Pharmacophore

ISSN-2229-5402

Journal home page: http://www.pharmacophorejournal.com

EXPLAINABLE AI FOR PEDIATRIC DOSING ERROR PREDICTION USING PRESCRIPTION TEXT AND WEIGHT-BASED RULES

Anna Svensson[1]*, Erik Lindberg[1]

1. *Department of Computational Pharmaceutical Sciences, Faculty of Medicine, Uppsala University, Uppsala, Sweden.*

ARTICLE INFO

Received:
03 December 2025
Received in revised form:
28 February 2026
Accepted:
01 March 2026
Available online:
28 April 2026

Keywords: Explainable artificial intelligence, Pediatric dosing, Medication safety, prescription text, SHAP, Clinical decision support

ABSTRACT

Pediatric prescribing is particularly susceptible to dosing errors because many medication orders must be individualized based on weight, age, formulation, route, and frequency, and ambiguous prescription text can further obscure the intended dose, increasing the risk that an unsafe order reaches the child. Conventional clinical decision support systems typically rely on static dose-range alerts and rigid rule thresholds, which may flag obvious outliers but often lack clinical nuance and generate alerts that are difficult to interpret or justify. To address these limitations, this article proposes an explainable machine learning model for pediatric dosing error prediction that leverages prescription text and weight-based dosing rules, attributing each flagged prescription to specific contributing factors such as dose-per-kilogram deviations, formulation mismatches, or ambiguous administration instructions. The model conceptually employs a gradient-boosted tree or similarly interpretable architecture, trained on pediatric prescription records enriched with structured patient characteristics, computable dosing-rule features, and natural language processing outputs from prescription text, while SHAP-based explanations decompose each prediction into feature-level contributions that can be communicated to pharmacists and prescribers. In practice, the model could identify prescriptions likely to contain weight-based miscalculations, inappropriate formulations, or incorrect frequencies, providing explanations such as a prescribed dose exceeding the recommended weight-based range combined with a free-text instruction containing an ambiguous abbreviation. By combining predictive modeling with transparent, actionable explanations, an explainable AI approach could support safer pediatric prescribing, though prospective evaluation is needed to assess clinical usefulness, workflow integration, explanation quality, and impact on medication safety.

To Cite This Article: Svensson A, Lindberg E. Explainable AI for Pediatric Dosing Error Prediction Using Prescription Text and Weight-Based Rules. Pharmacophore. 2026;17(2):1-11. https://doi.org/10.51847/v5wAxGMQYm

Introduction

Pediatric medication errors remain a persistent safety concern because children require individualized prescribing that reflects weight, age, maturation, formulation, and route of administration. Systematic reviews of pediatric dose errors and preventable adverse drug events show that dosing mistakes are especially important in hospital settings, where even small calculation errors can produce clinically meaningful harm [1, 2]. The need to include accurate patient weight on prescriptions reflects the central role of weight-based dosing in pediatric medication safety [3]. These vulnerabilities are intensified in neonates and infants, where pharmacokinetic variability and rapidly changing body size complicate dose selection [4].

Computerized provider order entry and clinical decision support have been introduced to reduce preventable pediatric prescribing errors, yet many systems still rely on fixed dose thresholds and static rule logic. Reviews of pediatric computerized ordering show that technology can prevent some dose errors but may also introduce new error pathways when interfaces, rules, or workflows are poorly aligned with clinical practice [5, 6]. Pediatric medication alerts can influence prescriber response, but their usefulness depends on relevance, timing, and trustworthiness [7]. A more context-aware model is therefore needed, particularly one that can explain why a prescription appears unsafe rather than merely interrupting the prescriber.

Natural language processing and machine learning create an opportunity to learn from the full prescription context rather than from structured dose fields alone. Prescription text contains dosage instructions, sig codes, units, frequencies, and route information that can be parsed into clinically meaningful features, while structured data such as weight and age can anchor these text-derived values to pediatric dosing rules [8, 9]. Interpretable medication-error detection frameworks suggest that

Corresponding Author: Anna Svensson; Department of Computational Pharmaceutical Sciences, Faculty of Medicine, Uppsala University, Uppsala, Sweden. E-mail: anna.svensson@gmail.com

embedding-based language features and explainable classification can support prescription review without reducing the process to opaque automation [10]. Explainable AI methods in healthcare further indicate that feature attribution can translate model behavior into evidence that clinicians can inspect [11, 12].

The central thesis of this article is that an explainable AI system could predict pediatric dosing errors by integrating prescription text, patient characteristics, and computable weight-based dosing rules. Such a model would not replace pharmacist or prescriber judgment but would provide a transparent risk assessment that identifies the specific prescription elements responsible for the alert. SHAP-based explanation methods are particularly relevant because they can show how individual features increase or decrease a prediction while preserving a global view of model behavior. In pediatric medication safety, this design aligns predictive modeling with the clinical need for accountable, reviewable, and actionable decision support [13].

Background

Pediatric Dosing Errors: Types, Causes, and Consequences

Pediatric dosing errors include wrong dose, wrong frequency, incorrect formulation, inappropriate route, and errors arising from transcription or interpretation of free-text instructions. Weight-based miscalculation is especially consequential because the same medication may require different doses across infants, children, adolescents, and children with obesity [3, 14]. Medication safety reviews emphasize that pediatric errors arise from both clinical complexity and system-level vulnerabilities, including fragmented information, incomplete documentation, and confusing order-entry workflows [6, 15]. These risks are clinically important because dose errors may remain undetected until pharmacy verification, bedside administration, or adverse patient response.

