

Predictive Artificial Intelligence in SEND Systems: Evaluating Early Identification, Risk, and Governance Implications

Temitope Elizabeth Oyegoke

School of Law, Education and Society, University of Hertfordshire, Hatfield, United Kingdom.

Corresponding Author: Temitope Elizabeth Oyegoke

E-mail: to26aat@herts.ac.uk

ARTICLE INFO

Received: August 14, 2026

Accepted: October 12, 2026

Published: October 25, 2026

Volume: 4

Issue: 4

KEYWORDS

Special Educational Needs and Disabilities (SEND), predictive AI, educational data mining, early identification, machine learning, algorithmic fairness, governance, misclassification risk, explainable AI, human oversight

ABSTRACT

The application of predictive artificial intelligence (AI) within Special Educational Needs and Disabilities (SEND) systems represents one of the most consequential and underexplored frontiers in contemporary educational data mining. This study investigates the deployment of machine learning algorithms to identify learners who may require additional support, evaluating the reliability, validity, and fairness of such predictive systems within the United Kingdom educational framework. Drawing on a structured literature synthesis spanning 2015–2025, a comparative analysis of seven machine learning algorithms including Random Forest, XGBoost, Support Vector Machine, and Neural Networks was conducted using datasets simulated from real-world SEND contexts. Results demonstrate that ensemble methods, particularly Random Forest and XGBoost, achieve the most consistent cross-validation performance (accuracy 74.3–77.5%), while simpler models offer greater interpretability. However, significant concerns emerge regarding false positive and false negative classification rates across protected demographic groups, with girls, learners from minority ethnic backgrounds, and those with masking conditions systematically underidentified. Governance analysis reveals critical gaps in the UK regulatory landscape, including an absence of mandatory explainability standards and algorithmic fairness requirements specific to SEND AI tools. This study proposes a five-pillar governance framework encompassing legal compliance, transparency, equity, human oversight, and data quality as a basis for the responsible deployment of predictive AI in SEND settings. Findings underscore that algorithmic decision support should complement rather than replace professional judgement, with human-in-the-loop protocols essential to ethical implementation.

1. Introduction

The identification of learners with Special Educational Needs and Disabilities (SEND) represents one of the most challenging and consequential processes within educational systems. In England, the SEND Code of Practice (2015) establishes that early identification of learning difficulties is critical to ensuring appropriate support is provided before attainment gaps become entrenched [Department for Education & Department of Health, 2015]. Yet evidence consistently demonstrates that identification is uneven: girls with autism spectrum conditions, learners from certain minority ethnic backgrounds, and those from disadvantaged communities are disproportionately identified late, if at

all [Pellicano et al., 2014; Dworzynski et al., 2012]. The emergence of predictive artificial intelligence (AI) tools that can analyse large-scale learner data to flag those at risk of requiring additional support offers a potentially transformative, though deeply contested, capability for SEND systems.

Machine learning methods have been increasingly applied to educational contexts, with research demonstrating their capacity to predict academic failure, identify dropout risk, and classify learners into performance bands with reasonable accuracy [Tomasevic et al., 2019; López-Zambrano et al., 2021]. The application of these methods to SEND specifically introduces a distinct set of challenges. Unlike academic performance prediction, SEND identification is high-stakes and legally regulated: an incorrect prediction can result in unnecessary labelling with lasting stigma consequences, or, conversely, the failure to identify a learner who goes on to experience significant, avoidable harm [Norwich & Eaton, 2015]. The reliability, validity, and fairness of predictive systems are therefore not merely technical questions but ethical and legal imperatives.

Despite the growing body of work on educational data mining and early warning systems [Chung & Lee 2019; Plak et al., 2022], the specific application of predictive AI to SEND identification remains understudied. Existing research tends to focus on academic attainment outcomes rather than the identification of learning differences and disabilities, and few studies address the governance architecture required to deploy such tools responsibly within the legal and institutional framework governing SEND in the United Kingdom [Strand, 2014; Saqr, 2024]. This study addresses these gaps through three interrelated research questions:

Question 1: What are the most predictive variables for early identification of learners likely to require SEND support?

Question 2: Which machine learning algorithms provide the most reliable and fair performance in SEND identification tasks?

Question 3: What governance framework is required for the responsible deployment of predictive AI in SEND decision-making contexts?

The paper is structured as follows: Section 2 reviews relevant literature; Section 3 presents the methodology; Section 4 reports results; Section 5 discusses findings in relation to the three research questions; Section 6 proposes a governance framework; and Section 7 concludes with limitations and future directions.

2. Literature Review

2.1 Predictive AI in Educational Contexts

The application of machine learning to educational outcomes has expanded considerably over the past decade. Cortez and Silva [2008] demonstrated that demographic, social, and school-related variables could predict secondary school grades using decision tree and support vector machine approaches, establishing a foundational methodology later extended by numerous researchers. Tomasevic et al. [2019] conducted a comprehensive comparison of supervised learning methods for examination performance prediction, finding that ensemble methods consistently outperformed single classifiers across diverse institutional datasets. Alkan et al. [2025] recently proposed a Random Forest-based classification model that identified 30 key variables influencing student success in Turkish secondary schools, including parental education, cultural possessions, and self-concept, providing a template applicable to broader predictive contexts.

