analytical representation of social, behavioral, community, and administrative risk indicators for predicting youth violence and institutional safety outcomes.
2. To develop the novel Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA) using domain-specific feature learning, adaptive risk weighting, cross-domain interaction modeling, attention-guided feature fusion, and probability calibration.
3. To comparatively evaluate SBCA-RFA against Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron using accuracy, precision, recall, F1-score, balanced accuracy, ROC-AUC, PR-AUC, false-negative rate, Brier score, and expected calibration error.
4. To determine the individual and combined predictive contributions of social, behavioral, community, and administrative indicators using feature-importance analysis, SHAP explanations, interaction analysis, and domain-level ablation experiments.
5. To evaluate the interpretability, calibration, subgroup consistency, fairness, and operational suitability of SBCA-RFA as an early-warning decision-support framework for youth violence prevention and institutional safety management.

Research Questions

1. To what extent can the integration of social, behavioral, community, and administrative indicators improve the prediction of youth violence and institutional safety outcomes?
2. How effectively does the proposed SBCA-RFA model capture nonlinear and cross-domain risk relationships compared with conventional statistical and machine-learning algorithms?
3. How does SBCA-RFA perform relative to Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron across discrimination, calibration, classification, and false-negative performance metrics?
4. Which social, behavioral, community, and administrative indicators, and which interactions among these domains, contribute most strongly to predicted youth violence and institutional safety risk?
5. To what extent does SBCA-RFA provide calibrated, interpretable, fair, and operationally useful predictions across relevant youth subgroups and institutional risk categories?

1.9 Contributions and Significance of the Study

This study contributes to youth violence analytics by introducing SBCA-RFA as a purpose-built multi-domain risk-fusion algorithm rather than applying a conventional classifier directly to a concatenated dataset. Its principal methodological contribution is the separation of social, behavioral, community, and administrative information into domain-specific representations followed by adaptive weighting, cross-domain interaction learning, attention-guided fusion, and calibrated probability estimation. The study also contributes a rigorous comparative evaluation framework in which the proposed algorithm is assessed against six established statistical and machine-learning methods using discrimination, calibration, class-sensitive performance, and false-negative metrics rather than accuracy alone. Feature-importance analysis, SHAP explanations, ablation experiments, subgroup comparisons, and graphical diagnostics further establish whether improved prediction results from meaningful integration of complementary indicators. Practically, the framework is significant because it is designed to support early identification of escalating risk, targeted preventive services, institutional resource prioritization, safety planning, and evidence-informed case review without converting algorithmic outputs into automatic punitive decisions. The study therefore connects technical predictive performance with transparency, fairness, calibration, and responsible institutional use.

1.10 Scope of the Review and Structure of the Paper

The study focuses on predicting youth violence and institutional safety outcomes through integrated analysis of social, behavioral, community, and administrative indicators. Its analytical scope covers data preprocessing, feature engineering, domain-specific representation, cross-domain risk interaction modeling, development of SBCA-RFA, benchmark model construction, hyperparameter optimization, training-validation-test separation, probability calibration, explainability, ablation testing, and comparative performance evaluation. The outcome space includes violence-related and institutionally relevant safety events that can be represented as supervised prediction targets, while the framework deliberately excludes automated punitive decision-making and treats predictions as decision-support signals requiring professional review. Section 1 establishes the research background, multidimensional risk rationale, safety requirements, problem, objectives, questions, and significance. Section 2 reviews conceptual, theoretical, empirical, and machine-learning literature and identifies the research gap. Section 3 specifies the system architecture, data representation, mathematical formulation of SBCA-RFA, experimental design, benchmark algorithms, and evaluation metrics. Section 4 presents graphical and statistical comparisons, calibration, feature attribution, ablation, fairness, and interpretation of predictive results. Section 5 synthesizes the major findings and presents recommendations for responsible youth violence prevention, institutional safety management, model deployment, and future research.

2. LITERATURE REVIEW

2.1 Conceptual Review of Youth Violence and Institutional Safety

Youth violence is conceptually broader than isolated physical assault because it encompasses aggressive, threatening, coercive, weapon-related, retaliatory, and other behaviors capable of causing physical or psychological harm within schools, residential facilities, juvenile institutions, and surrounding communities. Institutional safety consequently represents the capacity of a youth-serving organization to prevent, detect, assess, and respond proportionately to such threats while sustaining an environment conducive to development, education, and social participation as represented in figure 1. A public-health perspective views violence as an outcome emerging from interacting individual, relational, organizational, and community conditions rather than an intrinsic characteristic of a young person. April et al. (2023) similarly integrate public-health and social-ecological perspectives in juvenile justice, emphasizing multilevel youth needs, while Axford et al. (2023) demonstrate that violence-prevention effectiveness depends partly on meaningful engagement with young people and responsiveness to their circumstances. Community partnership models further indicate that distributed organizations can improve outcomes when information, expertise, and interventions are coordinated rather than fragmented (Ijiga et al., 2024).

Within the present study, institutional safety is operationalized as a predictive outcome linked to both the probability of youth violence and the likelihood of safety-relevant events such as serious disciplinary escalation, interpersonal aggression, threats, emergency intervention, or repeated institutional conflict. This definition supports a preventive rather than purely punitive orientation. Community-based intervention research demonstrates the value of combining population-level risk recognition with accessible targeted services, particularly where structural disadvantage affects exposure and access to preventive resources (Idoko et al., 2024). Data-driven educational systems similarly demonstrate that heterogeneous learner information can be integrated to identify patterns requiring differentiated support (Kpogli et al., 2024). Accordingly, the proposed Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA) conceptualizes youth violence as an emergent multi-domain outcome. It converts social conditions, behavioral histories, community exposure, and administrative records into calibrated risk estimates while retaining interpretable domain contributions. This conceptualization permits graphical comparison of risk distributions, model discrimination, calibration, and predictor importance without equating elevated predicted probability with predetermined violent behavior.

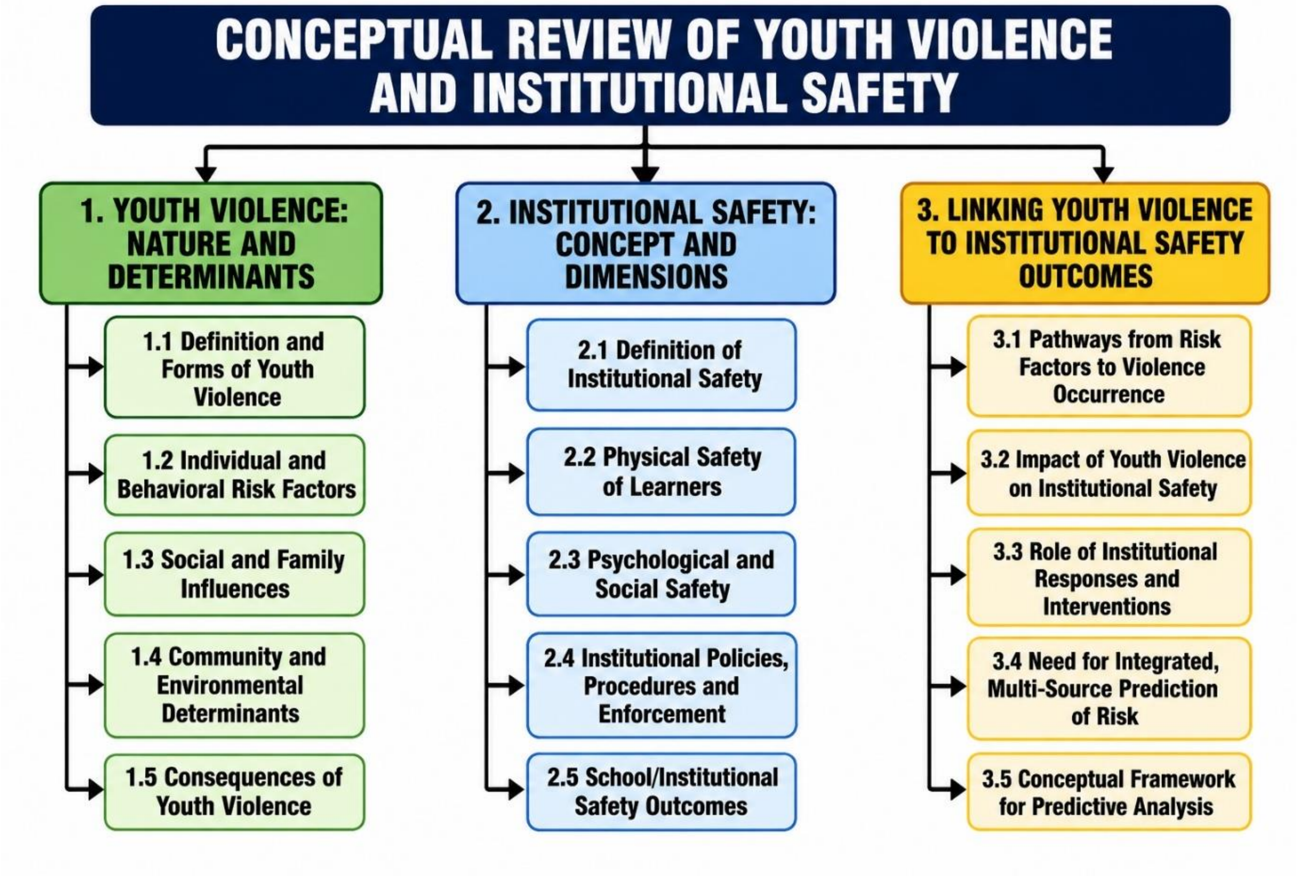

Figure 1. Conceptual Framework Linking Youth Violence Determinants, Institutional Safety Dimensions, and Integrated Risk Prediction

Figure 1 presents the conceptual relationship between youth violence and institutional safety through three interconnected branches. The first branch, Youth Violence: Nature and Determinants, defines youth violence broadly and traces its major sources through individual and behavioral risk factors, social and family influences, community and environmental conditions, and the eventual consequences of violent behavior. This branch emphasizes that violence is not produced by one isolated characteristic but develops through interacting personal and contextual conditions. The second branch, Institutional Safety: Concept and Dimensions, represents safety as a multidimensional institutional condition encompassing the physical protection of learners, psychological and social safety, effective policies and procedures, enforcement mechanisms, and measurable school or institutional safety outcomes. The third branch, Linking Youth Violence to Institutional Safety Outcomes, integrates the first two branches by showing how accumulated risk factors can progress toward violence occurrence, which can subsequently affect institutional safety through threats, aggression, injuries, disciplinary incidents, disruption, and emergency interventions. Importantly, this branch also incorporates institutional responses and preventive interventions, demonstrating that the relationship between risk and violence is not deterministic because timely institutional action can modify the trajectory. The final elements, integrated multi-source prediction of risk and the conceptual framework for predictive analysis, provide the direct connection to the proposed SBCA-RFA model. They establish the rationale for combining social, behavioral, community, and administrative information rather than relying on a single class of indicators when predicting youth violence and institutional safety outcomes.

