Medicina Katastrof

(ISSN:2070-1004)
(E-ISSN:2686-7966)
(Publishing Date: 19/12/2025)

Health Professions **Open Access**

Artificial Intelligence in Medication Safety Systems for the Prevention of Perioperative and Inpatient Medication Errors

Mohammed Hudaiban Al-Rashidi[1], Sami Mohammed Alhazemi[2], Mustafa Jaffer Hyyan[3], Mohammed Khalaf Mashan Alharbi[4], Naemah Rashed Bin Reshedan[5], Haidar Mohammed Alsaygeh[6], Nouf Shafi Alshammari[7]

Background:

Medication errors are a major source of preventable harm in perioperative and inpatient care, and artificial intelligence (AI) is increasingly used to strengthen medication safety systems.

Methods:

PubMed was searched from inception to September 2025 for English-language human studies evaluating AI-enabled medication safety interventions in hospital perioperative or inpatient settings. Two reviewers screened records, extracted data in duplicate, and appraised risk of bias using Joanna Briggs Institute checklists; results were synthesized narratively without meta-analysis.

Results:

Eleven studies, predominantly retrospective cohorts or pre–post implementations, were included across adult wards, intensive care, pediatric/neonatal intensive care, and intravenous administration monitoring, with sample sizes ranging from 311 ICU prescribing episodes to 3,481,634 alert events. Discrimination for identifying high-risk prescriptions or patients was moderate to high (AUROC 0.74–0.97). Improvements were reported in key process measures, including reduced pediatric ICU dosing deviations (RR 0.21; 95% CI 0.05–0.96), higher reconciliation discrepancy yield with admission prioritization (45% vs 21%; RR 2.13; 95% CI 1.40–3.24), and lower alert burden (54.1% reduction at sensitivity 0.99; precision 0.192); one deployment triggered alerts for 0.4% of orders and 43% prompted order changes.

Conclusions:

AI-enabled medication safety systems were associated with improved identification and management of high-risk medication scenarios and more efficient safety surveillance. Evidence remained heterogeneous, and preventable adverse drug events and longer-term patient outcomes were infrequently measured.

Keywords:

Medication Errors, Patient Safety, Artificial Intelligence, Machine Learning, Clinical Decision Support Systems, Perioperative Care

Author details:
[1] *Nursing Technician, Al-Nairiyah General Hospital, Al-Nairiyah, Saudi Arabia.*
[2] *Nursing Technician, Al-Nairiyah General Hospital, Al-Nairiyah, Saudi Arabia.*
[3] *Dental Hygienist, Al-Nairiyah General Hospital, Al-Nairiyah, Saudi Arabia.*
[4] *Anesthesia Technician, Maternity Hospital, Hafar Al-Batin, Saudi Arabia.*
[5] *Anesthesia Technologist, Al-Nairiyah General Hospital, Al-Nairiyah, Saudi Arabia.*
[6] *Laboratory Technician, Al-Nairiyah General Hospital, Al-Nairiyah, Saudi Arabia.*
[7] *Medical Sterilization Technician, Al-Nairiyah General Hospital, Al-Nairiyah, Saudi Arabia.*

Introduction

Medication errors can be conceptualized as preventable failures anywhere along the medication-use process (e.g., prescribing, order communication/transcription, dispensing, administration, and monitoring) that may lead to inappropriate medication use and/or patient harm. Across healthcare systems, observational meta-analytic synthesis has estimated that preventable medication harm occurs in roughly 3% of patient-incidence records (95% CI 2%–4%), while overall medication harm occurs in 9% (95% CI 7%–11%), with substantial heterogeneity across settings and study designs [1]. Importantly for the present protocol, the same meta-analysis observed higher rates of preventable medication harm in intensive care (7%, 95% CI 4%–12%) and highly specialized or surgical care (6%, 95% CI 3%–11%) compared with many other clinical contexts, and it identified prescribing and monitoring as prominent sources of the highest prevalence rates of preventable harm [1].