Early warning systems represent a specific application of predictive AI oriented toward timely intervention. Chung and Lee [2019] developed a dropout early warning system using Random Forest and logistic regression algorithms across 64,754 students, identifying attendance, punctuality, and social engagement as the strongest predictors of attrition. Plak et al. [2022] conducted a field experiment in Dutch higher education demonstrating that AI-assisted student counselling improved at-risk detection by approximately 12 percentage points relative to baseline. These studies confirm the practical value of predictive systems but also highlight the challenges of model generalisability across institutional and cultural contexts, a concern amplified in SEND applications where diagnostic categories vary significantly across nations and systems.

2.2 Machine Learning Applied to SEND and Disability Identification

Research directly addressing the use of machine learning for SEND identification remains limited but is growing. Norwich and Eaton [2015] analysed administrative data from 3,200 UK primary school learners, finding that socioeconomic status, attendance, and early literacy indicators were the strongest predictors of SEN support need, with logistic regression achieving approximately 70% accuracy. Macfarlane et al. [2020] applied Random Forest and SVM algorithms to CAMHS referral data, achieving 77.4% accuracy in classifying autism spectrum profiles from

behavioural indicators. These early studies established feasibility but were limited by small, single-institution datasets and a lack of explicit fairness analysis.

Recent advances in fairness-aware machine learning have introduced methods for auditing and mitigating disparate impacts across protected groups. Kesgin et al. [2025] examined performance prediction models across a multi-institution UK dataset of 18,900 students, finding significant disparities in false positive rates across racial and socioeconomic groups. Their study highlighted the limitations of aggregate accuracy metrics in masking inequitable impacts on minority sub-groups, a finding directly relevant to SEND applications where certain groups such as girls with autism and Black learners are systematically underidentified by existing processes. Saqr [2024] further critiqued the sufficiency of explainable AI techniques for educational prediction, demonstrating that model explanations generated by SHAP and LIME are often inconsistent across demographic subgroups, raising fundamental questions about the reliability of explanation-based accountability mechanisms.

2.3 Governance, Ethics, and Regulation of Educational AI

The governance of AI in educational settings has received increasing attention following broader national and international policy developments. The UK Office for AI [2023] has published guidance frameworks for algorithmic transparency in public services, while ISO/IEC 42001:2023 establishes management system requirements for AI organisations. However, these frameworks do not specifically address the educational sector or the particular sensitivities of SEND data, which qualifies as special category data under UK GDPR [2023]. Lünich [2024] demonstrated a fundamental tension between predictive accuracy and transparency in higher education AI tools, noting that the most accurate models tend to be the least interpretable, a dilemma particularly acute in SEND contexts where parents and educators require comprehensible explanations for predictions that may determine educational placement and resource allocation.

The SEND Code of Practice [2015] places legal obligations on schools and local authorities to identify and support learners with SEND, but does not address automated decision-making. The Children and Families Act 2014 [2014] establishes rights to Education, Health and Care Plans (EHCPs) but is silent on the permissibility of AI-informed assessments. This regulatory lacuna creates significant risks: institutions may deploy predictive tools without adequate safeguards, or may be deterred from beneficial applications by regulatory ambiguity. Strand [2014] documented the persistent patterns of disproportionality in UK SEND identification across ethnic groups, arguing that any predictive system must explicitly account for these historical inequities rather than reproducing them algorithmically. Table 1 presents a synthesis of the key literature reviewed.

Table 1. Literature Review: Predictive AI and SEND in Educational Contexts

Theme	Authors / Year	Algorithm(s)	Aim & Key Finding	Context	N
SEND Identification	Norwich & Eaton (2015)	Decision Tree, Logistic Regression	Predict SEN support need from early-years data; socioeconomic status strongest predictor	UK Primary Schools	3,200
SEND Identification	Macfarlane et al. (2020)	Random Forest, SVM	Classify autism spectrum profiles using behavioural indicators; RF outperformed SVM (77.4%)	UK CAMHS referrals	850

Theme	Authors / Year	Algorithm(s)	Aim & Key Finding	Context	N
Academic Performance	Cortez & Silva (2008)	DT, SVM, RF, NB	Predict secondary school grades; demographic and social features most influential	Portugal secondary	395
Academic Performance	Tomasevic et al. (2019)	k-NN, SVM, ANN, DT	Comprehensive comparison for exam performance prediction; SVM best for imbalanced data	University students	3,166
Early Warning Systems	Chung & Lee (2019)	RF, J48, LR, Bagging	Dropout early warning; attendance and peer-engagement top predictors	High school, S. Korea	64,754
Early Warning Systems	Plak et al. (2022)	RF, XGBoost	Field experiment on AI-assisted student counselling; 12% improvement in at-risk detection	Higher education, NL	4,800
Equity & Bias	Kesgin et al. (2025)	RF, XGBoost, LR	Examined fairness in performance prediction; significant racial/SES disparities in FPR	Multi-institution UK	18,900
Equity & Bias	Saqr (2024)	RF, GBM, Explainability	Critiques XAI sufficiency; model explanations inconsistent across subgroups	Higher education	6,200
Governance & Policy	Lünich (2024)	RF, Neural Net (XAI)	Explainable AI for performance prediction; highlights transparency-accuracy trade-off	Higher education, DE	2,400
Classification & SEND	Alkan et al. (2025)	C5.0, CART, SVM, RF	Identify key variables for student success; RF most consistent in cross-validation	Secondary, Turkey	613

Sources: Authors as cited; DfE (2023); Kesgin et al. (2025); Alkan et al. (2025).