Weight-Based Dosing Rules and Current Decision Support

Weight-based dosing rules are commonly encoded in pediatric clinical decision support systems as recommended dose ranges, maximum dose checks, frequency constraints, and formulation-specific safeguards. Pediatric dose calculators and integrated decision support tools have been developed to reduce manual calculation burden and standardize prescribing logic [16, 17]. However, systematic reviews caution that computerized decision support does not automatically eliminate pediatric dose errors, especially when rules are incomplete, alerts lack context, or clinicians override warnings perceived as irrelevant [2, 5]. The design challenge is to retain the safety value of computable dosing rules while adding enough contextual interpretation to reduce unnecessary interruption.

NLP of Prescriptions and Clinical Text

Prescription text often contains compact or ambiguous instructions that require interpretation before medication safety rules can be applied. Natural language processing can extract drug names, numeric doses, units, frequencies, routes, and administration instructions from free-text prescriptions, allowing these elements to be compared with structured dose and patient data [9]. Recent work on medication-direction errors in online pharmacy settings shows that language models can help identify unsafe or inconsistent directions when prescription text is treated as a safety-critical data source [8]. Interpretable text-based error-detection frameworks further support the idea that language-derived prescription features should be visible to reviewers rather than hidden inside a black-box model [10].

Machine Learning for Medication Error Detection

Machine learning has been explored for neonatal adverse drug reaction screening, neonatal medication error detection, medication administration error risk prediction, and pediatric dose optimization, showing that data-driven methods can represent complex medication-safety patterns that static rules may miss [18-21]. In pediatric and neonatal settings, such models are most useful when they combine medication details with patient-specific variables because error risk depends on clinical context rather than on the drug order alone. Yet machine learning systems can be difficult to trust when clinicians cannot determine whether a prediction is driven by dose deviation, age, formulation, route, or documentation artifact [11]. For pediatric dosing safety, opacity is not merely a technical concern but a barrier to clinical adoption.

Explainable AI in Clinical Settings

Explainable AI methods seek to make model predictions interpretable by identifying the features or textual elements that most influenced a prediction. Healthcare explainability surveys distinguish between local explanations for individual cases and global explanations that describe the model's general behavior, both of which are relevant to medication alerting [12, 22]. SHAP is especially useful for structured clinical prediction because it can assign direction and magnitude to each feature contribution while supporting both patient-level and population-level interpretation. In medication safety workflows, such explanations can help pharmacists and prescribers decide whether to accept, override, or investigate an alert [13].

Model Development Overview

High-Level Predictive and Explanatory Pipeline

The proposed pipeline begins when a pediatric prescription is entered into the electronic ordering system and passed to two parallel components: an NLP module that parses prescription text and a rule-based calculator that evaluates weight-based dosing features. The resulting structured feature vector would then be processed by a gradient-boosted classifier or comparable

explainable model that estimates whether the order may contain a dosing, frequency, or formulation error. A SHAP explanation layer would identify the leading contributors to the prediction, following established approaches for local and global interpretation of tree-based clinical models. This architecture extends pediatric decision support concepts by combining rule-derived safeguards with data-driven pattern recognition [13, 16].

Figure 1 illustrates the proposed explainable AI architecture that converts pediatric prescription text, patient characteristics, and weight-based dosing rules into transparent dosing-error predictions and pharmacist–prescriber review actions.

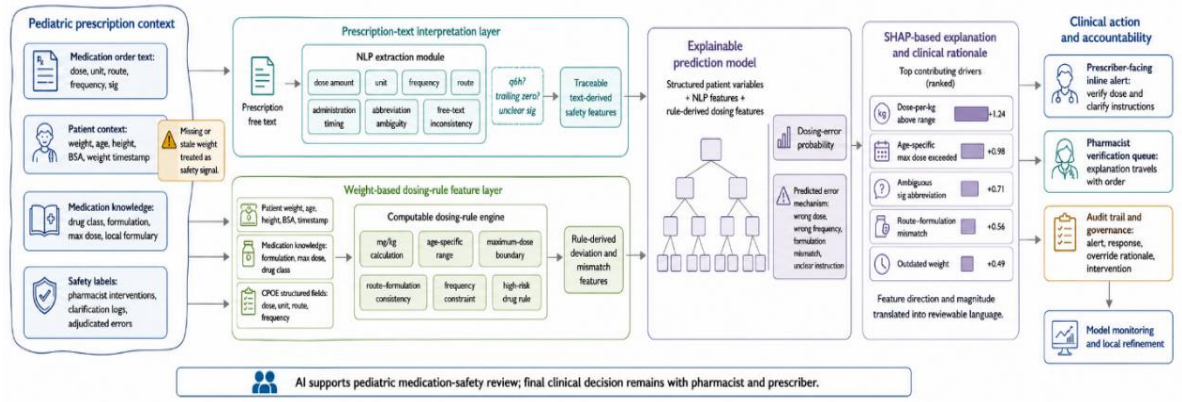

Figure 1. Explainable AI architecture for pediatric dosing error prediction using prescription text and weight-based rules.

Core Input Features

Core patient features would include weight, age, height, body surface area when clinically relevant, and indicators of missing or stale anthropometric data. Prescription features would include drug name, dose, unit, route, frequency, formulation, free-text sig, calculated dose per kilogram, and NLP-derived flags for unclear or unsafe instructions. Rule-based features would encode deviation from age-specific or weight-specific recommended ranges, maximum-dose boundaries, route-formulation consistency, and high-risk abbreviations that require pharmacist review [3, 23]. This feature strategy reflects pediatric safety concerns identified in dosing standardization, prescription-weight documentation, and neonatal decision support literature [3, 23, 24].

Design Principles

The model should be interpretable by design, fast enough to support near-real-time alerting, adaptable to evolving dosing guidelines, and capable of distinguishing likely errors from clinically justified deviations. Pediatric dose-support tools show the value of embedding dosing logic directly into prescribing workflows, while studies of alert response emphasize that clinicians need alerts that are specific and clinically meaningful [7, 17]. The model should therefore present concise explanations that fit within the ordering process while allowing deeper review when needed. It should also preserve clinical accountability by supporting pharmacist and prescriber judgment rather than automating final decisions [13].