These findings motivate a systems-oriented lens: "medication safety systems" in modern hospitals increasingly rely on layered safeguards, workflow checks, standardization, and digital infrastructure, to reduce the probability that a single human lapse propagates into patient harm [1]. While medication errors are frequently described as ubiquitous but "mostly low harm" events, high-quality national modeling illustrates that even modest per-event harm probabilities translate into major aggregate burden. In England, an economic analysis estimated 237 million medication errors annually across the medication process, with 38.4% occurring in primary care; 72% had little or no potential for harm, yet approximately 66 million were considered potentially clinically significant [2]. The same analysis estimated that definitely avoidable adverse drug events cost the National Health Service £98,462,582 per year, consumed 181,626 bed days, and caused or contributed to 1,708 deaths (including both primary-care adverse drug events leading to admission and secondary-care events leading to longer hospital stay) [2]. These quantitative estimates support the rationale for prevention strategies that are not limited to "rare catastrophic" events, but instead strengthen system resilience against common, workflow-driven failures that can scale into large societal impact [2]. Evidence synthesized specifically for high-acuity inpatient environments also reinforces that perioperative and critical-care care pathways have distinctive error mechanisms and measurement challenges. A systematic review focusing on operating rooms and intensive care units reported an operating theatre medication error rate range of 7.3%–12% and an intensive care unit medication error rate range of 1.32%–31.7%, derived from studies published between 2017 and 2023 [3].

In Saudi Arabia, national database studies provide complementary process-stage insight at scale: a nationwide observational analysis of reports submitted to a Ministry of Health pharmaceutical-care database (March 2018–June 2019) identified 71,332 medication error events, with 84.8% occurring during prescribing, 5.8% during transcribing, and 5.7% during dispensing; 5.8% of reported errors reached the patient, and reporters associated 31.6% of errors with work overload and 22.7% with lack of experience [4]. More recent analysis of reported medication errors from January 2023 to December 2024 across nine Ministry of Health hospitals evaluated 19,645 errors, reporting that 69.1% were intercepted before reaching patients, that prescribing constituted 94.8% of events, and that pharmacists reported more than 90% of cases [5]. Although individual drug-class analyses are not directly generalizable to all inpatient medications, national database work on clozapine-related medication caused

errors further illustrates that most reported events may not reach patients (92.3% did not reach the patient) and that harm is uncommon among those that do (reported harm 0.3% in this dataset), reinforcing both the value and the limitations of incident-reporting signals for estimating "true incidence." [6]. Taken together, contemporary quantitative evidence indicates that medication error prevention is a high-yield patient safety target globally and within Saudi Arabia, but the burden manifests differently depending on denominators, detection methods, and care context. Globally, preventable medication harm prevalence estimates (3%, 95% CI 2%–4%) and setting-specific peaks in intensive care and surgical/specialized care (e.g., intensive care 7%, 95% CI 4%–12%; specialized/surgical care 6%, 95% CI 3%–11%) support prioritization of inpatient and perioperative pipelines where exposure to high-alert intravenous agents and rapid titration is common [1].

In parallel, system-level modeling in England translates medication errors into concrete resource and mortality impacts (e.g., £98.5 million/year for definitely avoidable adverse drug events and 1,708 deaths/year in the modeled framework), making the case that even incremental prevention at population level can have meaningful downstream benefits [2]. Within Saudi Arabia, the sheer scale of reported errors (71,332 events in a 16-month national reporting window, with most originating at prescribing and <10% reaching the patient) suggests that prevention strategies should be designed both to reduce upstream error generation and to strengthen downstream interception reliability [4]. More recent multi-hospital analysis (19,645 reported errors across 2023–2024, with 69.1% intercepted before reaching the patient and prescribing representing 94.8%) highlights a contemporary pattern in which prescribing-stage vulnerability remains dominant, while interception is frequent but not universal [5].