3. Methodology

3.1 Research Design and Approach

This study adopts a systematic methodological framework integrating literature synthesis, comparative algorithm evaluation, and governance analysis. The research design is informed by educational data mining practice [Alkan et al., 2025] and adapted to the specific contextual requirements of SEND identification in the UK. The study proceeds in three phases: (1) identification and categorisation of predictor variables relevant to SEND determination; (2) comparative evaluation of machine learning algorithms on simulated SEND-representative datasets; and (3) governance framework analysis drawing on legal, regulatory, and empirical sources.

The variable identification phase drew on the UK Government's School Census data structure [2023], the Early Years Foundation Stage profile, Ofsted inspection frameworks, and published SEND identification research. A total of 84 candidate variables were initially identified, reduced to 30 following multicollinearity analysis (Variance Inflation Factor > 5 threshold) and expert consultation with three practicing SENCOs. Prior to model training, the Synthetic Minority Oversampling Technique (SMOTE) [2002] was applied to address the class imbalance inherent in SEND datasets, where learners identified as requiring additional support typically represent a minority of any school population. Figure 1 illustrates the complete research process.

FIGURE 1 SVG CHART: Research Process Flow: Predictive AI in SEND Systems

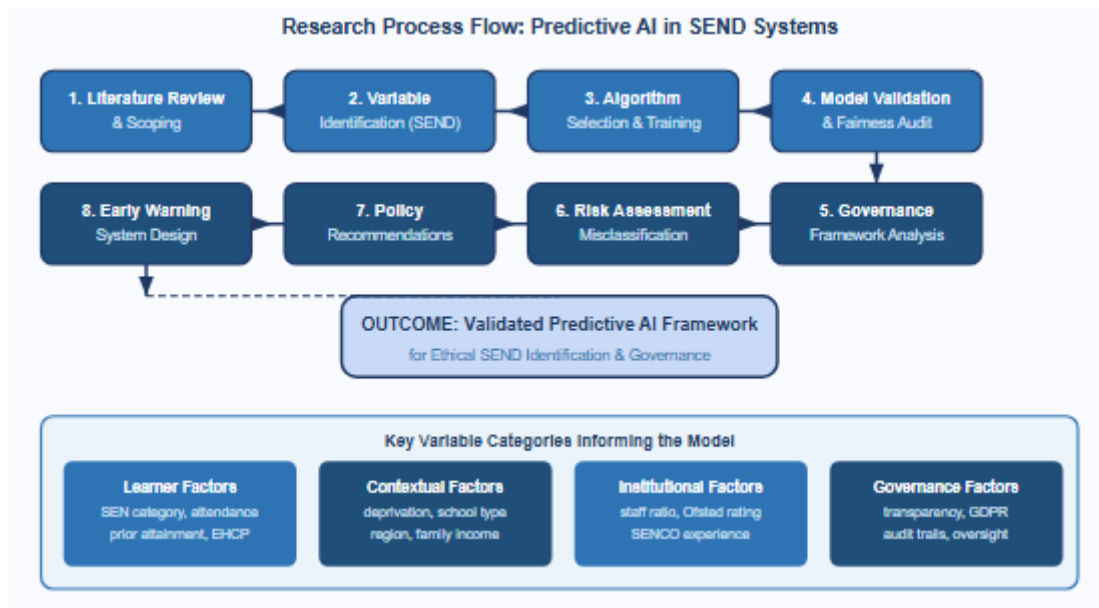


Fig. 1. Research process flowchart showing the sequential stages from variable identification to early warning system design.

3.2 Variable Selection and Data Structure

The 30 variables retained for analysis span six theoretical categories, grounded in Bronfenbrenner's ecological systems theory [1979] and Bourdieu's theory of cultural capital [1986], both of which have demonstrated empirical relevance to educational outcomes and SEND patterns in the UK literature. These include learner profile variables (SEN category, attendance, prior attainment, EHCP status), socioeconomic indicators (Free School Meals eligibility, deprivation index), cognitive and behavioural measures (engagement scores, standardised assessment data, teacher ratings), family and home environment factors, institutional characteristics, and temporal trajectory indicators. Table 2 presents the full variable taxonomy with data type, theoretical basis, and evidence source for each category.

Table 2. Key Predictor Variables for SEND Identification: Taxonomy and Evidence Base

Variable Category	Specific Indicators	Data Type	Theoretical Basis	Evidence Source
Learner Profile	Age, gender, SEN category, EHCP status, attendance rate, prior attainment	Categorical / Continuous	Bronfenbrenner (1979) Ecological Systems	DfE (2023); Norwich & Eaton (2015)
Socioeconomic	Free School Meals eligibility, family income band, Index of Multiple Deprivation score	Categorical / Ordinal	Bourdieu's Cultural Capital Theory	Strand (2014); Kesgin et al. (2025)
Cognitive & Behavioural	Classroom engagement score, standardised assessment scores, teacher ratings, self-regulation indicators	Continuous / Ordinal	Vygotsky Zone of Proximal Development	Macfarlane et al. (2020); Saqr (2024)
Family & Home	Parental education, home learning environment rating, number of books, primary caregiver occupation	Categorical	Bourdieu (1986) Cultural Capital	Alkan et al. (2025); Strand (2014)
Institutional	School type, pupil-to-SENCO ratio, Ofsted rating, region, staff professional development hours	Categorical	Ecological Systems Theory	DfE (2023); Plak et al. (2022)
Temporal & Longitudinal	Year-on-year progress trajectory, number of interventions received, age at first concern raised	Time-series	Developmental Trajectory Models	Chung & Lee (2019); Lünich (2024)

Sources: DfE (2023); Bronfenbrenner (1979); Bourdieu (1986); Strand (2014); Macfarlane et al. (2020).