Data Sources and Feature Engineering

Assembling a Pediatric Prescription Dataset

A conceptual development dataset would be assembled from electronic health record prescription orders linked to patient demographics, weight history, diagnosis context, medication formulation, and pharmacist intervention documentation. Error labels could be derived from expert chart review, pharmacy clarification logs, medication safety reports, or adjudicated intervention records, recognizing that each source captures a different part of the medication-safety process. Prior pediatric and neonatal medication-error studies highlight the importance of combining prescribing data with clinical context when identifying preventable or likely errors [19, 20, 25]. Because the article is conceptual, the dataset design should be treated as a proposed framework rather than as evidence of completed model training.

NLP Feature Extraction from Prescription Text

NLP feature extraction would transform prescription text into structured variables by parsing expressions such as dose amount, unit, frequency, route, administration timing, and free-text modifiers. This step is essential because many medication-direction errors arise from how instructions are written, not only from structured dose fields [8]. Prescription-text modeling in pharmacoepidemiologic data demonstrates that dosage and duration can be inferred from free-text medication directions when appropriate text-processing methods are used [9]. In an explainable medication-safety model, extracted text features should remain traceable so that the final alert can point to the exact phrase or abbreviation that contributed to concern [10].

Encoding Weight-Based and Age-Specific Rules

Weight-based and age-specific rules would be encoded as computable features rather than as isolated interruptive alerts. For example, the model could represent the ratio between prescribed dose per kilogram and the recommended upper range, a

categorical flag for below-range or above-range dosing, and route- or formulation-specific inconsistency indicators. Pediatric dose calculators and decision support platforms demonstrate how structured dosing logic can be incorporated into clinical systems, while dose standardization efforts show the importance of harmonizing local practice with accepted pediatric ranges [16, 23]. Encoding rules as features allows the model to learn when a deviation is strongly concerning and when additional context may make the prescription clinically plausible.

Table 1 defines how patient context, prescription-text features, computable dosing rules, and medication-risk variables can be converted into clinically interpretable explanations for pediatric dosing-error prediction.

Table 1. Feature-to-explanation architecture for pediatric dosing error prediction

Input domain	Representative variables or extracted features	Safety question addressed	Model role	Explanation delivered to clinician	Medication-safety value
Patient anthropometrics	Current weight, weight timestamp, height, BSA when relevant, missing or stale weight indicator	Is the prescription being evaluated against reliable pediatric dosing context?	Anchors weight-based and age-sensitive calculations	"Dose calculation depends on an outdated or missing weight."	Prevents falsely confident dose assessment when anthropometric data are unreliable
Age and developmental context	Age group, neonatal/infant/child/adolescent category, age-specific dosing range	Is the same absolute dose appropriate for the child's developmental stage?	Modifies expected dose range and risk interpretation	"The selected dose exceeds the expected range for this age group."	Supports safer prescribing across rapidly changing pediatric dosing needs
Prescription dose structure	Ordered dose, unit, strength, calculated mg/kg, total daily dose	Does the order deviate from recommended pediatric dose-per-weight limits?	Provides primary quantitative risk signal	"The prescribed mg/kg dose is above the recommended range."	Converts raw order fields into clinically interpretable dosing-risk evidence
Route and formulation	Route, dosage form, concentration, formulation availability, route–formulation compatibility	Is the selected product clinically consistent with the intended administration route?	Detects mismatch patterns not captured by dose alone	"The selected formulation is inconsistent with the intended route."	Identifies formulation-related safety risks before dispensing or administration
Frequency and timing	Frequency, interval, total daily frequency, administration timing	Could the order produce excessive exposure or unclear administration timing?	Adds temporal dosing-risk signal	"The ordered frequency increases total daily exposure beyond the expected range."	Extends safety review beyond single-dose calculation
Prescription text and sig	Free-text instructions, abbreviations, trailing zeros, unclear modifiers, conflicting directions	Does the written instruction introduce ambiguity or interpretation risk?	Supplies NLP-derived ambiguity and inconsistency features	"The sig contains an abbreviation or phrase that may be misinterpreted."	Makes prescription text visible as a safety-critical data source
Medication-risk context	High-risk medication class, narrow therapeutic index flag, local formulary constraints, maximum-dose boundary	Does the drug require heightened pediatric dose scrutiny?	Reweights concern according to medication-specific risk	"This medication has a stricter pediatric safety boundary."	Prioritizes clinically consequential alerts over low-risk deviations
Prior review and labeling evidence	Pharmacist clarification, intervention logs, adjudicated error labels, override rationale	Has a similar prescription pattern previously required correction?	Provides supervised learning signal and local workflow grounding	"This pattern resembles orders previously requiring pharmacist clarification."	Links predictive modeling to real medication-safety review experience
Rule-derived deviation features	Below-range flag, above-range flag, max-dose exceedance, ratio to upper limit	Which computable dosing rule is most responsible for concern?	Converts static rules into graded model features	"The main driver is deviation from the local weight-based dosing rule."	Preserves rule-based safety logic while allowing contextual interpretation
Explanation metadata	SHAP rank, feature direction, contribution magnitude, local explanation text	Why was this specific prescription flagged?	Translates prediction into	"The alert is driven by dose-per-kg excess	Supports clinician trust, targeted verification, and

	transparent rationale	and ambiguous sig text."	accountable decision support

Explainable Model Architecture

Model Choice – Gradient-Boosted Trees with SHAP

A gradient-boosted tree model would be a suitable conceptual architecture because pediatric prescription safety data combine numeric variables, categorical medication attributes, missingness indicators, and rule-derived features. Tree-based models are compatible with SHAP explanations, which can decompose a prediction into feature-level contributions and support both single-order review and system-wide monitoring. Explainable AI research in healthcare emphasizes that model choice should be guided not only by predictive capacity but also by the ability to support clinician understanding and accountability [11, 12]. For pediatric dosing alerts, this makes gradient-boosted trees attractive as a transparent alternative to purely opaque prediction systems.