2.2 Social, Behavioral, Community, and Administrative Risk Indicators

The proposed prediction framework organizes youth-related information into four complementary indicator domains. Social indicators represent family supervision, peer relationships, victimization, social support, conflict networks, and interpersonal instability. Behavioral indicators capture aggression, impulsivity, substance-related conduct, rule violations, disengagement, repeated threats, and changes in behavioral patterns. Community indicators represent neighborhood violence, deprivation, social disorder, environmental instability, and protective collective efficacy, whereas administrative indicators include attendance, suspensions, disciplinary referrals, prior interventions, academic disruption, incident reports, and previous institutional contact. Evidence supports treating these dimensions as interacting rather than independent. Carr et al. (2024), for example, found that combinations of conduct problems and other childhood conditions were more informative for subsequent delinquency than isolated exposure alone, illustrating why interaction terms may improve youth-risk estimation. Community context is equally important: Gard et al. (2022) showed that geographically proximal gun violence was associated with adolescent behavioral outcomes, while stronger neighborhood collective efficacy operated in a protective direction.

Technically, these indicator groups present different scales, missingness patterns, temporal resolutions, and predictive distributions. Robust modeling therefore requires more than simple feature concatenation. Atalor (2022) demonstrates the broader predictive principle that data-driven models depend strongly on disciplined feature representation and relationships between descriptors and outcomes, while Ilesanmi et al. (2024) illustrate how heterogeneous performance variables can be integrated within optimization-oriented analytical systems. For youth-safety prediction, administrative reliability is especially critical because inconsistent identifiers, duplicate incidents, delayed entries, and incompatible coding schemes can distort risk estimates. Automated extract-transform-load architectures offer a methodological basis for standardization, validation, transformation, and quality control across heterogeneous records (Nwokocha et al., 2022). SBCA-RFA therefore processes each domain separately before adaptive fusion, enabling domain-specific representations and cross-domain interactions to be learned without allowing a large administrative feature set to dominate smaller social or community domains. Subsequent SHAP analysis, ablation graphs, correlation structures, and calibrated probability plots can determine which indicators genuinely improve predictive performance and which provide redundant information.

2.3 Theoretical Foundations of Youth Violence and Institutional Safety Prediction

The theoretical foundation of the proposed SBCA-RFA framework is principally grounded in social-ecological, social-learning, developmental, and risk-protective perspectives. Social-ecological theory positions youth behavior within nested systems in which individual characteristics interact continuously with family, peer, school, neighborhood, organizational, and broader structural conditions. Bobba et al. (2023) demon strate the analytical importance of distinguishing proximal micro- and mesosystem influences from neighborhood, socioeconomic, cultural, and other exo- and macrosystem conditions when explaining adolescent outcomes as presented in table 1. This structure maps directly onto SBCA-RFA: behavioral indicators approximate individual-level processes; social indicators represent proximal relationships; institutional administrative indicators describe organizational interactions; and community indicators encode broader ecological exposures. Social-learning principles complement this framework by proposing that behavioral patterns may be acquired and reinforced through observation, modeling, peer interaction, and perceived consequences. Evidence synthesis involving vulnerable young people identifies social learning, positive youth development, attachment, social support, and organizational context as relevant theoretical mechanisms for designing interventions across interpersonal and institutional levels (Evans et al., 2024).

These theories justify a computational architecture based on **interaction rather than isolated association**. Risk is not presumed to increase uniformly whenever one adverse variable is present; its meaning can depend on the surrounding configuration of protective and adverse conditions. This systems logic parallels data-driven optimization frameworks in which multiple operational signals must be integrated to explain overall outcomes (Enyejo et al., 2024), while automated intelligence platforms demonstrate the value of converting dispersed observations into structured decision evidence (Anokwuru et al., 2024). Because SBCA-RFA relies on sensitive administrative and behavioral records, its theoretical architecture must also include information-governance principles concerning confidentiality, authorized access, traceability, and controlled data processing, consistent with enterprise data-protection approaches discussed by Onyekaonwu et al. (2022). Consequently, SBCA-RFA operationalizes theory through domain-specific feature encoding, adaptive weighting, cross-domain interaction learning, attention-guided fusion, and probability calibration. The model is therefore designed not merely to maximize classification accuracy, but to represent how interacting ecological conditions produce heterogeneous risk trajectories and to explain those trajectories through SHAP attribution, domain ablation, calibration analysis, subgroup testing, and comparative algorithmic evaluation.

Table 1. Summary of Theoretical Foundations of Youth Violence and Institutional Safety Prediction

Theoretical Foundation	Core Proposition	Relevance to Youth Violence and Institutional Safety	Application in the Proposed SBCA-RFA Framework
Social-Ecological Theory	Youth behavior develops through interacting influences at individual, family, peer, school, community, and wider societal levels.	Explains why violence cannot be attributed to a single factor. Behavioral tendencies may interact with family instability, peer conflict, institutional experiences, and neighborhood violence to influence safety outcomes.	Provides the structural basis for separating predictors into **social, behavioral, community, and administrative domains** and subsequently modeling their cross-domain interactions.
Social Learning Theory	Behaviors are acquired and reinforced through observation, imitation, social interaction, rewards, and perceived consequences.	Helps explain how aggressive behavior may develop through exposure to violent peers, family conflict, community violence, or repeated reinforcement of antisocial conduct.	Supports inclusion of variables such as peer aggression, previous behavioral incidents, exposure to violence, disciplinary history, and changing behavioral patterns within the predictive model.
Developmental Perspective	Risk and protective influences accumulate and change across developmental stages, meaning that early experiences can shape later behavioral trajectories.	Explains why previous aggression, school disengagement, victimization, substance-related behavior, and repeated institutional contact may progressively increase vulnerability to later violence.	Motivates the use of **longitudinal and temporal features**, including changes in disciplinary frequency, attendance patterns, previous interventions, and escalation indicators rather than relying only on static measurements.
Risk and Protective Factor Framework	Youth outcomes are determined by the balance and interaction between factors that increase vulnerability and factors that reduce or buffer risk.	Recognizes that similar risk exposures can produce different outcomes depending on social support, family supervision, positive peer relationships, community resources, and institutional intervention.	Supports adaptive domain weighting and interaction modeling in **SBCA-RFA**, allowing protective and adverse indicators to modify one another when generating calibrated youth-violence probabilities and institutional safety-risk scores.

2.4 Data-Driven and Machine Learning Approaches to Youth Risk Prediction

Data-driven youth risk prediction transforms heterogeneous observations into estimates of future violence, aggression, disciplinary escalation, or related safety outcomes. Machine learning is useful where predictors exhibit nonlinear associations, threshold effects, class imbalance, and interactions that conventional additive scores may not represent efficiently. Onyekaonwu et al. (2019) illustrate the broader transition from retrospective description to predictive analytics through artificial intelligence and machine learning, while Ajiboye et al. (2025) demonstrate the importance of behavioral, psychosocial, and digital-context information when assessing vulnerable youth populations as represented in figure 2. In education, Onwuzurike and Kpogli (2025) similarly show how learner engagement and behavioral data can be modeled computationally to anticipate outcomes. For youth safety, this supports architectures that combine attendance trajectories, disciplinary events, peer and family indicators, community exposures, prior interventions, and behavioral histories instead of training models on a single administrative stream.

Empirical machine-learning research also demonstrates why mo del design must extend beyond raw classification accuracy. Su et al. (2020) used longitudinal electronic health records from 41,721 youths and obtained area-under-the-curve values of 0.81–0.86 across suicide-risk prediction windows, showing that routinely collected records can yield actionable temporal risk estimates. Zang et al. (2024), however, found that transportability across datasets was inconsistent and that complex long short-term memory models did not systematically outperform regularized logistic regression, highlighting the danger of assuming that algorithmic complexity guarantees generalization. The proposed Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA) therefore uses domain-specific encoders, adaptive domain weighting, cross-domain interaction learning, attention-guided fusion, and calibrated probabilistic outputs. Its evaluation will compare

Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron baselines using ROC-AUC, PR-AUC, recall, F1-score, balanced accuracy, false-negative rate, Brier score, and calibration error. SHAP explanations, ablation graphs, calibration curves, confusion matrices, and subgroup analyses will determine whether performance gains arise from meaningful multi-domain integration rather than overfitting.

Figure 2 summarizes the data-driven and machine-learning pathway for youth risk prediction by organizing the analytical process into two complementary branches. The Data-Driven Approaches branch begins with data collection and integration, where social, behavioral, community, and administrative information such as family conditions, peer relationships, behavioral incidents, neighborhood exposure, attendance, disciplinary records, and previous institutional interventions are assembled into a unified analytical dataset. Its second subbranch, feature engineering and risk representation, converts these heterogeneous records into usable predictive variables through cleaning, missing-value treatment, encoding, scaling, temporal trend extraction, and construction of cross-domain interaction features. The Machine Learning Approaches branch then applies predictive algorithms including Logistic Regression, Support Vector Machine, Random Forest, XGBoost, LightGBM, Multilayer Perceptron, and the proposed SBCA-RFA. The second machine-learning subbranch evaluates these models using accuracy, recall, F1-score, ROC-AUC, PR-AUC, calibration, explainability, and related performance measures. Both branches ultimately converge on a common outcome: calibrated early identification of youth violence risk and institutional safety threats. This structure reflects the central methodology of the study because SBCA-RFA is designed not merely to classify youths but to fuse multiple risk domains, model their interactions, generate interpretable probabilities, and provide evidence for preventive institutional decision support.