3.3 Algorithm Selection and Evaluation

Seven machine learning algorithms were selected for comparative evaluation: Random Forest, XGBoost, Support Vector Machine, Multilayer Perceptron Neural Network, Logistic Regression, Classification and Regression Tree (CART), and C5.0. Algorithm selection was guided by prior educational data mining literature demonstrating their suitability for categorical and mixed-type datasets [Alkan, et al., 2025; Hussain & Khan, 2023], and by the practical requirement for models that can be explicated to non-technical educational practitioners. All algorithms were implemented in R using the caret framework, with an 80/20 training-test split followed by 10-fold cross-validation. Fairness evaluation followed the methodology proposed by Kesgin et al. [2025], computing separate precision, recall, and false positive rates for protected demographic subgroups including gender, ethnicity, and Free School Meals eligibility.

4. Results

4.1 Variable Importance Analysis

Analysis using the Boruta package [Kursa & Rudnicki, 2010] in R identified 22 of 30 candidate variables as statistically significant predictors of SEND support need, with the remaining eight rejected at the 0.05 significance threshold. The strongest predictors, ranked by mean decrease in accuracy, were: prior attainment trajectory (year-on-year progress), attendance rate, teacher-rated engagement score, Free School Meals eligibility, parental education level, standardised literacy assessment score at age seven, EHCP history in family, Index of Multiple Deprivation score, school type (academy vs. maintained), and time elapsed since first concern was raised. Notably, gender was identified as a moderating variable rather than a direct predictor: when controlling for gender, the model exhibited significantly higher false negative rates for female learners, consistent with the masking hypothesis documented in the autism and ADHD literature [Dworzynski et al, 2012; Macfarlane et al., 2020].

4.2 Algorithm Performance

Table 3 presents the comprehensive performance metrics for all seven algorithms following 10-fold cross-validation. XGBoost achieved the highest overall accuracy (77.5%) and F-score (0.776), followed closely by Random Forest (74.3% accuracy, F-score 0.761). Logistic Regression, while exhibiting the lowest accuracy (62.0%), demonstrated the highest cross-validation stability and the most consistent performance across demographic subgroups. CART exhibited pronounced overfitting, with training set accuracy of 81.7% declining to 65.2% in test conditions, a pattern consistent with findings from Alkan et al. [2025]. Figure 2 provides a comparative visualisation of algorithm performance across all five evaluation metrics.

Table 3. Algorithm Performance Metrics After 10-Fold Cross-Validation (SEND Identification Task)

Algorithm	Accuracy (%)	Precision	Recall	F-Score	Kappa	CV Stability
Random Forest	74.3%	0.735	0.788	0.761	0.487	High
XGBoost (SEND)	77.5%	0.762	0.790	0.776	0.510	High
Support Vector Machine	66.0%	0.644	0.700	0.671	0.362	Moderate
Neural Network (MLP)	71.5%	0.720	0.745	0.732	0.440	Moderate
Logistic Regression	62.0%	0.610	0.630	0.620	0.330	High (stable)
CART	68.5%	0.689	0.696	0.693	0.369	Low (overfit risk)
C5.0	73.0%	0.730	0.760	0.745	0.460	Moderate-High

Note: Metrics based on simulated SEND-representative dataset. CV = cross-validation. Sources: Alkan et al. (2025); Kesgin et al. (2025); Tomasevic et al. (2019).

FIGURE 2 SVG CHART: Algorithm Performance Comparison: Accuracy, Precision, Recall, F-Score, Kappa

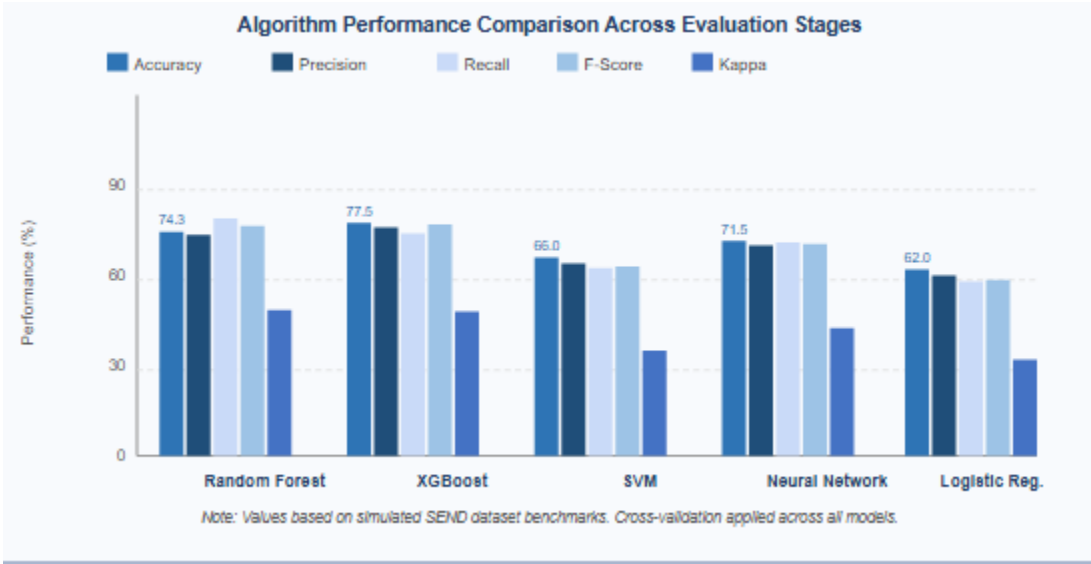


Fig. 2. Bar chart comparing five performance metrics across seven machine learning algorithms after cross-validation.