Specialized analyses of high-risk medications (e.g., clozapine) indicate similar stage concentration and high interception proportions, but also point to policy, process standardization, and reporting culture as recurring contributors that are actionable targets for informatics-enabled safeguards [6]. Perioperative medication safety has distinct human-factors stressors, time pressure, parallel tasking, rapid physiological consequences, and frequent use of look-alike/sound-alike (LASA) preparations, that can place exceptional reliance on local workflow design and team experience. In Saudi Arabia, a cross-sectional survey of anesthesia clinicians reported that 69% had experienced an anesthetic drug error at least once in their career, with "haste" and "heavy workload" each reported by 60.3% as primary causal factors, and with less experience associated with more committed errors; respondents also indicated that fear of medicolegal issues was a major barrier to reporting [7]. Large retrospective perioperative data from Japan similarly emphasizes provider-level experience as a measurable risk factor: intraoperative medication errors occurred in 102 of 100,093 procedures (0.10%), and compared with attending anesthesiologists, the adjusted odds of medication error were higher when care involved residents (odds ratio [OR]. 2.713, 95% CI 1.283–6.815) or interns (OR 3.272, 95% CI 1.508–8.368) [8]. Importantly, the target condition for this protocol, perioperative and inpatient medication errors, spans both "rare but high-severity" events that nonetheless generate downstream costs [2].

From a prevention-technology standpoint, operating-room clinical decision support is increasingly recognized as a credible intervention layer: in a retrospective cross-sectional review of self-reported intraoperative medication errors, 76 of 80 (95%) were classified as preventable by clinical decision support algorithms, with wrong-medication and wrong-dose errors among those most consistently rated preventable, while inadvertent bolus errors were least preventable by this approach [9]. These findings collectively support a hybrid prevention model in which technology is not treated as a substitute for expertise, but as an adaptive barrier that strengthens weak points in human memory, attention, and communication, especially during high workload and high stakes [2,7–9]. However, current evidence indicates important translational gaps.

A perioperative digital-health viewpoint highlights that AI-enabled decision support and health information technology tools (e.g., computerized provider order entry, electronic medication administration records, barcode medication administration) are plausible mechanisms to reduce perioperative medication errors, but it also emphasizes measurement and implementation challenges specific to the perioperative domain [10]. In the inpatient prescribing context, a scoping review of AI used to optimize medication alerts found only 10 quantitative studies; positive predictive value ranged from 9% to 100%, only 30% reported both statistical performance and clinical outcomes, only two studies implemented AI-based alerts in hospital current

practice, and none performed external validation [11]. A large retrospective study of an implemented machine learning–based clinical decision support system (MedGuard) reported 1,206,895 prescriptions with a 2.36% alert rate (28,536 alerts), 48.88% alert acceptance, and identification of 470 intercepted errors (1.64% of alerts; 16.4 intercepted errors per 1,000 alerted orders), suggesting a potentially favorable balance between alert volume and actionable yield, but generalizability remains uncertain [12]. Meta-analytic work on determinants of medication-related alert handling further indicates that acceptance behavior is systematically associated with contextual factors (e.g., fellows vs residents OR 1.14, 95% CI 1.09–1.20; weekday vs weekend OR 1.25, 95% CI 1.11–1.40), underscoring that "human-in-the-loop" behavior is not random noise but a measurable system component that AI optimization efforts must address [13].

Beyond alerts, machine learning has been deployed to prioritize admission medication reconciliation: a predictive tool trained on 7,200 patients achieved an area under the receiver operating characteristic curve of 0.74 and, in retrospective "real-life" simulation, identified 45% of selected patients as having at least one unintended discrepancy versus 21% using an existing tool (reported improvement 113%) [14]. Finally, natural language processing and machine learning applied to unstructured electronic health record text have shown promise for detecting under-reported adverse drug events and safety signals, but evidence remains limited and heterogeneous, with relatively few studies and variable evaluation methods [15]. Collectively, these findings suggest that the field is advancing, but still lacks consistent external validation, perioperative-specific evaluation, and clear linkage between AI-enabled medication safety systems and patient-important outcomes across inpatient and perioperative settings [10–16]. To systematically review and synthesize evidence on artificial intelligence–enabled medication safety systems that prevent perioperative and inpatient medication errors, evaluating their designs, implementation contexts, effectiveness on error-related outcomes, and limitations in validation and generalizability.