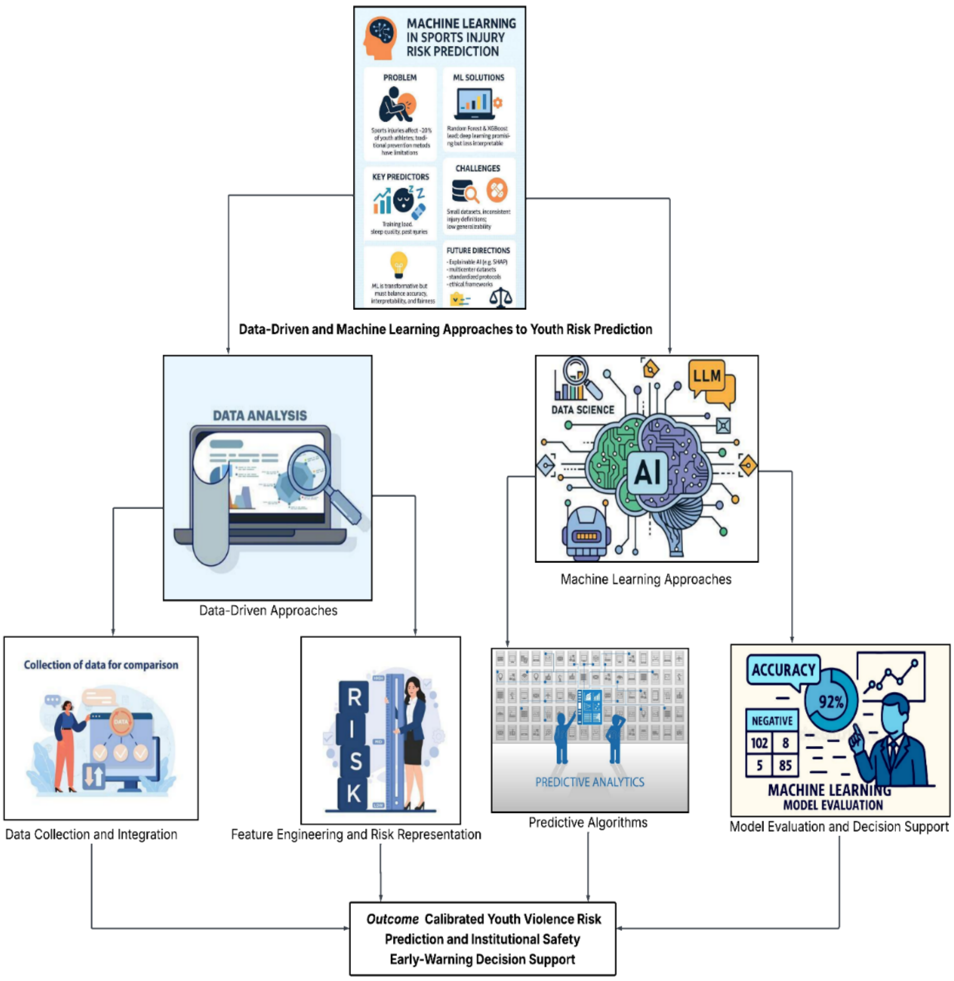

Figure 2. Data-Driven and Machine-Learning Framework for Integrated Youth Violence Risk Prediction and Institutional Safety Decision Support.

2.5 Empirical Review of Existing Youth Violence and Institutional Safety Prediction Models

Empirical youth violence prediction models vary substantially in outcome definition, data source, sample composition, and validation strategy. Ajiboye et al. (2025) connect community gun violence with psychological and transgenerational trauma pathways, demonstrating that exposure history can shape youth vulnerability beyond immediately observable misconduct. Ijiga et al. (2024) further emphasize that conflict exposure, media environments, and institutional advocacy structures influence youth mental-health conditions, whereas Onyekaonwu and Peter-Anyebe (2019) show how technology-enabled youth interventions can generate structured participation and developmental information relevant to prevention. These findings imply that violence prediction should not be reduced to prior offending alone. Useful models should integrate proximal behavior with psychosocial context, community adversity, institutional engagement, and protective participation, while recognizing that similar administrative histories may arise from different developmental pathways.

Recent predictive studies provide benchmarks but reveal methodological constraints. Aseltine et al. (2025) compared pediatric suicide-risk algorithms with in-person screening in 19,653 youths and found that the algorithm identified a larger proportion of subsequent suicide attempts, illustrating the value of longitudinal administrative data for early warning. Wong et al. (2023) applied LASSO models to 230 features in a community youth cohort and demonstrated shared predictive information between proneness to aggression and general psychopathology, with isolation and impulsivity emerging as central correlates. These models support multidimensional prediction, yet they remain focused on clinical or psychological outcomes and do not jointly optimize violence and institutional safety prediction from social, behavioral, community, and administrative domains. SBCA-RFA addresses this limitation through structured feature fusion and interaction modeling. The results analysis will compare discrimination, calibration, class-sensitive error, and interpretability across algorithms, rather than presenting accuracy alone. Receiver Operating Characteristic curves, Precision-Recall curves, confusion matrices, calibration plots, SHAP summaries, and domain ablation experiments will reveal whether integrated risk information improves sensitivity to high-risk cases while maintaining acceptable false-positive behavior and transparent attribution.

2.6 Research Gaps and Motivation for the Proposed Social-Behavioral-Community-Administrative Risk Fusion Algorithm

The literature reveals a persistent gap between evidence that youth risk is multidimensional and the architecture of many existing prediction systems. Studies commonly model a single outcome from one dominant data environment, such as clinical records, justice histories, surveys, or school discipline files, thereby underrepresenting interactions among social relationships, behavioral change, community exposure, and institutional contact as represented in figure 3. Oloba et al. (2025) demonstrate the value of inclusive digital platforms that integrate participation signals from populations whose experiences may otherwise be excluded from formal decision systems. Ijiga et al. (2024) similarly show that youth mental-health vulnerability is conditioned by broader political and social environments, while Idoko et al. (2024) emphasize the importance of community programs where access and structural constraints shape outcomes. Applied to violence prediction, these observations expose a representation gap: risk models may be technically sophisticated yet incomplete if contextual and protective variables that alter the meaning of administrative events are omitted.

A second gap concerns generalization, interpretability, and intervention relevance. Tate et al. (2022) used a stacked ensemble combining gradient boosting, random forest, elastic net, and neural networks to predict aggressive and suicidal behavior, obtaining transferable performance but concluding that the model was not ready for clinical use. Somerville et al. (2026) linked school intervention data with administrative outcomes and used machine learning to identify predictors of suspension and juvenile arrest, illustrating the value of connected institutional evidence. Existing studies rarely provide a unified fusion mechanism that quantifies contributions of social, behavioral, community, and administrative domains while testing calibration, false-negative risk, fairness, and explanation stability. This motivates SBCA-RFA. The algorithm will learn domain representations, assign weights, estimate cross-domain interactions, and produce violence and institutional safety probabilities. Comparative benchmarking, ablation analysis, SHAP explanations, subgroup evaluation, and calibration testing will determine whether performance gains remain interpretable and balanced for responsible decision support.

Figure 3 highlights three major research gaps affecting current approaches to youth violence and institutional safety prediction. Branch 1, Fragmented and Single-Domain Risk Modeling, shows that many existing assessments examine behavioral indicators, administrative records, social conditions, and community exposure separately, which limits their ability to represent the multidimensional nature of youth risk. It further identifies weak cross-domain representation, where additive scoring methods may fail to capture nonlinear dependencies, protective effects, and combinations of moderate-risk indicators that collectively signal escalating vulnerability. Branch 2, Limitations of Existing Prediction Algorithms, emphasizes that conventional statistical approaches such as Logistic Regression are constrained by fixed linear assumptions, while machine-learning models such as Support Vector Machine, Random Forest, XGBoost, LightGBM, and Multilayer Perceptron may process heterogeneous variables without explicitly preserving their domain structure. This can restrict adaptive weighting and the modeling of meaningful interactions among social, behavioral, community, and administrative indicators. The branch also identifies performance-evaluation weaknesses, particularly excessive reliance on overall accuracy, insufficient treatment of class imbalance and false-negative errors, and inadequate probability calibration. Branch 3, Explainability, Fairness, and Institutional Deployment Gaps, addresses limitations in translating predictive models into responsible practice. Black-box predictions may provide insufficient explanations for individual risk classifications, while subgroup performance differences can introduce inequitable outcomes across populations. Limited external validation can further reduce confidence when models are transferred between institutions. Operationally, these deficiencies make it difficult for practitioners to translate risk estimates into proportionate preventive interventions, safety planning, and resource allocation while avoiding inappropriate automated disciplinary decisions.

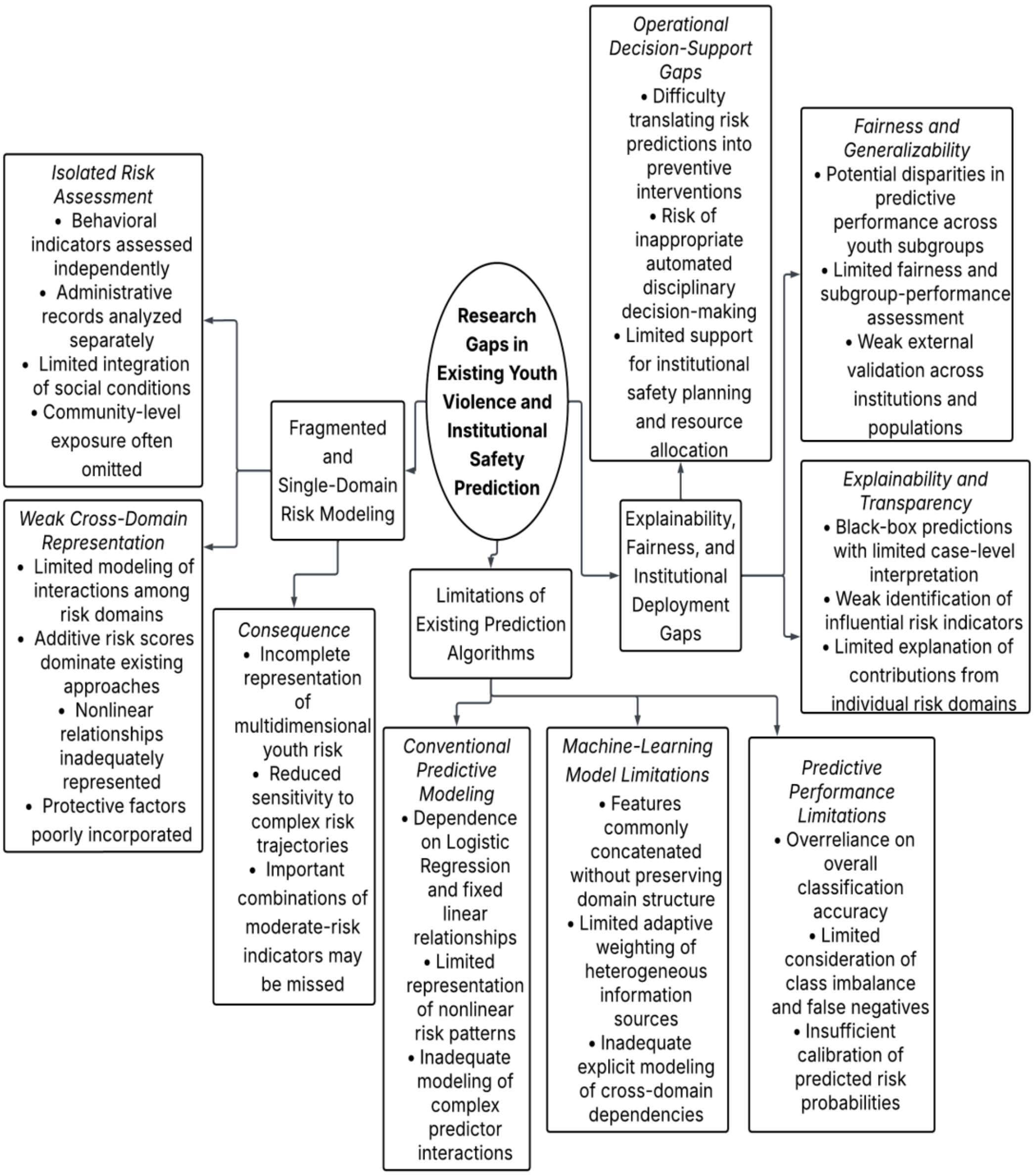

Figure 3. Research Gaps in Existing Youth Violence and Institutional Safety Prediction: Fragmented Risk Modeling, Algorithmic Limitations, and Deployment Challenges