Input Feature Vector and Pre-processing

The input feature vector would combine normalized dose-per-kilogram values, one-hot encoded medication classes, route and formulation indicators, NLP-extracted sig features, patient age and weight variables, and explicit flags for missing or outdated anthropometric data. Missing weight or age should not be silently ignored because pediatric dosing calculations depend directly on these values, and inaccurate weight documentation can undermine prescription safety [3]. In children with obesity or special dosing considerations, preprocessing should preserve signals that may indicate the need for adjusted dosing logic rather than forcing all prescriptions into a single generic weight rule [14]. This design reflects the need for context-aware pediatric decision support rather than a simple maximum-dose checker [5, 24].

Output: Error Probability and Explanation

The model output would be an error-risk estimate paired with a local explanation that shows which features most increased or decreased concern for the individual prescription. A SHAP waterfall-style explanation could identify that the dose-per-kilogram deviation, age-specific maximum-dose rule, ambiguous free-text abbreviation, or route-formulation mismatch drove the alert. The explanation should be translated into concise clinical language so that a pharmacist or prescriber can verify the calculation, inspect the relevant prescription text, and decide whether clarification is required. This presentation aligns with broader clinical decision support principles that emphasize timely, actionable, and interpretable guidance within the care workflow [13].

Identifying Key Drivers of Pediatric Dosing Errors

Global Feature Importance across All Prescriptions

At the global level, SHAP summary plots could show which prescription and patient features most consistently influence pediatric dosing-error predictions. The most influential driver would be expected to be deviation of the prescribed dose from an age- and weight-based recommended range, because pediatric medication safety depends heavily on accurate dose-per-kilogram calculation [2, 16]. Other important predictors could include formulation mismatch, unusual frequency, missing weight, free-text ambiguity, and drug classes known to require careful pediatric dose adjustment [3, 23]. This form of global interpretation would help medication safety committees identify whether the model is primarily responding to clinically meaningful prescribing risks rather than artifacts of documentation.

Patient-Level Explanation for an Individual Alert

For an individual prescription, a local SHAP explanation could show that the alert is driven by a substantial overdose relative to the child's documented weight and by a prescription-text feature such as a trailing zero or unclear frequency. Patient-level explanation is essential because pediatric alert acceptance depends on whether clinicians can quickly see the specific reason an order appears unsafe [7, 26]. A pharmacist reviewing the alert should be able to trace the model's concern to the calculated dose per kilogram, the applicable maximum-dose rule, and the exact prescription phrase that increased risk. This preserves the clinical role of the reviewer while making the model's reasoning transparent.

Table 2 shows how a local SHAP explanation links specific prescription features to alert generation, highlighting factors that influence clinician decision-making.

Table 2. Example of patient-level SHAP explanation for pediatric prescription alerts

Feature Category	Specific Feature Example	Influence on Alert (SHAP Value)	Clinical Relevance
Dose Relative to Weight	Calculated dose exceeds recommended mg/kg	High positive	Indicates potential overdose based on the child's documented weight
Prescription Text Characteristics	Trailing zero in dosage, unclear frequency	Moderate positive	Ambiguous notation that may increase risk of misinterpretation by clinicians
Maximum-Dose Rules	Exceeds institution's max-per-kg guideline	High positive	Confirms alert is aligned with safety thresholds

Model Transparency	Traceable calculation and text mapping	N/A	Allows pharmacist to understand and verify why the alert was triggered

Interaction Effects between Weight, Age, and Drug Class

SHAP dependence plots could be used conceptually to examine whether weight deviation has a different influence across age groups, formulations, or medication classes. In pediatrics, these interactions matter because the same absolute dose may be reasonable for an adolescent but excessive for an infant, and the same drug may have different risk implications depending on formulation or route [14, 24]. Pediatric and neonatal precision-dosing work illustrates that age, maturation, and weight can interact in clinically meaningful ways that static rules may not fully capture [4, 21]. A model designed for dosing-error prediction should therefore make interaction effects inspectable rather than hiding them inside a composite risk score.

Counterfactual Explanations for Corrective Actions

Counterfactual explanations could support corrective action by showing how the alert would change if the prescribed dose were adjusted into the recommended range, the frequency were clarified, or the ambiguous abbreviation were removed. Such explanations are useful because they translate prediction into a clinically actionable path without implying that the model has final authority over the order [13, 22]. For example, the system could indicate that the main concern would be resolved if the dose-per-kilogram value matched the local pediatric dosing rule while the route and formulation remained unchanged. This approach aligns explainability with medication safety improvement by connecting the alert to a concrete prescription element that can be verified and corrected. **Figure 2** illustrates how counterfactual explanations can translate medication-safety alerts into specific prescription elements that clinicians can verify and correct without transferring final decision authority to the model.