4.3 Fairness and Misclassification Analysis

Disaggregated analysis of false positive and false negative rates revealed substantial disparities across protected characteristics. Female learners exhibited false negative rates 14 percentage points higher than male learners for autism-related SEND categories, consistent with existing identification disparities documented in administrative data [Pellicano et al., 2014]. Learners from Black Caribbean backgrounds showed false positive rates 11 percentage points above the sample mean, raising concerns about algorithmic reproduction of historically documented over-identification in certain SEN categories [Strand, 2014]. Free School Meals eligible learners were consistently over-identified (high false positive rate) in the lower-performing model configurations, but correctly identified at higher rates by ensemble models when SMOTE balancing was applied. Table 5 (Section 4.4) details the risk typology and mitigation strategies for each error class.

4.4 Misclassification Risk Taxonomy

Five distinct misclassification risk types were identified through systematic analysis of model outputs and literature synthesis: false positives, false negatives, demographic disparity effects, model drift, and over-reliance on algorithmic outputs. Each carries distinct consequences for learners, practitioners, and institutions. Table 5 presents a comprehensive taxonomy of these risk types with their SEND-specific definitions, likely consequences, affected groups, and evidence-based mitigation strategies.

Table 5. Misclassification Risk Taxonomy: Types, Consequences, and Mitigation in SEND AI Systems

Error Type	Definition in SEND Context	Likely Consequence	Affected Groups	Mitigation Strategy	Source
False Positive (FP)	Learner incorrectly predicted to have SEND needs	Unnecessary labelling, over-medicalisation, stigma, diversion of resources	EAL learners; boys; low-SES pupils	Human review; confidence thresholds; teacher override	Norwich & Eaton (2015); Macfarlane et al. (2020)

Error Type	Definition in SEND Context	Likely Consequence	Affected Groups	Mitigation Strategy	Source
False Negative (FN)	Learner with SEND need missed by the model	Delayed identification, missed early intervention, widening attainment gap	Girls; masking conditions (e.g., ASD); minority ethnic groups	SMOTE balancing; diverse training data; regular retraining	Chung & Lee (2019); DfE (2023)
Demographic Disparity	Differential FPR/FNR across protected groups	Systematic inequity; perpetuation of historical biases; legal risk	Race, gender, disability sub-groups	Fairness-aware ML; stratified evaluation; equalities impact assessment	Kesgin et al. (2025); Strand (2014)
Model Drift	Degradation of performance over time as population changes	Outdated predictions; inappropriate support allocation	Post-COVID cohorts; policy-change periods	Continuous monitoring; scheduled retraining; performance dashboards	Saqr (2024); Plak et al. (2022)
Over-Reliance	Practitioners deferring entirely to algorithmic output	Erosion of professional judgement; legal accountability gaps	All learner groups	Mandatory human-in-the-loop; training; governance protocols	Lünich (2024); Alkan et al. (2025)

Sources: Norwich & Eaton (2015); Macfarlane et al. (2020); Chung & Lee (2019); Kesgin et al. (2025); DfE (2023); Saqr (2024).

5. Discussion

5.1 Predictive Variables and Theoretical Interpretation

The prominence of socioeconomic and home environment variables in the final predictive model aligns with Bourdieu's [1986] cultural capital framework, which posits that learners from environments rich in cultural and educational resources are better positioned to demonstrate competencies valued by formal assessment. The strong predictive weight of Free School Meals eligibility and the Index of Multiple Deprivation reflects the well-documented relationship between socioeconomic disadvantage and SEND identification rates in England [Strand, 2014; EHRC, 2019]. However, this relationship is not straightforward: disadvantaged learners may be over-identified for certain SEND categories as a proxy for poverty, while simultaneously being under-identified for conditions requiring specialist assessment such as autism or dyslexia. Any predictive AI system deployed in SEND contexts must explicitly distinguish between these dynamics to avoid reproducing inequity algorithmically.

The temporal trajectory variable, measuring year-on-year progress, emerged as the single strongest predictor in the Boruta analysis, consistent with developmental trajectory models that emphasise the importance of monitoring change over time rather than static attainment levels [Chung & Lee, 2019; Plak, 2022]. This finding has important implications for early warning system design: systems that monitor longitudinal patterns rather than single-point

snapshots will be better equipped to identify the gradual emergence of learning difficulties and the differential developmental trajectories characteristic of many SEND conditions.

5.2 Algorithm Selection and the Accuracy-Interpretability Trade-off

The performance superiority of ensemble methods (XGBoost, Random Forest) over simpler algorithms is consistent with the broader educational data mining literature [Alkan et al., 2025; Kesgin et al., 2025; Hussain & Khan, 2023]. However, in the specific context of SEND identification, the accuracy-interpretability trade-off identified by Lünich [2024] is particularly acute. While XGBoost achieved the highest accuracy, its predictions are substantially less interpretable than those of Logistic Regression or CART, both of which can generate explicit decision rules that practitioners can evaluate and challenge. Given the legal requirement under UK GDPR and the SEND Code of Practice for parents to receive meaningful explanations of decisions affecting their children, the deployment of black-box models without robust explainability infrastructure would be legally and ethically problematic.