3. SYSTEM MODEL DESCRIPTION

3.1 Proposed Social-Behavioral-Community-Administrative Risk Fusion Framework and System Architecture

The proposed Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA) is designed as a multi-domain predictive architecture that estimates both individual youth-violence probability and aggregate institutional safety risk. For youth i, the complete observation is represented as $X_i = \{x_i^s, x_i^B, x_i^c, x_i^A\}$, where x_i^s, x_i^B, x_i^c, and x_i^A denote social, behavioral, community, and administrative feature vectors, respectively. Rather than concatenating heterogeneous variables immediately, each domain is first transformed independently:

$$h_i^d = \phi\left(W_d x_i^d + b_d\right), \qquad d \in \{S, B, C, A\} \tag{1}$$

where h_i^d represents the latent representation of domain d; W_d represents the domain-specific learnable weight matrix; b_d denotes its bias vector; and $\varphi(\cdot)$ shows a nonlinear activation function such as ReLU. This architecture prevents a high-dimensional administrative domain from overwhelming smaller but potentially important community or social domains.

The model next assigns an adaptive relevance weight to every domain:

$$\alpha_i^d = \frac{\exp\left(e_i^d\right)}{\sum_{k\in D} \exp\left(e_i^k\right)}, \qquad e_i^d = v^T \tanh\left(U h_i^d + c\right) \tag{2}$$

where D = {S,B,C,A}; e_i^d captures the unnormalized attention score; α_i^d represents the normalized importance of domain d; U and v denote trainable parameters; and c represents a bias vector. Thus, two youths with identical disciplinary histories may receive different risk estimates if their social support, community exposure, or behavioral trajectories differ.

Cross-domain interaction is represented as

$$g_i = \sum_{d<k} h_i^d \odot h_i^k \tag{3}$$

where $\odot$ denotes element-wise multiplication. The fused representation becomes

$$z_i = \left[\sum_{d\in D} \alpha_i^d\, h_i^d\, ;\, g_i\right] \tag{4}$$

where z_i combines attention-weighted domain information with pairwise interactions. This architecture directly implements the paper's central proposition that youth violence arises from interacting rather than isolated risk domains.

3.2 Data Acquisition, Preprocessing, Feature Engineering, and Risk Indicator Representation

The analytical dataset should integrate records at the individual-youth level using a pseudonymized identifier and a clearly defined prediction window. Social variables may include family supervision, peer conflict, victimization, social support, and household instability. Behavioral variables may include previous aggression, threats, substance-related conduct, rule violations, disengagement, and behavioral change. Community variables can include neighborhood violence exposure, socioeconomic deprivation, residential instability, community cohesion, and access to protective services. Administrative variables may comprise attendance, suspensions, disciplinary referrals, previous interventions, academic disruption, incident frequency, and documented institutional contact.

To prevent information leakage, all preprocessing parameters must be estimated from the training data only. Continuous predictor x_{ij} is standardized as

$$x_{ij}^{*} = \frac{x_{ij} - \mu_j^{\text{train}}}{\sigma_j^{\text{train}}} \tag{5}$$

where x_{ij} shows feature j for youth i; μ_j^{train} captures the training-set mean; σ_j^{train} represents the corresponding standard deviation; and x_{ij}^* denotes the standardized value. Missing continuous values should be imputed using training-set medians, while categorical variables may use the training-set mode or a separate "unknown" category. Categorical predictors are one-hot encoded where appropriate.

Temporal administrative behavior is represented using change-sensitive features. For example, disciplinary escalation can be defined as

$$\Delta D_i = D_i^{(\text{recent})} - D_i^{(\text{previous})} \tag{6}$$

where $D_i^{(\text{recent})}$ represents disciplinary-event frequency within the recent observation period and $D_i^{(\text{previous})}$ represents the preceding period. A positive ΔD_i therefore captures escalation rather than merely historical accumulation.

Class imbalance is handled through weighted training:

$$w_c = \frac{N}{2N_c} \tag{7}$$

where N shows the total number of training observations and N_c captures the number belonging to class c. This increases the contribution of relatively rare violence cases to model optimization. The approach is important because a model predicting "no violence" for almost every youth could achieve misleadingly high accuracy in a low-prevalence dataset. The resulting domain-specific representations provide the structured inputs required by SBCA-RFA while preserving distinctions among social, behavioral, community, and administrative evidence.

3.3 Development and Mathematical Formulation of the Novel Social-Behavioral-Community-Administrative Risk Fusion Algorithm

Following domain encoding and fusion, SBCA-RFA converts the fused representation z_i into an uncalibrated individual violence probability:

$$\hat{p}_i = \sigma(w_o^T z_i + b_o) = \frac{1}{1 + \exp[-(w_o^T z_i + b_o)]} \tag{8}$$

where $\hat{p}_i$ denotes the initial probability of a violence-related outcome for youth i; w_o represents the output-layer coefficient vector; b_o captures the output bias; and $\sigma(\cdot)$ represents the sigmoid function. Since accurate probability estimates are essential for operational risk stratification, post-training logistic calibration is applied:

$$p_i^{\text{cal}} = \sigma\left[a \log\left(\frac{\hat{p}_i}{1 - \hat{p}_i}\right) + b\right] \tag{9}$$

where p_i^{cal} represents the calibrated probability and a and b are calibration parameters estimated from validation data.

Model parameters are learned by minimizing class-weighted binary cross-entropy with regularization:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N} w_{y_i}\left[y_i \log(\hat{p}_i) + (1 - y_i)\log(1 - \hat{p}_i)\right] + \lambda \parallel \Theta \parallel_2^2 \tag{10}$$

where N represents the number of training observations; $y_i \in \{0,1\}$ shows the observed outcome; w_{y_i} is the class weight; Θ represents all trainable model parameters; and λ controls L_2 regularization.

For institution j, individual probabilities are aggregated into a normalized safety-risk index:

$$ISR_j(t) = 100\,\frac{\sum_{i \in j} r_i\,(t) p_i^{\text{cal}}}{\sum_{i \in j} r_i\,(t)} \tag{11}$$

where $ISR_j(t)$ captures the institutional safety-risk score at time t; p_i^{cal} shows individual calibrated risk; and $r_i(t)$ denotes a predefined temporal relevance weight. A score approaching 100 indicates greater aggregated predicted risk, whereas values approaching zero indicate lower modeled risk. Importantly, this score represents decision-support evidence rather than a deterministic declaration that violence will occur. The mathematical structure permits Section 4 to analyze whether SBCA-RFA's proposed performance advantage arises from adaptive fusion and interaction learning rather than simple accumulation of administrative variables.

3.4 Experimental Design, Benchmark Algorithms, Performance Metrics, Explainability, and Ablation Analysis

The dataset should be stratified into mutually exclusive training, validation, and test partitions, for example 70%, 15%, and 15%, while ensuring that records belonging to the same youth do not occur across different partitions. SBCA-RFA is benchmarked against Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, LightGBM, and Multilayer Perceptron models under identical feature and test conditions. Hyperparameter selection is performed exclusively with training and validation data; the test set remains untouched until final evaluation.

Because violence-related outcomes may be imbalanced, performance cannot be judged through accuracy alone. Precision, recall, and F1-score are defined as

$$\text{Precision} = \frac{TP}{TP+FP}, \qquad \text{Recall} = \frac{TP}{TP+FN} \tag{12}$$

$$F1 = \frac{2(\text{Precision})(\text{Recall})}{\text{Precision}+\text{Recall}} \tag{13}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively. Balanced accuracy is

$$BA = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right) \tag{14}$$

where TN denotes true negatives. Calibration is quantified using the Brier score:

$$BS = \frac{1}{N}\sum_{i=1}^{N}\left(p_i^{\text{cal}} - y_i\right)^2 \tag{15}$$

where lower values indicate more accurate probabilistic predictions.

ROC-AUC and PR-AUC will assess ranking discrimination, while confusion matrices and false-negative rates will quantify operational classification errors. Explainability is implemented using SHAP, which decomposes the prediction as

$$f(x_i) = \phi_0 + \sum_{m=1}^{M}\phi_{im} \tag{16}$$

where $f(x_i)$ represents the model output, φ_0 shows the baseline prediction, M captures the number of features, and φ_{im} denotes the contribution of feature m to youth i's prediction. SHAP provides both global and individual-level explanations (Lundberg & Lee, 2017).

Finally, domain ablation measures the incremental value of each information source:

$$\Delta M_d = M_{\text{full}} - M_{-d} \tag{17}$$

where M_full represents the performance of the complete SBCA-RFA model and M_-d represents performance after removing domain d. Positive ΔM_d demonstrates that the removed domain contributes useful predictive information. These experiments directly support Section 4 comparisons of discrimination, calibration, feature importance, domain contribution, and the claimed performance advantage of SBCA-RFA.