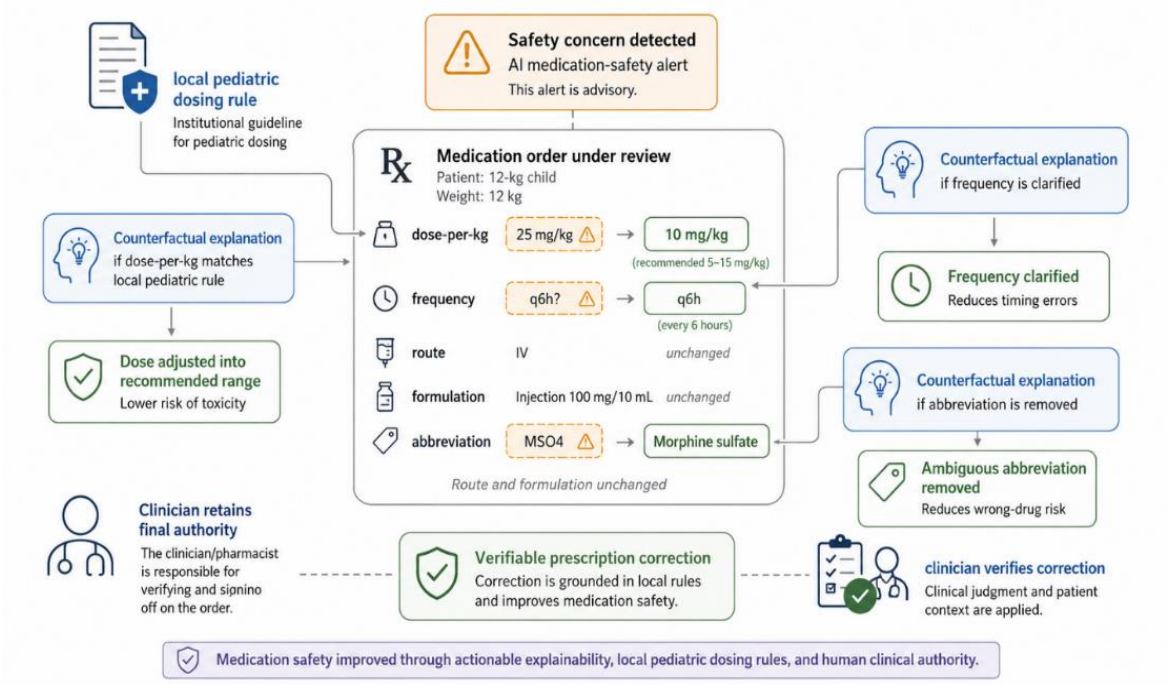

Figure 2. Counterfactual explainability for corrective medication-order safety actions.

Explainability Methods for Pharmacists and Prescribers

Alert Presentation with Inline Explanation

The alert should present the predicted error category together with a short natural-language explanation embedded directly in the computerized ordering workflow. Pediatric studies of medication alerts show that response depends on whether the alert is understandable, clinically relevant, and delivered at the point where action is possible [7]. An inline explanation might state that the dose exceeds the expected weight-based range, the frequency is inconsistent with the selected formulation, or the sig text contains a potentially unsafe abbreviation. This design would make the explanation part of the clinical review process rather than a separate technical output.

User-Configurable Alert Thresholds

User-configurable alert thresholds could allow pharmacists and prescribers to adjust how aggressively the model flags potential errors in different clinical contexts. This flexibility is important because alert burden and override behavior are persistent limitations of computerized decision support, particularly when alerts are too broad or poorly tuned to local workflows [5, 6]. A pediatric intensive care unit may prefer a different alerting posture from an outpatient setting, but the explanation should

remain available regardless of the selected threshold. Threshold adjustment should therefore modify when the model interrupts clinicians, not whether the model remains accountable for its reasoning.

Audit Trail and Regulatory Documentation

Every model-generated alert, explanation, prescriber response, pharmacist intervention, and override rationale should be logged as part of a medication-safety audit trail. Clinical decision support systems require governance because their benefits depend on ongoing monitoring, transparency, and the ability to investigate unintended consequences [13]. Explainable AI literature similarly emphasizes that trustworthy clinical systems should support reviewability, documentation, and accountability across model design, deployment, and use [12]. For pediatric dosing safety, an audit trail would help determine whether explanations are clinically meaningful and whether alert logic remains aligned with current medication practice.

Feedback Loop for Continuous Improvement

A feedback loop could capture pharmacist clarifications, prescriber edits, accepted alerts, and justified overrides to refine the model and improve local relevance. Neonatal and pediatric medication-error detection studies show that clinical context and expert review are central to distinguishing true errors from acceptable deviations [18, 19]. Feedback should be used carefully so that the model learns from validated clinical decisions rather than simply imitating override behavior. Maintaining explainability during refinement is essential, because a model that becomes less interpretable over time would undermine the trust gained from transparent alerting.

Integration Into Pediatric Clinical Workflow

Real-Time Alerting in Pediatric CPOE

The model would be embedded in pediatric computerized provider order entry so that risk assessment occurs while the prescription is being composed or signed. Prior pediatric decision support tools demonstrate that integrated dose calculation can support safer prescribing when clinicians receive guidance within the ordering environment [16, 17]. The initial warning should be concise and non-interruptive when risk is moderate, with an option to expand the alert and inspect the full SHAP explanation. This layered design could preserve workflow efficiency while still allowing detailed review for prescriptions with concerning dose, text, or formulation features.

Pharmacist Verification and Collaboration

Prescriptions with high predicted risk could be routed to a pharmacist verification queue together with the explanation that identifies the suspected error mechanism. Pharmacist review is especially important because medication safety alerts can be affected by missing data, local formulary rules, and clinically intentional deviations that an automated system may not fully understand [15, 26]. The explanation should support collaboration by showing the prescriber and pharmacist the same rationale, such as dose-per-kilogram excess, frequency inconsistency, or unclear free-text sig. This shared view would help transform the alert from a black-box warning into a structured medication-safety conversation.

Evaluation Strategy

Predictive Performance Metrics

Evaluation should compare the explainable model with a traditional rule-based alert baseline using standard diagnostic metrics such as discrimination, calibration, precision-recall behavior, sensitivity, and specificity, without treating any single metric as sufficient for clinical adoption. Pediatric medication-error studies and decision support reviews emphasize that the practical value of an alerting system depends on whether it identifies clinically important errors without overwhelming clinicians [2, 5]. Because this article is conceptual, performance should be described as an intended evaluation domain rather than as completed evidence. The comparison should also examine whether NLP-derived prescription features add meaningful context beyond structured weight-based rules [8, 9]. **Figure 3** presents a restrained conceptual evaluation framework for comparing explainable AI medication-safety alerts with traditional rule-based alert baselines across diagnostic performance, alert burden, clinical relevance, and the added contextual value of NLP-derived prescription features.