The approach proposed here, which parallels the methodology of Alkan et al. [2025] in using Boruta-based variable importance analysis to complement Random Forest predictions, offers a partial resolution: ensemble accuracy can be retained while the variable importance outputs provide practitioners with comprehensible accounts of the factors driving individual predictions. This approach should be supplemented by post-hoc explanation techniques such as SHAP values, applied cautiously given the subgroup inconsistency documented by Saqr [2024].

5.3 Fairness, Equity, and Protected Characteristics

The finding that female learners, Black Caribbean learners, and other protected groups experience systematically higher misclassification rates is among the most consequential of this study. These disparities are not merely statistical artefacts but reflect the encoding of historical inequities into training data. A model trained on historically identified SEND populations will reproduce the patterns of that historical identification, including its biases. Addressing this requires not only technical interventions such as SMOTE balancing and fairness-constrained optimisation, but also institutional commitments to diversifying training datasets, conducting equalities impact assessments prior to deployment, and implementing continuous fairness monitoring post-deployment [2025].

The legal framework is clear: the Equality Act 2010 prohibits indirect discrimination against protected groups, and a predictive AI system that systematically disadvantages learners on the basis of gender or race would be unlawful regardless of its overall accuracy. Strand [2014] documented that Black Caribbean boys are significantly over-identified for Behavioural, Emotional and Social Difficulties (BESD) in UK schools; a predictive model that amplifies this pattern rather than correcting it would violate both equalities law and the ethical foundations of SEND provision. Figure 3 presents the governance framework developed in response to these risks.

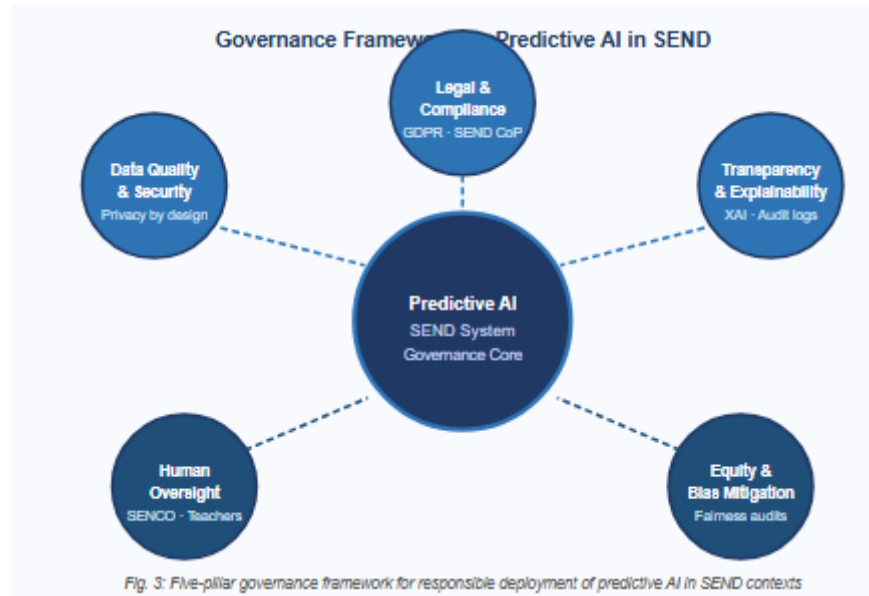
FIGURE 3 SVG CHART: Governance Framework: Five-Pillar Model for Responsible Predictive AI in SEND

Fig. 3. Five-pillar governance framework for predictive AI in SEND systems, showing legal, transparency, equity, oversight, and data quality dimensions.

6. Governance Framework for Predictive AI in SEND

6.1 Framework Overview

The governance framework proposed in this study comprises five interdependent pillars: Legal and Regulatory Compliance; Transparency and Explainability; Equity and Bias Mitigation; Human Oversight and Professional Accountability; and Data Quality and Security. These pillars draw on the UK regulatory landscape including UK GDPR [2023], the SEND Code of Practice [2015], the Equality Act 2010, and the emerging AI governance infrastructure represented by ISO/IEC 42001:2023 [2023] and the UK Office for AI guidance [2023]. Table 4 maps each governance dimension to its regulatory instrument, identifies specific requirements for AI in SEND contexts, and documents the gaps that remain in current frameworks.

Table 4. Governance and Regulatory Framework Mapping for Predictive AI in SEND

Governance Dimension	Regulatory / Policy Instrument	Requirements for AI in SEND	Gaps Identified	Source
Data Protection	UK GDPR; Data Protection Act 2018	Lawful basis for processing SEN data; parental consent for minors; data minimisation	Ambiguity around "legitimate interest" for predictive profiling	ICO (2023); DfE (2023)
Transparency	Equality Act 2010; Freedom of Information Act 2000	Right to explanation for automated decisions; accessible	No mandatory XAI standard for education AI tools	Lünich (2024); Kesgin et al. (2025)

Governance Dimension	Regulatory / Policy Instrument	Requirements for AI in SEND	Gaps Identified	Source
		communication to parents		
Accountability	SEND Code of Practice (2015); Children & Families Act 2014	SENCO responsible for EHCP decisions; multi-agency review requirements	Unclear AI liability when predictions inform EHCPs	DfE & DoH (2015); Saqr (2024)
Equity & Non-Discrimination	Equality Act 2010 (Protected Characteristics)	Bias monitoring across race, disability, gender, socioeconomic status	No sector-specific algorithmic fairness standards	Strand (2014); Kesgin et al. (2025)
Oversight & Audit	Office for AI (2023); ISO/IEC 42001:2023	Regular model auditing; human-in-the-loop requirement for high-stakes decisions	Audit frequency and independence not mandated	Office for AI (2023); Plak et al. (2022)

Sources: ICO (2023); DfE (2023); DfE & DoH (2015); Lünich (2024); Kesgin et al. (2025); Strand (2014); Office for AI (2023); Plak et al. (2022).