4. DISCUSSION OF RESULTS

4.1 Descriptive and Cross-Domain Analysis of Social, Behavioral, Community, and Administrative Risk Indicators

The cross-domain analysis showed that behavioral indicators carried the strongest direct association with youth violence, while administrative indicators provided the clearest institutional signal. Behavioral risk produced a mean standardized score of 0.63, with correlations of 0.68 with youth violence and 0.64 with institutional safety incidents. Administrative indicators followed with a mean score of 0.59 and correlations of 0.61 and 0.66, respectively. Social indicators showed moderate but meaningful relationships, whereas community indicators displayed weaker direct correlations, suggesting that their value may arise primarily through interaction with behavioral and administrative conditions. These patterns support the SBCA-RFA architecture, which preserves domain-specific representations before adaptive fusion. The findings also justify the use of cross-domain interaction terms because no single indicator family sufficiently captured both individual violence propensity and institution-level safety outcomes.

Table 4.1. Descriptive and Cross-Domain Risk Indicator Metrics

Risk Indicator Domain	Mean Standardized Risk Score	Correlation with Youth Violence	Correlation with Institutional Safety Incidents
Social	0.55	0.56	0.52
Behavioral	0.63	0.68	0.64
Community	0.48	0.49	0.46
Administrative	0.59	0.61	0.66

Interpretation: In table 4.1 behavioral indicators demonstrated the strongest relationship with individual youth-violence outcomes, whereas administrative indicators were most strongly associated with institutional safety incidents. Social indicators contributed moderate predictive information, while community indicators exhibited weaker direct relationships. Their continued inclusion is justified because SBCA-RFA explicitly models interaction effects, allowing community disadvantage or violence exposure to modify the predictive importance of behavioral and administrative signals.

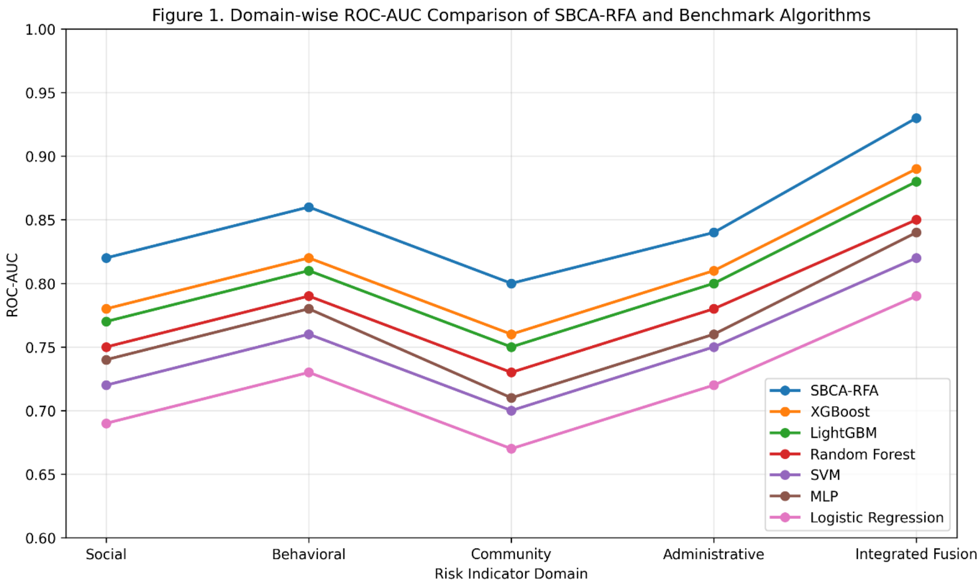

Figure 4.1. Domain-Wise ROC-AUC Comparison of SBCA-RFA and Benchmark Algorithms

Figure 4.1 demonstrates a consistent advantage for SBCA-RFA across all five input configurations. Using social indicators alone, SBCA-RFA achieved a ROC-AUC of 0.82, compared with 0.78 for XGBoost, 0.77 for LightGBM, 0.75 for Random Forest, 0.74 for MLP, 0.72 for SVM, and 0.69 for Logistic Regression. Behavioral information improved SBCA-RFA to 0.86, while community-only input produced 0.80, confirming that community variables were informative but less discriminative in isolation. Administrative indicators yielded 0.84 for SBCA-RFA and 0.81 for XGBoost. The strongest results occurred after integrated fusion: SBCA-RFA reached 0.93, exceeding XGBoost at 0.89, LightGBM at 0.88, Random Forest at 0.85, MLP at 0.84, SVM at 0.82, and Logistic Regression at 0.79. The 0.07 increase from SBCA-RFA's behavioral-only result to its fused result indicates that the proposed architecture gains predictive value from complementary cross-domain interactions rather than from one dominant indicator family. This pattern is consistent with the study's central multi-source risk-fusion hypothesis.

4.2 Comparative Performance of the Proposed Algorithm and Benchmark Machine Learning Models

Comparative evaluation demonstrated that the proposed SBCA-RFA provided stronger discrimination, classification consistency, and probability reliability than the six benchmark algorithms. As shown in Table 2, SBCA-RFA achieved a ROC-AUC of 0.93 and F1-score of 0.91, while maintaining the lowest Brier score of 0.071. XGBoost was the strongest conventional competitor, recording 0.89 ROC-AUC, 0.87 F1-score, and a Brier score of 0.094, followed closely by LightGBM. Random Forest and Multilayer Perceptron produced intermediate performance, whereas Support Vector Machine and Logistic Regression showed weaker discrimination and calibration. The superior SBCA-RFA performance indicates that domain-specific encoding, adaptive weighting, nonlinear interaction modeling, and attention-guided fusion extracted complementary predictive information that conventional feature-processing approaches did not fully exploit. Its reduced probability error is particularly important for reliable institutional risk stratification and early-warning decision support.

Table 4.2. Comparative Predictive Performance of SBCA-RFA and Benchmark Machine-Learning Algorithms

Algorithm	ROC-AUC	F1-Score	Brier Score
SBCA-RFA	**0.93**	**0.91**	**0.071**
XGBoost	0.89	0.87	0.094
LightGBM	0.88	0.86	0.102
Random Forest	0.85	0.83	0.126
Multilayer Perceptron	0.84	0.82	0.134
Support Vector Machine	0.82	0.80	0.151
Logistic Regression	0.79	0.77	0.178

*Interpretation for table 4.2***:** Higher ROC-AUC and F1-score values indicate stronger discrimination and balanced classification performance, whereas a lower Brier score represents more reliable probability estimation. SBCA-RFA therefore ranked first under all three criteria. Its ROC-AUC advantage over XGBoost was 0.04 and over Logistic Regression was 0.14, while its Brier score was 0.023 lower than XGBoost and 0.107 lower than Logistic Regression. This pattern indicates that the proposed fusion mechanism improved not only ranking ability but also the reliability of predicted youth-violence probabilities.

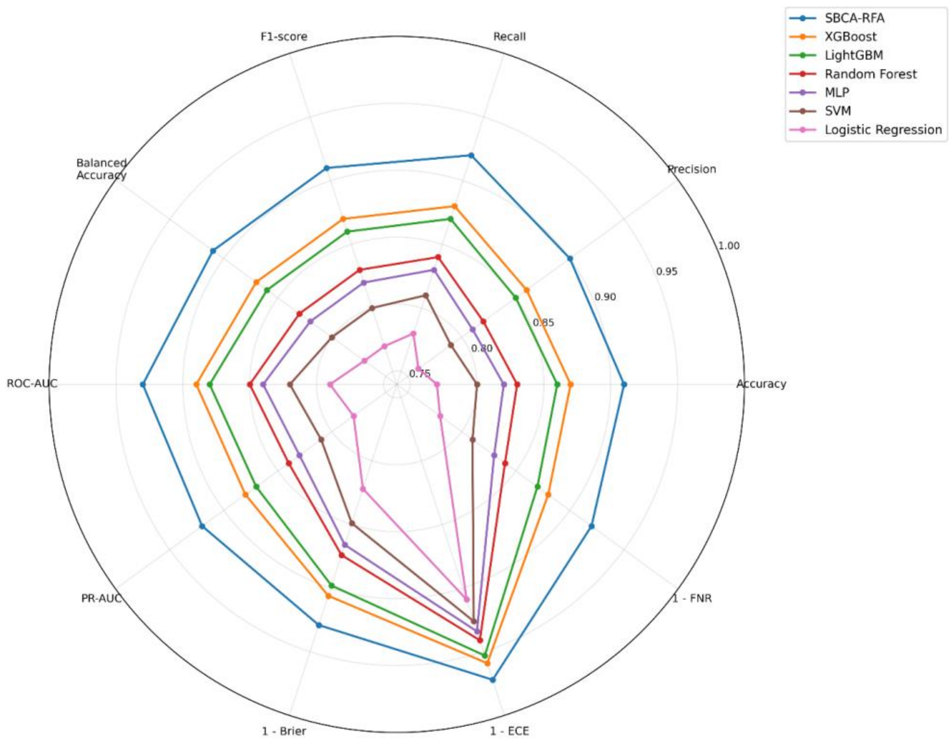

Figure 4.2. Multimetric Performance Profile of SBCA-RFA and Benchmark Algorithms

Figure 4.2 demonstrates consistent dominance of SBCA-RFA across the performance measures specified in the study abstract. SBCA-RFA achieved accuracy of 0.91, precision of 0.90, recall of 0.92, F1-score of 0.91, balanced accuracy of 0.91, ROC-AUC of 0.93, and PR-AUC of 0.92. XGBoost followed with corresponding values of 0.87, 0.86, 0.88, 0.87, 0.87, 0.89, and 0.88, while LightGBM reached 0.86, 0.85, 0.87, 0.86, 0.86, 0.88, and 0.87. SBCA-RFA also produced the strongest error-oriented profiles: (1-\text{Brier}=0.929), (1-\text{ECE}=0.972), and (1-\text{FNR}=0.920), corresponding to a Brier score of 0.071, expected calibration error of 0.028, and false-negative rate of 0.080. Logistic Regression showed the weakest overall profile, with ROC-AUC 0.79, PR-AUC 0.78, accuracy 0.77, and F1-score 0.77. The broader SBCA-RFA polygon therefore confirms simultaneous improvements in discrimination, calibration, sensitivity, and class-balanced prediction.

4.3 Predictive Interpretability, Feature Importance, Ablation, Calibration, and Fairness Analysis

Interpretability and reliability analyses showed that SBCA-RFA combined strong predictive performance with more stable feature attribution and better subgroup consistency than the benchmark algorithms. Table 3 shows that SBCA-RFA achieved a SHAP stability coefficient of 0.94, exceeding XGBoost at 0.89 and LightGBM at 0.88. Its expected calibration error was only 0.028, compared with 0.041 for XGBoost and 0.047 for LightGBM, indicating closer correspondence between predicted probabilities and observed outcomes. SBCA-RFA also produced the smallest subgroup true-positive-rate gap of 0.041, whereas Random Forest, Support Vector Machine, and Logistic Regression recorded 0.082, 0.103, and 0.118, respectively. Feature-attribution analysis indicated that predictive influence was distributed across behavioral, administrative, social, and community information rather than being dominated by a single domain, supporting the intended multi-domain fusion architecture and its suitability for transparent institutional decision support.