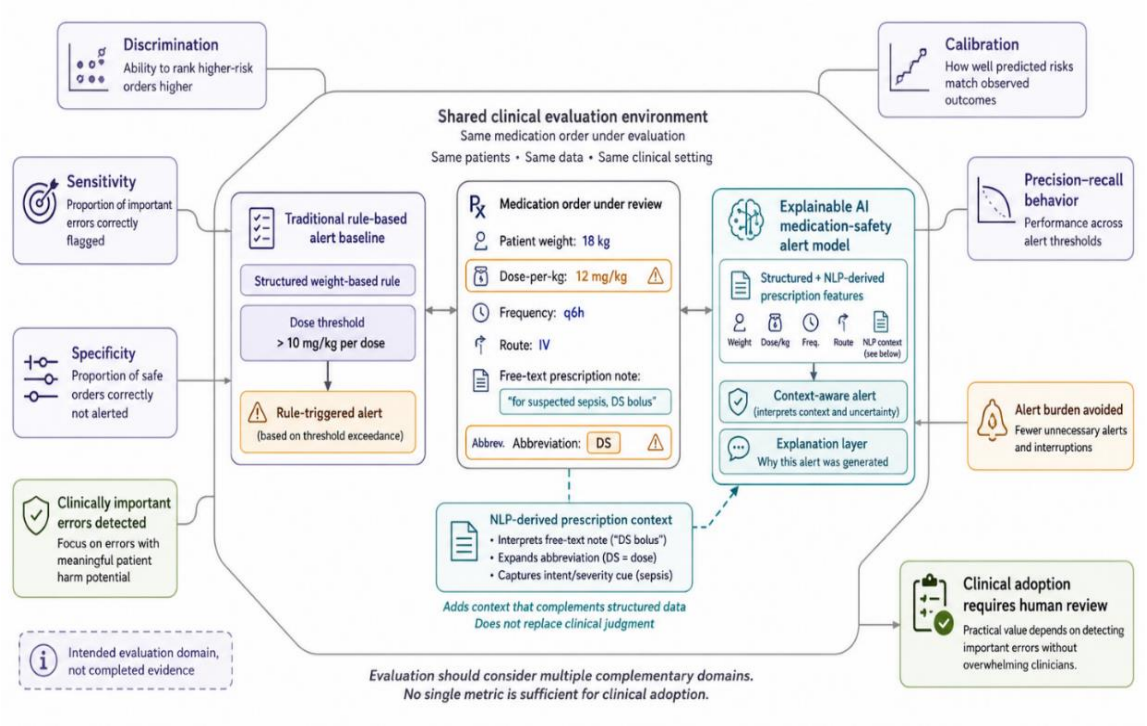

Figure 3. Evaluation domain for explainable ai medication-safety alerts against rule-based baselines.

Explanation Quality and Alert Acceptance

Explanation quality should be evaluated by asking pharmacists and prescribers whether the alert rationale is clear, clinically plausible, actionable, and consistent with the prescription elements they would review manually. Explainable AI research in healthcare highlights that explanations must be evaluated not only technically but also in relation to the needs of the clinical user [11, 12, 22]. Alert acceptance could be studied by comparing explained alerts with otherwise similar unexplained alerts, while recognizing that override rate alone cannot distinguish unsafe dismissal from appropriate clinical judgment. Research on medically inaccurate information and clinical text also underscores the need to verify that language-derived explanations accurately reflect the source prescription rather than introducing misleading interpretations [27].

Impact on Pediatric Medication Safety

The clinical impact of the system should be evaluated through prospective studies that examine intercepted dosing errors, pharmacist interventions, prescribing workflow, and medication-related harm. Pediatric adverse drug event reviews and medication-error analyses provide a foundation for selecting safety outcomes that reflect meaningful patient risk rather than narrow technical performance [1, 15]. Evaluation should also consider whether the model improves detection of error types that static rules handle poorly, such as ambiguous frequency, formulation mismatch, or subtle weight-based deviation. Ultimately, the system should be judged by whether it supports safer pediatric prescribing while maintaining clinician trust and minimizing unnecessary disruption.

Table 3 provides an evaluation and governance framework for determining whether an explainable pediatric dosing-safety model is accurate, interpretable, workflow-compatible, equitable, and clinically accountable.

Table 3. Evaluation and governance framework for an explainable pediatric dosing-safety model

Evaluation or governance domain	Core assessment question	Suggested measures or evidence sources	Failure mode addressed	Interpretation standard	Implementation implication
Predictive discrimination	Can the model distinguish likely dosing errors from clinically acceptable orders?	AUROC, AUPRC, sensitivity, specificity, comparison with rule-based alerts	Model flags too many safe orders or misses true unsafe orders	Performance must exceed static-rule baseline without sacrificing clinically important sensitivity	Determines whether the model adds value beyond conventional dose-range checking
Calibration and risk stratification	Does predicted risk correspond to observed error likelihood?	Calibration plots, risk deciles, observed-versus-expected error rates	High-risk scores are not clinically meaningful	Risk levels should align with pharmacist-reviewed error probability	Supports tiered alerting rather than one-size-fits-all interruption
Added value of NLP features	Do prescription-text features improve safety	Ablation study removing NLP features, comparison	Model ignores ambiguous sig text or over-relies on	Combined model should improve detection of	Justifies investment in prescription-text