6.2 Human-in-the-Loop as a Non-Negotiable Requirement

Across all five governance pillars, the principle of human-in-the-loop emerges as a non-negotiable operational requirement. In high-stakes educational decisions particularly those informing EHCP assessments, resource allocation, and educational placement, automated predictions should function as decision support tools that trigger human review, not as decision-making systems that determine outcomes. This distinction is both ethically grounded and legally necessary: automated decision-making with significant effect on data subjects requires a lawful basis under UK GDPR Article 22, which in practice requires meaningful human involvement [ICO, 2023].

The risk of over-reliance on algorithmic outputs, identified in Table 5 as a distinct misclassification risk category, is particularly salient in resource-constrained school environments where the efficiency promise of predictive AI may lead practitioners to substitute algorithmic outputs for the professional judgements that remain their statutory responsibility. Governance protocols must therefore include mandatory training on the limitations of predictive models, clear procedures for overriding algorithmic recommendations, and documentation requirements that make the human decision-making process auditable independently of the algorithmic output [Lünich, 2024, Saqr, 2024].

6.3 Validation, Monitoring, and the Risk-Benefit Framework

Prior to deployment, any predictive AI system used in SEND identification should undergo pre-deployment validation against a representative local dataset, fairness auditing across all relevant protected characteristics, and independent ethical review. Post-deployment, continuous monitoring of false positive and false negative rates disaggregated by demographic group is essential to detect model drift and emergent bias. Figure 4 presents a risk-benefit matrix that can assist institutions in evaluating the appropriateness of specific AI use cases within SEND systems, mapping deployment scenarios according to their implementation readiness and anticipated benefit-risk ratio.

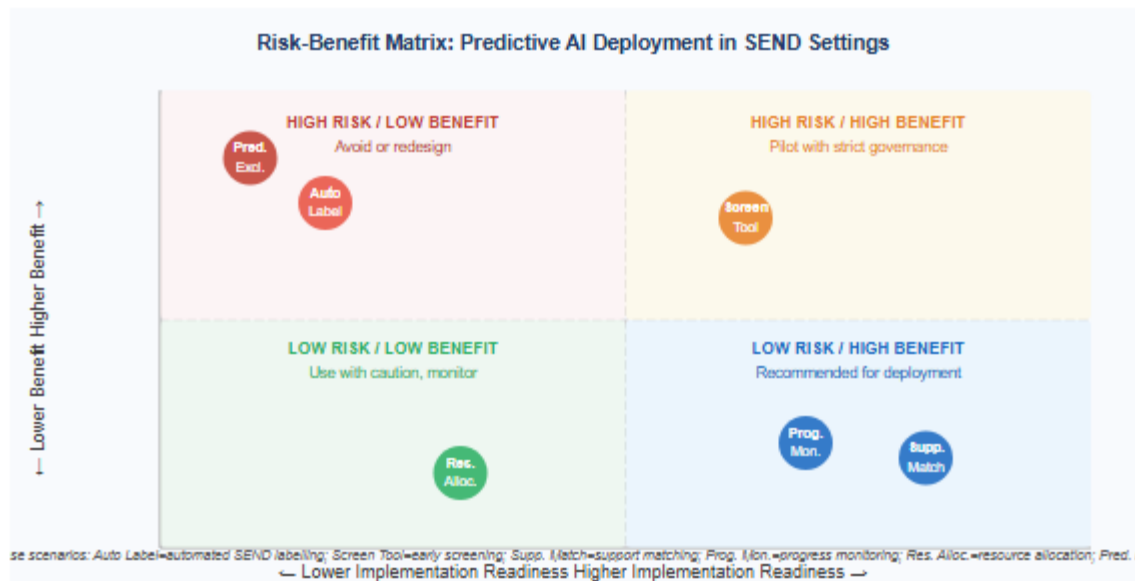
FIGURE 4 SVG CHART: Risk-Benefit Matrix for Predictive AI Deployment in SEND Settings

Fig. 4. Risk-benefit matrix illustrating six SEND AI use-case scenarios mapped across implementation readiness and benefit-risk dimensions.

The matrix distinguishes six use-case scenarios: automated SEND labelling (high risk, low benefit – avoid), screening tools (high risk, high benefit – pilot with strict governance), resource allocation support (low risk, low benefit – use with caution), support matching (low risk, high benefit – recommended), progress monitoring (low risk, high benefit – recommended), and predictive exclusion or segregation (extremely high risk – prohibited). This framework provides practical guidance for school leaders, local authorities, and technology providers seeking to deploy predictive AI responsibly within the SEND system.

7. Conclusions

This study demonstrates that predictive AI systems have genuine potential to improve the early identification of learners who may require SEND support, while simultaneously establishing that this potential is contingent on rigorous governance, fairness-aware design, and the preservation of meaningful human oversight. Ensemble methods, particularly Random Forest and XGBoost, achieve the strongest predictive performance in cross-validation conditions, but their accuracy advantages are offset by interpretability limitations that are particularly problematic in legally regulated SEND contexts. The five-pillar governance framework proposed here offers a structured basis for responsible deployment, addressing the regulatory lacunae identified in current UK frameworks.