Methods

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 in statement, with methods specified a priori but without protocol registration (PRISMA-Item-24). The review question addressed whether artificial intelligence–enabled medication safety systems reduced or prevented medication errors in perioperative and inpatient hospital care. Eligible studies enrolled human participants receiving perioperative care (preoperative, intraoperative, or immediate postoperative phases) and/or inpatient care (medical/surgical wards, intensive care units, emergency department admissions leading to hospitalization) and evaluated an artificial intelligence (AI) component embedded in a medication-use process (prescribing, transcribing, dispensing, administration, monitoring) or in a medication safety platform (e.g., decision support, surveillance, smart infusion oversight).

We included comparative and non-comparative primary research designs relevant to implementation in real-world clinical workflows (randomized trials, quasi-experimental studies, interrupted time series, cohort studies, cross-sectional evaluations, and prospective/retrospective system validations performed in clinical settings). We excluded editorials, narrative reviews, purely simulated datasets without clinical deployment, conference abstracts lacking sufficient methods/results for appraisal, and studies focused exclusively on outpatient/community pharmacy settings. Primary outcomes were medication error occurrence, interception/detection rates, and preventable adverse drug events; secondary outcomes included alert performance metrics (e.g., sensitivity, specificity, positive predictive value), workflow impact, and user-acceptance measures.

The primary database searched was PubMed, from inception to 30 September 2025 (PRISMA-Item-7). Searches were performed using a combination of Medical Subject Headings (MeSH) and free-text terms for (1) AI methods, (2) medication errors and safety, and (3) perioperative/inpatient hospital settings, with limits applied to English language and humans. The exact PubMed search string was: (("Artificial Intelligence"[Mesh]. OR "Machine Learning"[Mesh]. OR "Natural Language Processing"[Mesh]. OR "Clinical Decision Support Systems"[Mesh]. OR artificial intelligence[tiab]. OR machine learning[tiab]. OR deep learning[tiab]. OR neural network*[tiab]. OR natural language processing[tiab]. OR large language model*[tiab]. OR algorithm*[tiab]. OR predictive model*[tiab]) AND ("Medication Errors"[Mesh]. OR "Drug-Related Side Effects and Adverse Reactions"[Mesh]. OR medication error*[tiab]. OR prescribing error*[tiab]. OR administration error*[tiab]. OR dispensing error* [tiab]. OR adverse drug event* [tiab].

OR medication safety[tiab]) AND ("Perioperative Care"[Mesh]. OR "Surgical Procedures, Operative"[Mesh]. OR anesthesia[tiab]. OR perioperative[tiab]. OR intraoperative[tiab]. OR postoperative[tiab]. OR "Inpatients"[Mesh]. OR inpatient*[tiab]. OR hospital*[tiab]. OR intensive care[tiab]. OR ICU[tiab]) AND ("Medication Systems, Hospital"[Mesh]. OR "Computerized Provider Order Entry"[Mesh]. OR "Barcode Technology"[Mesh]. OR smart pump*[tiab]. OR infusion pump*[tiab]. OR medication system*[tiab]. OR order entry[tiab]. OR e-prescrib*[tiab]. OR surveillance[tiab]. OR monitoring system*[tiab])) AND (english[lang]) AND (humans[MeSH Terms]) AND ("1900/01/01"[Date - Publication]. : "2025/09/30"[Date-Publication]). Reference lists of included studies were also screened for additional eligible records (backward citation searching), and Scopus was searched as a secondary database using a conceptually aligned strategy.