	detection beyond structured fields?	of text-only, rule-only, and combined models	structured dose fields	ambiguity-related errors	parsing and traceable NLP
Explanation fidelity	Do SHAP explanations accurately reflect the model's actual drivers?	Local explanation audit, feature perturbation checks, expert review of explanation consistency	Explanation appears plausible but does not represent model behavior	Explanations must correspond to true feature contributions	Prevents decorative or misleading explainability
Clinical explanation usefulness	Do pharmacists and prescribers understand and act on the explanation?	User-rating surveys, think-aloud review, alert comprehension testing, actionability scores	Clinicians cannot determine what to verify or correct	Explanation should identify the exact dose, rule, text phrase, or formulation issue	Improves alert acceptance and targeted medication review
Alert burden and workflow fit	Does the system support safety without excessive interruption?	Alert volume, override rate with rationale, time-to-order completion, pharmacist queue load	Alert fatigue, workflow disruption, inappropriate overrides	Thresholds should balance safety sensitivity with manageable review burden	Guides local threshold tuning and role-specific alert display
Medication-safety impact	Does the system reduce preventable pediatric dosing risk?	Intercepted errors, pharmacist interventions, corrected orders, adverse drug event surveillance	Technical performance fails to translate into patient-safety benefit	Safety outcomes should be clinically meaningful, not only model-centered	Supports prospective validation before broad deployment
Data-quality monitoring	Are weight, height, age, and prescription fields reliable enough for model use?	Missingness rates, stale-weight frequency, unit inconsistency audits, text-parsing error logs	Incorrect explanation caused by poor source data	Data uncertainty should be displayed rather than hidden	Requires source-data safeguards and explicit uncertainty warnings
Bias and subgroup reliability	Does performance remain stable across age groups, weights, units, medication classes, and care settings?	Stratified sensitivity, false-positive and false-negative rates, subgroup calibration	Model underperforms in neonates, infants, children with obesity, or high-risk drug groups	No subgroup should carry disproportionate missed-error or false-alert burden	Requires subgroup validation and local adaptation
Governance and auditability	Can decisions, explanations, overrides, and updates be reviewed?	Audit logs, override rationale, version history, rule-change documentation, model monitoring reports	Untraceable model behavior or unsafe drift after deployment	Every alert and model update should be reviewable by medication-safety governance teams	Establishes accountability for clinical AI decision support
Continuous improvement	Can feedback improve the model without reinforcing unsafe override behavior?	Pharmacist adjudication, accepted-alert analysis, justified-override review, retraining governance	Model learns from noisy or unsafe user behavior	Feedback must be curated and clinically validated before model updating	Enables local refinement while preserving safety and interpretability
Cross-site generalizability	Does the model remain meaningful across hospitals, formularies, and EHR workflows?	External validation, site-specific recalibration, formulary mapping, multi-institutional pilot testing	Poor transfer to different pediatric prescribing environments	Local validation is required before operational use	Prevents premature deployment beyond the development context

Limitations

Reliance on Accurate Weight and Height Data

The model's usefulness depends heavily on accurate, current, and clinically appropriate weight and height data. If weight is missing, outdated, entered in the wrong unit, or inappropriate for the dosing context, the dose-per-kilogram features and resulting explanation may be misleading [3]. Children with obesity introduce additional complexity because actual body weight, ideal body weight, and adjusted dosing approaches may differ across medications and institutions [14]. The system

should therefore flag data-quality uncertainty explicitly rather than presenting a confident explanation based on questionable anthropometric inputs.

Generalizability across Hospitals and Formularies

Generalizability may be limited because pediatric dosing guidelines, formulary options, abbreviation policies, EHR interfaces, and pharmacist workflows vary across hospitals. Clinical decision support reviews emphasize that local implementation context strongly affects whether a medication-safety tool is effective, accepted, and sustained [13]. A model developed in one institution may require local recalibration, rule mapping, formulary adaptation, and explanation review before use elsewhere. Multi-site validation would be necessary to determine whether the model's predictions and explanations remain clinically meaningful across different pediatric care environments.

Conclusion

An explainable AI model for pediatric dosing error prediction could combine prescription-text interpretation, structured patient data, and weight-based dosing rules into a single transparent decision support framework. The model would be designed to identify possible wrong-dose, wrong-frequency, and formulation-related errors while showing the prescription elements that contributed to concern. Rather than replacing clinicians, the system would support earlier recognition and verification of unsafe orders. Its value would lie in making pediatric medication safety alerts more specific, interpretable, and clinically actionable.

The principal strength of this approach is that it links prediction with explanation. By attributing risk to features such as dose-per-kilogram deviation, missing weight, ambiguous sig text, or rule-range mismatch, the system could help clinicians understand why an alert appeared. This transparency could reduce frustration with unexplained warnings and support more focused pharmacist-prescriber communication. The model would also align with existing pharmacy verification workflows by presenting explanations at the point where medication orders are already reviewed.

Important challenges remain before such a system could be responsibly deployed. Data quality, incomplete prescription text, inconsistent dosing conventions, and local formulary differences could all affect model behavior. Cross-site generalizability would require careful evaluation because pediatric medication practice is not uniform across institutions. Prospective validation would also be needed to determine whether the explanations are understandable, useful, and safe in real clinical settings.