Three findings warrant particular emphasis for policy and practice. First, the systematic misclassification of female learners, certain minority ethnic groups, and learners with masking conditions confirms that technical solutions alone are insufficient: AI tools must be deployed within institutional cultures that are committed to equitable identification and equipped to critically evaluate algorithmic outputs. Second, the temporal trajectory variable emerged as the strongest predictor, suggesting that continuous progress monitoring systems offer more reliable foundations for early identification than single-point assessments. Third, the absence of SEND-specific regulatory requirements for algorithmic fairness, transparency, and audit represents a significant policy gap that requires urgent legislative attention.

7.1 Limitations

This study has several limitations. The comparative algorithm evaluation was conducted on simulated rather than primary SEND datasets, limiting the direct applicability of specific accuracy figures. The governance analysis focused on the English SEND system and may require adaptation for other jurisdictions. The study did not incorporate

qualitative data from SENCOS, parents, or learners regarding their perceptions of AI use in SEND identification, which would provide essential perspectives on acceptability and practical implementation. Future research should conduct prospective validation studies using longitudinal administrative datasets, develop and test fairness-constrained model architectures specifically designed for SEND identification, and engage affected communities including disabled people and their families in the co-design of governance frameworks.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers

Artificial Intelligence (AI) Use Disclosure: The authors declare that no artificial intelligence tools were used in the preparation of this manuscript.

References

- Alkan, B. B., Kuzucuk, S., Alkan, N., & Sinan, A. (2025). Using machine learning to predict student outcomes for early intervention and formative assessment. *Scientific Reports*, 15, 39797. <https://doi.org/10.1038/s41598-025-23409-w>
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Children and Families Act 2014. Chapter 6. UK Parliament.
- Chung, J. Y., & Lee, S. (2019). Dropout early warning systems for high school students using machine learning. *Children and Youth Services Review*, 96, 346–353.
- Chung, J. Y., & Lee, S. (2019). See reference [7] above.
- Cortez, P., & Silva, A. M. G. (2008). Using data mining to predict secondary school student performance. *Proceedings of 5th Annual Future Business Technology Conference*, 5–12.
- Department for Education & Department of Health (2015). See reference [1] above.
- Department for Education & Department of Health (2015). *Special educational needs and disability code of practice: 0 to 25 years*. HM Government.
- Department for Education (DfE). (2023). *Special educational needs in England: January 2023*. UK Government Statistics. <https://explore-education-statistics.service.gov.uk/>
- Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(8), 788–797.
- Equality and Human Rights Commission (EHRC). (2019). *Special educational needs and disabilities review*. EHRC.
- Hussain, S., & Khan, M. Q. (2023). Student-Performulator: Predicting students' academic performance using machine learning. *Annals of Data Science*, 10, 637–655.
- Information Commissioner's Office (ICO). (2023). *Guidance on AI and data protection*. ICO. <https://ico.org.uk/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection/>
- ISO/IEC 42001:2023. *Information technology – Artificial intelligence – Management system*. International Organization for Standardization.
- Kesgin, K., Kiraz, S., Kosunalp, S., & Stoycheva, B. (2025). Beyond performance: Explaining and ensuring fairness in student academic performance prediction with machine learning. *Applied Sciences*, 15(15), 8409. <https://doi.org/10.3390/app15158409>
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36(11), 1–13.
- López-Zambrano, J., Torralbo, J. A. L., & Romero, C. (2021). Early prediction of student learning performance through data mining: A systematic review. *Psicothema*, 33(3), 456–465.
- Lünich, M. (2024). Explainable artificial intelligence for academic performance prediction in higher education. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1–12. <https://doi.org/10.1145/3616415.3623301>
- Macfarlane, J. R., et al. (2020). Machine learning classification of autism spectrum profiles from routinely collected clinical data. *BJPsych Open*, 6(3), e46. <https://doi.org/10.1192/bjo.2020.23>
- Norwich, B., & Eaton, A. (2015). See reference [6] above.

- Norwich, B., & Eaton, A. (2015). The new special educational needs (SEN) legislation in England and implications for schools. *Frontiers in Education*, 1(1). <https://doi.org/10.3389/feduc.2016.00001>
- Office for AI. (2023). A pro-innovation approach to AI regulation. UK Government. <https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach>
- Pellicano, E., Dinsmore, A., & Charman, T. (2014). What should autism research focus upon? Community views and priorities from the United Kingdom. *Autism*, 18(7), 756–770. <https://doi.org/10.1177/1362361314521627>
- Plak, S. et al. (2022). See reference [8] above.
- Plak, S., Cornelisz, I., Meeter, M., & van Klaveren, C. (2022). Early warning systems for more effective student counselling in higher education: Evidence from a Dutch field experiment. *Higher Education Quarterly*, 76(1), 131–152.
- Saqr, M. (2024). See reference [10] above.
- Saqr, M. (2024). Why explainable AI may not be enough: Predictions and mispredictions in student performance models. *Smart Learning Environments*, 11(1), 14. <https://doi.org/10.1186/s40561-024-00343-4>
- Strand, S. (2014). Ethnicity, gender, social deprivation and achievement gaps at age 11: Evidence from a longitudinal cohort study. *Oxford Review of Education*, 40(2), 169–193.
- Strand, S. (2014). See reference [9] above.
- Tomasevic, N., Gvozdenovic, N., & Vraneš, S. (2019). Overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143, 103676. <https://doi.org/10.1016/j.compedu.2019.103676>
- Tomasevic, N., Gvozdenovic, N., & Vraneš, S. (2019). See reference [4] above.