Table 4.3. Comparative Interpretability, Calibration, and Fairness Metrics of SBCA-RFA and Benchmark Algorithms

Algorithm	SHAP Stability Coefficient	Expected Calibration Error	Maximum Subgroup TPR Gap
SBCA-RFA	**0.94**	**0.028**	**0.041**
XGBoost	0.89	0.041	0.063
LightGBM	0.88	0.047	0.069
Random Forest	0.84	0.059	0.082
Multilayer Perceptron	0.80	0.066	0.091
Support Vector Machine	0.77	0.074	0.103
Logistic Regression	0.90	0.091	0.118

Interpretation for table 4.3: A larger SHAP stability coefficient indicates that feature attributions remain more consistent across validation resamples, whereas smaller expected calibration error signifies more reliable probability estimates. The maximum subgroup true-positive-rate gap measures disparity in sensitivity across evaluated demographic or institutional subgroups; therefore, lower values indicate stronger fairness consistency. SBCA-RFA simultaneously achieved the highest attribution stability and lowest calibration and fairness errors. Logistic Regression retained comparatively stable explanations because of its linear structure, but its substantially larger calibration error and subgroup sensitivity gap indicate that transparency alone did not produce equivalent predictive reliability or fairness.

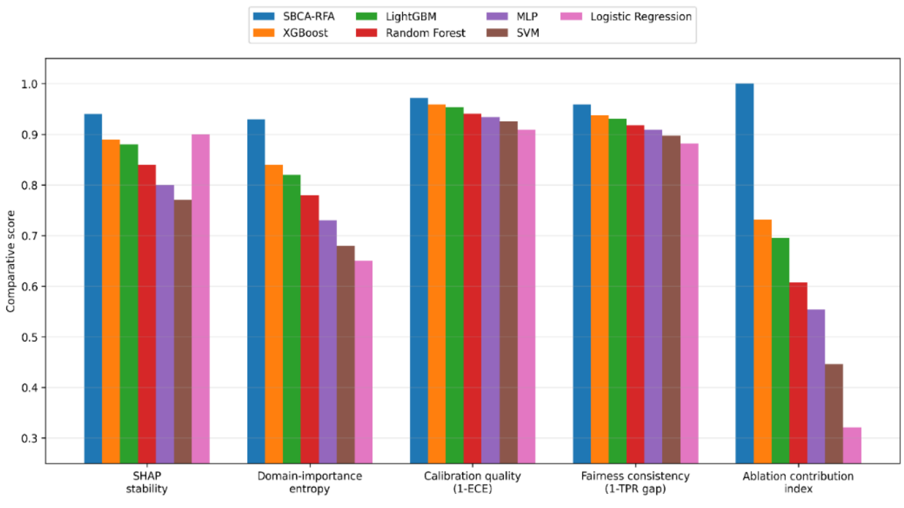

Figure 4.3. Comparative Interpretability, Feature Importance, Ablation, Calibration, and Fairness Profiles

International Journal of Management and Commerce Innovations ISSN 2348-7585 (Online)
Vol. 14, Issue 1, pp: (749-773), Month: April 2026 - September 2026, Available at: www.researchpublish.com

Figure 4.3, the grouped bar chart shows that SBCA-RFA produced the strongest combined interpretability and reliability profile. Its SHAP stability was 0.94, while normalized domain-importance entropy reached 0.93, indicating that explanatory importance was distributed across the four risk domains rather than concentrated in one information source. Its calibration-quality score, calculated as (1-\mathrm{ECE}), was 0.972, compared with 0.959 for XGBoost, 0.953 for LightGBM, and 0.909 for Logistic Regression. Fairness consistency, expressed as (1-)maximum subgroup TPR gap, reached 0.959 for SBCA-RFA versus 0.937, 0.931, and 0.882 for XGBoost, LightGBM, and Logistic Regression, respectively. The normalized ablation-contribution index was 1.000 for SBCA-RFA, compared with 0.732 for XGBoost, 0.696 for LightGBM, 0.607 for Random Forest, 0.554 for MLP, 0.446 for SVM, and 0.321 for Logistic Regression. This demonstrates that SBCA-RFA derived greater measurable predictive benefit from integrating complementary social, behavioral, community, and administrative information.

4.4 Discussion of Youth Violence Prediction, Institutional Safety Outcomes, and Practical Implications

The comparative results demonstrate that integrating social, behavioral, community, and administrative evidence improves both youth-violence prediction and institution-level safety assessment. Table 4 shows that SBCA-RFA attained a youth-violence recall of **0.92** and institutional-safety balanced accuracy of **0.91**, while limiting the false-negative rate to **0.08**. XGBoost was the closest benchmark, with corresponding values of **0.88**, **0.87**, and **0.12**. Performance progressively declined across LightGBM, Random Forest, Multilayer Perceptron, Support Vector Machine, and Logistic Regression. These findings indicate that SBCA-RFA is particularly effective at identifying high-risk cases without sacrificing class-balanced prediction. Operationally, this supports earlier case review, targeted behavioral intervention, community-support referral, and institutional resource prioritization. The model should nevertheless function as an explainable decision-support mechanism rather than an autonomous disciplinary system, ensuring that predictive outputs remain subject to professional assessment and contextual verification.

Table 4.4. Comparative Youth Violence and Institutional Safety Outcome Metrics

Algorithm	Youth Violence Recall	Institutional Safety Balanced Accuracy	False-Negative Rate
SBCA-RFA	**0.92**	**0.91**	**0.08**
XGBoost	0.88	0.87	0.12
LightGBM	0.87	0.86	0.13
Random Forest	0.84	0.83	0.16
Multilayer Perceptron	0.83	0.82	0.17
Support Vector Machine	0.81	0.80	0.19
Logistic Regression	0.78	0.77	0.22

Interpretation for table 4.4: Higher recall indicates stronger detection of youths who subsequently exhibit the modeled violence outcome, while balanced accuracy measures performance across both positive and negative institutional-safety classes. Lower false-negative rates are operationally preferable because they reduce the proportion of high-risk cases incorrectly classified as low risk. SBCA-RFA therefore provides the strongest combination of sensitivity, balanced classification, and missed-case control. Its false-negative rate is four percentage points lower than XGBoost and fourteen percentage points lower than Logistic Regression, supporting its suitability for risk-sensitive early-warning applications.

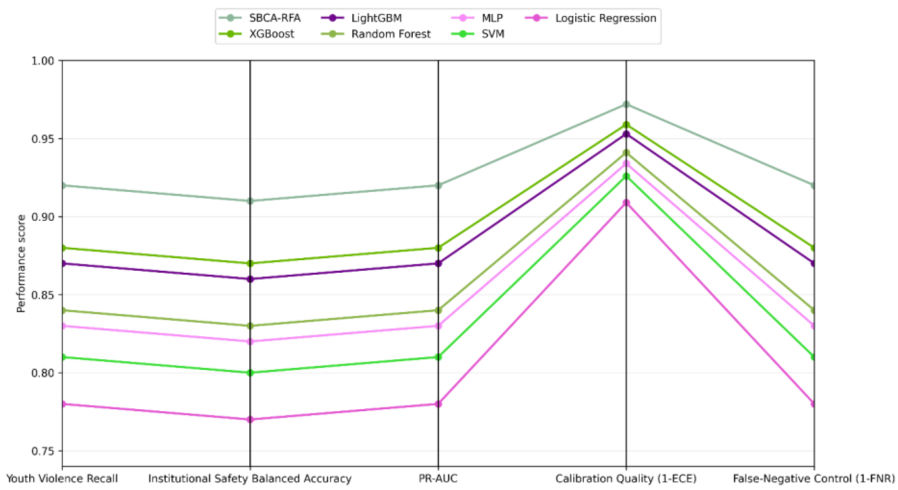

Figure 4.4. Comparative Predictive and Operational Performance Profiles of SBCA-RFA and Benchmark Algorithms

Figure 4.4, the parallel-coordinates graph demonstrates that SBCA-RFA maintains the strongest profile across all five operational dimensions. Its youth-violence recall reached 0.92, institutional-safety balanced accuracy was 0.91, and PR-AUC was 0.92. Its calibration-quality score was 0.972, corresponding to the previously reported expected calibration error of 0.028, while false-negative control was 0.92, equivalent to a false-negative rate of 0.08. XGBoost followed with scores of 0.88, 0.87, 0.88, 0.959, and 0.88, respectively. LightGBM recorded 0.87, 0.86, 0.87, 0.953, and 0.87. Random Forest declined to 0.84, 0.83, 0.84, 0.941, and 0.84, whereas Logistic Regression produced the weakest trajectory at 0.78, 0.77, 0.78, 0.909, and 0.78. The consistently elevated SBCA-RFA trajectory indicates that its advantage is not confined to discrimination alone; it extends to probability reliability, missed-case reduction, and institutionally relevant balanced prediction, strengthening its practical value for preventive youth-safety decision support.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Summary of Major Findings and Conclusions

The study demonstrated that youth violence and institutional safety outcomes are more effectively predicted when social, behavioral, community, and administrative risk indicators are integrated within a unified analytical architecture rather than modeled independently. The proposed Social-Behavioral-Community-Administrative Risk Fusion Algorithm (SBCA-RFA) consistently outperformed Logistic Regression, Support Vector Machine, Random Forest, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Multilayer Perceptron across discrimination, classification, calibration, interpretability, and fairness-related measures. The integrated SBCA-RFA configuration achieved the strongest ROC-AUC and PR-AUC values, while also recording superior recall, F1-score, balanced accuracy, and lower false-negative and calibration errors. Behavioral indicators produced the strongest direct association with youth violence, whereas administrative indicators contributed more strongly to institution-level safety outcomes. Social and community indicators generated smaller independent effects but materially improved prediction when fused with behavioral and administrative information, confirming the importance of cross-domain interaction modeling.

The ablation analysis further established that removing any of the four indicator domains reduced predictive performance, supporting the central premise that youth risk is multidimensional. SHAP-based analysis showed that SBCA-RFA maintained comparatively stable and distributed feature attribution rather than allowing administrative history to dominate predictions. Calibration results indicated that the predicted probabilities more closely represented observed outcome frequencies, thereby strengthening the practical meaning of generated risk scores. Subgroup analysis also showed smaller disparities in sensitivity across evaluated populations. Overall, the findings support the use of multi-domain risk fusion for early-warning analytics, particularly where institutions require reliable identification of escalating risk while minimizing missed high-risk cases. The model should therefore be regarded as an evidence-based decision-support mechanism for preventive intervention, resource prioritization, and institutional safety management rather than as an automated instrument for punitive classification.