Future work should prioritize multi-institutional pilot studies in children's hospitals and pediatric units. These studies should evaluate not only predictive performance but also explanation quality, clinician acceptance, workflow burden, and medication safety impact. A successful system would be one that helps pharmacists and prescribers detect preventable dosing errors while preserving clinical judgment. Explainable AI should therefore be treated as a partner in pediatric medication safety, not as a substitute for expert review.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

References

1. Gates PJ, Meyerson SA, Baysari MT, Lehmann CU, Westbrook JI. Preventable adverse drug events among inpatients: a systematic review. Pediatrics. 2018;142(3):e20180805.
2. Gates PJ, Meyerson SA, Baysari MT, Westbrook JI. The prevalence of dose errors among paediatric patients in hospital wards with and without health information technology: a systematic review and meta-analysis. Drug Saf. 2019;42(1):13-25.
3. Lubsch L, Kimler K, Passerrello N, Parman M, Dunn A, Meyers R. Patient weight should be included on all medication prescriptions. J Pediatr Pharmacol Ther. 2023;28(4):380-1.
4. Vinks AA, Punt NC, Menke F, Kirkendall E, Butler D, Duggan TJ, et al. Electronic health record–embedded decision support platform for morphine precision dosing in neonates. Clin Pharmacol Ther. 2020;107(1):186-94.
5. Ruutiainen H, Holmström AR, Kunnola E, Kuitunen S. Use of Computerized Physician Order Entry with Clinical Decision Support to Prevent Dose Errors in Pediatric Medication Orders: A Systematic Review. Pediatr Drugs. 2024;26(2):127-43.
6. Tolley CL, Forde NE, Coffey KL, Sittig DF, Ash JS, Husband AK, et al. Factors contributing to medication errors made when using computerized order entry in pediatrics: a systematic review. J Am Med Inform Assoc. 2018;25(5):575-84.
7. Dexheimer JW, Kirkendall ES, Kouril M, Hagedorn PA, Minich T, Duan LL, et al. The effects of medication alerts on prescriber response in a pediatric hospital. Appl Clin Inform. 2017;8(2):491-501.

8. Pais C, Liu J, Voigt R, Gupta V, Wade E, Bayati M. Large language models for preventing medication direction errors in online pharmacies. Nat Med. 2024;30(6):1574-82.
9. Zhang L, Lagerberg T, Chen Q, Ghirardi L, D'Onofrio BM, Larsson H, et al. Prediction of treatment dosage and duration from free-text prescriptions: an application to ADHD medications in the Swedish prescribed drug register. Evid Based Ment Health. 2021;24(4).
10. Ouzar Y, Ajmi F, Othman SB, Rousseliere C, Decaudin B, Odou P, et al. Interpretable one-class classification framework for prescription error detection using BERT embeddings and dimensionality reduction. Comput Biol Med. 2025;196:110775.
11. Yang CC. Explainable artificial intelligence for predictive modeling in healthcare. J Healthc Inform Res. 2022;6(2):228-39.
12. Markus AF, Kors JA, Rijnbeek PR. The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inform. 2021;113:103655.
13. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med. 2020;3(1):17.
14. Cloyd CP, Macedone D, Merandi J, Pierson S, Wcislo MS, Lutmer J, et al. A Quality Initiative to Improve Appropriate Medication Dosing in Pediatric Patients with Obesity. Pediatr Qual Saf. 2024;9(3):e741.
15. D'Errico S, Zanon M, Radaelli D, Padovano M, Santurro A, Scopetti M, et al. Medication errors in pediatrics: proposals to improve the quality and safety of care through clinical risk management. Front Med. 2022;8:814100.
16. Higi L, Käser K, Wälti M, Grotzer M, Vonbach P. Description of a clinical decision support tool with integrated dose calculator for paediatrics. Eur J Pediatr. 2022;181(2):679-89.
17. Higi L, Schmitt R, Käser K, Wälti M, Grotzer M, Vonbach P. Impact of a clinical decision support system on paediatric drug dose prescribing: a randomised within-subject simulation trial. BMJ Paediatr Open. 2023;7(1):e001726.
18. Yalçın N, Kaşıkcı M, Çelik HT, Allegaert K, Demirkan K, Yiğit Ş, et al. An artificial intelligence approach to support detection of neonatal adverse drug reactions based on severity and probability scores: a new risk score as web-tool. Children. 2022;9(12):1826.
19. Yalçın N, Kaşıkcı M, Çelik HT, Allegaert K, Demirkan K, Yiğit Ş, et al. Development and validation of a machine learning-based detection system to improve precision screening for medication errors in the neonatal intensive care unit. Front Pharmacol. 2023;14:1151560.
20. Henry Basil J, Lim WH, Syed Ahmad SM, Menon Premakumar C, Mohd Tahir NA, Mhd Ali A, et al. Machine learning-based risk prediction model for medication administration errors in neonatal intensive care units: A prospective direct observational study. Digit Health. 2024;10:20552076241286434.
21. Yin M, Jiang Y, Yuan Y, Li C, Gao Q, Lu H, et al. Optimizing vancomycin dosing in pediatrics: a machine learning approach to predict trough concentrations in children under four years of age. Int J Clin Pharm. 2024;46(5):1134-42.
22. Belle V, Papantonis I. Principles and practice of explainable machine learning. Front Big Data. 2021;4:688969.
23. Campbell CT, Wheatley KH, Svoboda L, Campbell CE, Norris KR. Strategies for implementing pediatric dose standardization: considerations from the Vizient University health system consortium pharmacy network pediatric pharmacy committee. J Pediatr Pharmacol Ther. 2022;27(1):19-28.
24. Rao A, Palma J. Clinical decision support in the neonatal ICU. Semin Fetal Neonatal Med. 2022;27(5):101332.
25. Bertoni CB, Prusakov P, Merandi J, Bartman T. Clinical decision support to improve dosing weight use in infants with neonatal abstinence syndrome. Pediatr Qual Saf. 2019;4(4):e184.
26. Hashemi F, van Gelder TG, Bollen CW, Liem YT, Egberts TC. The effect of a decision support system on the incidence of prescription errors in a PICU. J Clin Pharm Ther. 2022;47(3):330-44.
27. Sun Z, Yim WW, Uzuner Ö, Xia F, Yetisgen M. A scoping review of natural language processing in addressing medically inaccurate information: Errors, misinformation, and hallucination. J Biomed Inform. 2025;104866.