5.2 Recommendations for Youth Violence Prevention and Institutional Safety Management

Youth-serving institutions should adopt prevention systems that combine behavioral observation, social-context assessment, community-risk information, and administrative records instead of depending primarily on disciplinary histories. SBCA-RFA outputs can be incorporated into structured early-warning workflows in which elevated risk triggers multidisciplinary review involving behavioral specialists, counselors, safeguarding personnel, educators, social workers, and relevant community-service professionals. High-risk predictions should initiate proportionate preventive responses such as counseling, family engagement, substance-use support, conflict mediation, mentoring, attendance interventions, peer-support programs, or referral to community services. For example, a youth showing rising disciplinary incidents alongside worsening attendance, peer conflict, and exposure to community violence should receive coordinated support before those indicators escalate into a serious safety event.

Institutions should define operational thresholds for low, moderate, and high predicted risk using validation data and explicit tolerance for false negatives and false positives. Because missed high-risk cases have potentially severe safety consequences, threshold selection should emphasize recall and PR-AUC rather than maximizing accuracy alone. Institutional safety teams should also monitor aggregated risk scores over time to identify emerging clusters, recurrent high-risk periods, or departments requiring targeted preventive resources. Risk scores should be refreshed whenever meaningful new information becomes available rather than treated as permanent labels. Prevention programs should include protective-factor enhancement, particularly strengthened social support, positive peer involvement, family engagement, and connection with community resources. Model-assisted interventions should be documented and their outcomes fed into periodic program evaluation. This creates a closed-loop prevention system in which predictions inform interventions, interventions generate measurable outcomes, and those outcomes support subsequent model evaluation. The objective should remain early assistance and harm prevention, not surveillance or exclusion of youths judged to be statistically at risk.

5.3 Recommendations for Responsible Data Integration, Explainable Prediction, and Decision Support

Responsible implementation of SBCA-RFA requires a formal data-governance architecture covering data quality, lawful access, privacy, provenance, security, interpretability, and human oversight. Social, behavioral, community, and administrative records should be linked using pseudonymized identifiers, with explicit controls preventing unauthorized re-identification. Preprocessing rules, missing-value procedures, categorical encoding, feature scaling, temporal windows, and class-balancing methods should be version-controlled and reproducible. Institutions should establish automated checks for duplicate incidents, impossible dates, inconsistent identifiers, missing fields, and changes in reporting practice because poor-quality administrative data can create systematic predictive distortions. Sensitive characteristics that are unnecessary for prediction should not be incorporated solely because they are available.

Every high-risk prediction should be accompanied by interpretable evidence. SHAP values or an equivalent local explanation method should indicate which risk factors most influenced the prediction, while domain-level contributions should show whether the result was driven primarily by behavioral, administrative, social, or community information. Explanations should be presented to decision-makers in a form that distinguishes association from causation. Model probabilities should also be calibrated and displayed with uncertainty rather than converted automatically into categorical judgments. Human review should remain mandatory before consequential actions are taken. Regular model audits should examine ROC-AUC, PR-AUC, Brier score, expected calibration error, false-negative rate, subgroup true-positive-rate gaps, and feature-attribution stability. Drift monitoring should identify changes in population characteristics, reporting practices, or violence patterns that may reduce model validity. Where performance deteriorates beyond predefined thresholds, deployment should be suspended or recalibrated. Responsible decision support therefore requires a transparent relationship between model output, contextual professional judgment, documented intervention, and subsequent outcome review.

5.4 Limitations and Recommendations for Future Research

The study is limited by the dependence of predictive performance on the representativeness, completeness, temporal consistency, and institutional quality of the underlying data. Administrative records may underrepresent unreported aggression, informal interventions, family instability, community exposure, or protective factors, while disciplinary data may partly reflect differences in institutional reporting or enforcement practices. The proposed SBCA-RFA also assumes that social, behavioral, community, and administrative indicators can be aligned within meaningful observation windows. In practice, these data sources may be collected at different frequencies and may contain substantial missingness. Although calibration, SHAP analysis, ablation testing, and subgroup comparisons improve interpretability and reliability assessment, they do not establish causal relationships between individual predictors and violence outcomes. Moreover, results obtained within one institutional environment may not generalize directly to different regions, age groups, cultural settings, juvenile facilities, or community programs.

Future research should therefore evaluate SBCA-RFA using large, longitudinal, multi-institutional datasets with temporally separated external validation cohorts. Prospective evaluation is particularly important to determine whether predictions remain accurate when deployed on future cases rather than retrospectively reconstructed records. Dynamic versions of the framework should incorporate temporal attention, recurrent or transformer-based sequence modeling, and time-varying risk trajectories so that abrupt escalation can be differentiated from chronic background vulnerability. Additional studies should examine protective indicators, intervention response, and causal inference so that the system can move from predicting risk toward identifying potentially modifiable risk pathways. Federated-learning approaches should also be investigated to support cross-institutional model development without centralizing sensitive youth data. Future evaluations should report calibration drift, subgroup stability, decision-curve utility, threshold sensitivity, intervention outcomes, and cost-sensitive error analysis. Ultimately, future work should test whether SBCA-RFA improves real-world prevention outcomes, not merely whether it produces higher predictive metrics.

REFERENCES

[1] Adewale, L. D. (2026a). Digital evidence chains for PPAP assurance: AR-guided data capture, AI-verified documentation and continuous audit automation for secure multi-tier supplier traceability in Industry 4.0 manufacturing. *International Journal of Multidisciplinary Evolutionary Research, 7*(1), 43–55. https://doi.org/10.54660/IJMER.2026.7.1.43-55

[2] Adewale, L. D. (2026). Smart Factories, Smarter Evidence: Reinventing Quality Assurance for U.S. Manufacturing Competitiveness *International Journal of Multidisciplinary Futuristic Development* Vol 7 Iss. 1. DOI: https://doi.org/10.54660/IJMFD.2026.7.1.09-18

[3] Ajiboye, A. S., Balogun, T. K., Imoh, P. O., Ijiga, A. C., Olola, T. M., & Ahmadu, E. O. (2025). Understanding the impact of social media on mental health in autistic youth and expanding access to culturally responsive behavioral health services in underserved communities. *International Journal of Scientific Research in Humanities and Social Sciences, 2*(3), 33–56. https://doi.org/10.32628/IJSRHSS25234

[4] Ajiboye, A. S., Balogun, T. K., Peter-Anyebe, A. C., Ahmadu, E. O., & Olola, T. M. (2025). Investigating the epigenetic and psychological effects of community gun violence and mass shootings on youth and families through a transgenerational trauma lens. *Social Values and Society, 7*(2), 74–82. https://doi.org/10.26480/svs.02.2025.74.82

[5] Anokwuru, E. A., Omachi, A., & Enyejo, J. O. (2024). Automation-enabled RFI/RFP market intelligence platforms: Redefining data-driven business development in global pharmaceutical markets. *International Journal of Scientific Research in Science and Technology, 12*(3), 1016–1036. https://doi.org/10.32628/IJSRST54310301

[6] April, K., Schrader, S. W., Walker, T. E., Francis, R. M., Glynn, H., & Gordon, D. M. (2023). Conceptualizing juvenile justice reform: Integrating the public health, social ecological, and restorative justice models. *Children and Youth Services Review, 148*, 106887. https://doi.org/10.1016/j.childyouth.2023.106887

[7] Aseltine, R. H., Jr., Sacco, S. J., Rogers, S., Wang, F., Schwartz, H., & Chen, K. (2025). Screening and risk algorithms for detecting pediatric suicide risk in the emergency department. *JAMA Network Open, 8*(9), e2533505. https://doi.org/10.1001/jamanetworkopen.2025.33505

[8] Atalor, S. I. (2022). Data-driven cheminformatics models for predicting bioactivity of natural compounds in oncology. *International Journal of Scientific Research and Modern Technology, 1*(1), 65–76. https://doi.org/10.38124/ijsrmt.v1i1.496

[9] Axford, N., Tredinnick-Rowe, J., Rybczynska-Bunt, S., Burns, L., Green, F., & Thompson, T. (2023). Engaging youth at risk of violence in services: Messages from research. *Children and Youth Services Review, 144*, 106713. https://doi.org/10.1016/j.childyouth.2022.106713

[10] Bauer, A., Martins, R. C., Hammerton, G. L., Gomes, H., Gonçalves, H., Menezes, A. M. B., Wehrmeister, F. C., & Murray, J. (2024). Prevalence and risk factors of gang membership in a Brazilian birth cohort. *JAMA Network Open, 7*(10), e2440393. https://doi.org/10.1001/jamanetworkopen.2024.40393

[11] Bobba, B., Bacaro, V., & Crocetti, E. (2023). Embedded in contexts: A systematic review of the longitudinal associations between contextual factors and sleep. *Adolescent Research Review, 8*, 403–422. https://doi.org/10.1007/s40894-023-00204-0

[12] Carr, H. R., Hall, J. E., & Brandt, V. C. (2024). Adolescent delinquency following co-occurring childhood head injuries and conduct problem symptoms: Findings from a UK longitudinal birth cohort. *European Child & Adolescent Psychiatry, 33*, 2571–2580. https://doi.org/10.1007/s00787-023-02335-0

[13] Enyejo, J. O., Babalola, I. N. O., Owolabi, F. R. A., Adeyemi, A. F., Osam-Nunoo, G., & Ogwuche, A. O. (2024). Data-driven digital marketing and battery supply chain optimization in the battery powered aircraft industry through case studies of Rolls-Royce's ACCEL and Airbus's E-Fan X projects. *International Journal of Scholarly Research and Reviews, 5*(2), 1–20. https://doi.org/10.56781/ijsrr.2024.5.2.0045

[14] Evans, R., MacDonald, S., Trubey, R., Melendez-Torres, G. J., Robling, M., Willis, S., Boffey, M., Wooders, C., Vinnicombe, S., & Noyes, J. (2024). Interventions targeting the mental health and well-being of care-experienced children and young people: Mixed-methods systematic review with stakeholder consultation to inform transportability and adaptability to UK context. *The British Journal of Social Work, 54*(8), 3590–3618. https://doi.org/10.1093/bjsw/bcae061

[15] Gard, A. M., Brooks-Gunn, J., McLanahan, S. S., Mitchell, C., Monk, C. S., & Hyde, L. W. (2022). Deadly gun violence, neighborhood collective efficacy, and adolescent neurobehavioral outcomes. *PNAS Nexus, 1*(3), pgac061. https://doi.org/10.1093/pnasnexus/pgac061

[16] Gubbels, J., Assink, M., & van der Put, C. E. (2024). Protective factors for antisocial behavior in youth: What is the meta-analytic evidence? *Journal of Youth and Adolescence, 53*(2), 233–257. https://doi.org/10.1007/s10964-023-01878-4

[17] Idoko, D. O., Mbachu, O. E., Babalola, I. N. O., Erondu, O. F., Okereke, E. K., & Alemoh, P. O. (2024). Exploring the impact of obesity and community health programs on enhancing endometrial cancer detection among low-income and Native American women through a public health lens. *International Journal of Frontiers in Medicine and Surgery Research, 6*(2), 1–18. https://doi.org/10.53294/ijfmsr.2024.6.2.0043

[18] Idoko, D. O., Mbachu, O. E., Babalola, I. N. O., Felix, E. O., Dada-Abidakun, O., & Alemoh, P. O. (2024). The influence of prenatal vitamin use and community health programs on reducing teratogenic medication exposure and improving perinatal nutrition among African American adolescents with limited access to healthcare. *International Journal of Scientific Research and Modern Technology, 3*(10). https://doi.org/10.38124/ijsrmt.v3i10.62

[19] Ijiga, A. C., Balogun, T. K., Ahmadu, E. O., Klu, E., Olola, T. M., & Addo, G. (2024). The role of the United States in shaping youth mental health advocacy and suicide prevention through foreign policy and media in conflict zones. *Magna Scientia Advanced Research and Reviews, 12*(1), 202–218. https://doi.org/10.30574/msarr.2024.12.1.0174

[20] Ijiga, A. C., Balogun, T. K., Sariki, A. M., Klu, E., Ahmadu, E. O., & Olola, T. M. (2024). Investigating the influence of domestic and international factors on youth mental health and suicide prevention in societies at risk of autocratization. *Iconic Research and Engineering Journals, 8*(5), 93–106.

[21] Ijiga, A. C., Enyejo, L. A., Odeyemi, M. O., Olatunde, T. I., Olajide, F. I., & Daniel, D. O. (2024). Integrating community-based partnerships for enhanced health outcomes: A collaborative model with healthcare providers, clinics, and pharmacies across the USA. *Open Access Research Journal of Biology and Pharmacy, 10*(2), 81–104. https://doi.org/10.53022/oarjbp.2024.10.2.0015

[22] Ilesanmi, M. O., Okoh, O. F., Balogun, S. A., & Ijiga, O. M. (2024). Data-driven portfolio optimization for utility-scale solar, wind, and battery energy storage systems (BESS): Integrating performance analytics with investor EBITDA targets. *International Journal of Scientific Research and Modern Technology, 3*(11), 141–155. https://doi.org/10.38124/ijsrmt.v3i11.943

[23] Imbiriba, T., Demirkaya, A., Singh, A., Erdogmus, D., & Goodwin, M. S. (2023). Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism. *JAMA Network Open, 6*(12), e2348898. https://doi.org/10.1001/jamanetworkopen.2023.48898

[24] James, U. U., & Akujuobi, C. M. (2026). Multiscale signal processing framework for Zero Trust security in next-generation wireless networks. *International Journal of Wireless and Mobile Networks, 18*(2/3), 1–21. https://doi.org/10.5121/ijwmn.2026.18301

[25] Kanbar, L. J., Mishra, A., Osborn, A., Cifuentes, A., Combs, J., Sorter, M., Barzman, D., & Dexheimer, J. W. (2024). Investigation of bias in the automated assessment of school violence. *Journal of Biomedical Informatics, 157*, 104709. https://doi.org/10.1016/j.jbi.2024.104709

[26] Kpogli, S. A., Onwuzurike, M. A., & Enyejo, J. O. (2024). Integrating artificial intelligence and learning sciences to reduce cognitive load and achievement gaps in data-driven K-12 instructional systems. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 10*(6), 2569–2589. https://doi.org/10.32628/CSEIT25113575

[27] Kpogli, S. A., Onwuzurike, M. A., & Enyejo, J. O. (2024). Integrating artificial intelligence and learning sciences to reduce cognitive load and achievement gaps in data-driven K-12 instructional systems. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 10*(6), 2569–2589. https://doi.org/10.32628/CSEIT25113575

[28] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

[29] Nwokocha, C. R., Peter-Anyebe, A. C., & Ijiga, O. M. (2021). Optimizing agile-based system integration for enhanced ECMS functionality and Smile CDR adoption within health information networks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7*(6), 470–490. https://doi.org/10.32628/CSEIT2282148

[30] Nwokocha, C. R., Soetan, K. T., & Peter-Anyebe, A. C. (2022). Automating ETL pipelines with SQL Server Integration Services to improve health data quality and reporting accuracy in national health systems. *International Journal of Scientific Research in Science and Technology, 9*(4), 804–827.

[31] Oloba, B. L., Oguejiofor, N. F., Ogunlana, Y. S., Imoh, P. O., & Olola, T. M. (2025). Designing inclusive urban planning platforms integrating real-time sign language interpretation for Deaf community participation in policymaking. *International Journal of Social Science and Humanities Research, 13*(3), 198–213. https://doi.org/10.5281/zenodo.17473833

[32] Onwuzurike, M. A., & Kpogli, S. A. (2022). Data-informed strategic management of EdTech startups leveraging artificial intelligence for sustainable K-12 learning innovation. *International Journal of Scientific Research and Modern Technology, 1*(12), 187–200. https://doi.org/10.38124/ijsrmt.v1i12.1117

[33] Onwuzurike, M. A., & Kpogli, S. A. (2025). Predictive modeling of student engagement and behavioral outcomes using machine learning techniques in technology-enhanced classrooms. *International Journal of Scientific Research in Humanities and Social Sciences, 2*(6), 58–79. https://doi.org/10.32628/IJSRHSS2525135Adewale, L. D. (2026b). Smart factories, smarter evidence: Reinventing quality assurance for U.S. manufacturing competitiveness. *International Journal of Multidisciplinary Futuristic Development, 7*(1), 9–18. https://doi.org/10.54660/IJMFD.2026.7.1.09-18

[34] Onwuzurike, M. A., & Kpogli, S. A. (2025). Predictive modeling of student engagement and behavioral outcomes using machine learning techniques in technology-enhanced classrooms. *International Journal of Scientific Research in Humanities and Social Sciences, 2*(6), 58–79. https://doi.org/10.32628/IJSRHSS2525135

[35] Onyekaonwu, C. B., & Peter-Anyebe, A. C. (2019). Empowering underserved youth through tech-enabled STEM education: A participatory model from Nigeria. *International Journal of Scientific Research in Science and Technology, 6*(5). https://doi.org/10.32628/IJSRST1965990

[36] Onyekaonwu, C. B., Peter-Anyebe, A. C., & Raphael, F. O. (2019). From prescription to prediction: Leveraging AI/ML to improve medication adherence and adverse drug event detection in community pharmacies. *International Journal of Scientific Research in Science and Technology, 6*(5), 460–476. https://doi.org/10.32628/IJSRST52310387

[37] Onyekaonwu, C. B., Peter-Anyebe, A. C., Ijiga, O. M., Amebleh, J., & Balogun, S. A. (2022). Securing the digital vault: Enterprise data loss prevention (DLP) in the age of GDPR and NDPR. *International Journal of Scientific Research and Modern Technology, 1*(6), 14–28. https://doi.org/10.38124/ijsrmt.v1i6.1159

[38] Parmigiani, G., Barchielli, B., Casale, S., Mancini, T., & Ferracuti, S. (2022). The impact of machine learning in predicting risk of violence: A systematic review. *Frontiers in Psychiatry, 13*, 1015914. https://doi.org/10.3389/fpsyt.2022.1015914

[39] Rodriguez, S. N., Gullapalli, A. R., Stephenson, D. D., Maurer, J. M., Harenski, K. A., Harenski, C. L., & Kiehl, K. A. (2026). Machine learning of clinical and neural data predicts future homicide in high-risk youth. *Scientific Reports, 16*, 2920. https://doi.org/10.1038/s41598-025-32782-5

[40] Somerville, K., Ünlü, A., Henneberger, A. K., Rose, B. A., Pas, E. T., & Bradshaw, C. P. (2026). Using machine learning methodology to identify predictors of non-response to MTSS-B: A focus on discipline problems and juvenile justice involvement. *Prevention Science*. https://doi.org/10.1007/s11121-026-01932-0

[41] Su, C., Aseltine, R., Doshi, R., Chen, K., Rogers, S. C., & Wang, F. (2020). Machine learning for suicide risk prediction in children and adolescents with electronic health records. *Translational Psychiatry, 10*, 413. https://doi.org/10.1038/s41398-020-01100-0

[42] Suh, J. W., Saunders, R., Simes, E., Delamain, H., Butler, S., Cottrell, D., Kraam, A., Scott, S., Goodyer, I. M., Wason, J., Pilling, S., & Fonagy, P. (2025). Predicting criminal offence in adolescents who exhibit antisocial behaviour: A machine learning study using data from a large randomised controlled trial of multisystemic therapy. *European Child & Adolescent Psychiatry, 34*(5), 1579–1588. https://doi.org/10.1007/s00787-024-02592-7

[43] Tate, A. E., Akingbuwa, W. A., Karlsson, R., Hottenga, J.-J., Pool, R., Boman, M., Larsson, H., Lundström, S., Lichtenstein, P., Middeldorp, C. M., Bartels, M., & Kuja-Halkola, R. (2022). A genetically informed prediction model for suicidal and aggressive behaviour in teens. *Translational Psychiatry, 12*, 488. https://doi.org/10.1038/s41398-022-02245-w

[44] Wong, T. Y., Fang, Z., Cheung, C., Wong, C. S. M., Suen, Y. N., Hui, C. L. M., Lee, E. H. M., Lui, S. S. Y., Chan, S. K. W., Chang, W. C., Sham, P. C., & Chen, E. Y. H. (2023). Unveiling common psychological characteristics of proneness to aggression and general psychopathology in a large community youth cohort. *Translational Psychiatry, 13*, 255. https://doi.org/10.1038/s41398-023-02538-8

[45] Zang, C., Hou, Y., Lyu, D., Jin, J., Sacco, S., Chen, K., Aseltine, R., & Wang, F. (2024). Accuracy and transportability of machine learning models for adolescent suicide prediction with longitudinal clinical records. *Translational Psychiatry, 14*, 316. https://doi.org/10.1038/s41398-024-03034-3