All records retrieved were exported from PubMed and imported into a reference manager for duplicate removal (PRISMA-Item-8). Duplicates were identified using automated matching (title, authors, year, journal) followed by manual confirmation for ambiguous cases. Two reviewers independently screened titles and abstracts against eligibility criteria using a structured screening form, after completing a calibration exercise on a pilot set of records to harmonize interpretation of inclusion/exclusion rules. Full texts were retrieved for potentially eligible records, and the same two reviewers independently assessed full-text eligibility, documenting exclusion reasons in a standardized log to support PRISMA flow reporting (PRISMA-Item-16). Discrepancies at either stage were resolved by discussion; if consensus could not be reached, a third reviewer adjudicated. Where full texts could not be retrieved through institutional access, authors were contacted.

Data were extracted using a standardized electronic extraction form developed for AI-enabled medication safety interventions (PRISMA-Item-9). The form was piloted and refined before full extraction. Two reviewers performed double data extraction independently, capturing: study setting (country, hospital type, unit), population (age group, surgical vs medical admissions), clinical context (perioperative phase and/or inpatient service line), AI approach (rule-based vs machine learning vs deep learning vs natural language processing vs large language model–based methods), data inputs (electronic health record, computerized provider order entry logs, medication administration records, smart pump telemetry, pharmacy dispensing systems), comparator (usual care or non-AI clinical decision support), implementation characteristics (integration point, alerting logic, user workflow), and outcomes (medication error definitions, detection/interception rates, preventable adverse drug events, performance metrics, and process measures). When studies reported multiple timepoints or units, all eligible outcome measurements were extracted with corresponding denominators and time windows. Conflicts in extracted data were reconciled by consensus using the full-text source as the arbiter; persistent disagreements were resolved by a third reviewer. Where critical numerical details were missing (e.g., denominators for error rates), attempts were made to derive values from the report; if derivation was not possible, the item was recorded as not reported. Risk of bias was assessed at the individual study level using Joanna Briggs Institute (JBI) Critical Appraisal Checklists matched to study design.

Two reviewers independently appraised each study after a calibration exercise to standardize domain interpretation. Each checklist item was rated as “Yes,” “No,” “Unclear,” or “Not applicable,” and disagreements were resolved by discussion with third-reviewer adjudication if required. Overall risk-of-bias judgments were derived using a rule-based approach: low risk when key domains for the relevant design were met with no critical failures, moderate risk when one or more key domains were rated “Unclear” without critical failures, and high risk when one or more critical domains were rated “No.” No study was excluded solely on the basis of risk-of-bias ratings; instead, these judgments informed the strength and caution of narrative conclusions. Because of expected clinical and methodological heterogeneity across AI approaches, implementation contexts, outcome definitions, and study designs, quantitative meta-analysis was not performed and no statistical heterogeneity metrics were calculated (PRISMA-Item-13).

Findings were synthesized narratively using a structured framework that grouped studies by (1) care context (perioperative vs general inpatient vs intensive care), (2) medication-use process targeted (prescribing/order entry, pharmacy verification/dispensing, administration including smart infusion oversight, and monitoring/surveillance), and (3) AI method family (machine learning/deep learning predictive models, natural language processing for text-based signals, and hybrid systems combining rules with learned models). Within each subgroup, results were summarized by direction and consistency of effect on medication error reduction or interception, supported by extracted effect measures (rates, proportions, and diagnostic performance

metrics) reported in the original studies. Where studies reported multiple endpoints, priority was given to clinically anchored outcomes (medication errors and preventable adverse drug events) over intermediate alert metrics, while still documenting alert performance to contextualize feasibility and burden. Variation across studies was handled qualitatively by explicitly describing differences in populations, clinical workflow integration, baseline safety infrastructure, and outcome measurement; when discrepant findings occurred, plausible explanatory factors (e.g., alert fatigue, data drift, implementation maturity, or differing error definitions) were explored without pooling estimates. Risk-of-bias ratings were used to temper inferences within each subgroup, and conclusions emphasized patterns supported by multiple low-to-moderate risk evaluations.

Results

A PubMed search (inception–30 September 2025) identified 4,277 records, of which 1,003 duplicates were removed, leaving 3,274 unique records for title/abstract screening . After exclusion of 3,150 records at screening, 124 full texts were assessed for eligibility and 113 were excluded for non-hospital settings, non-AI/non-deployable interventions, simulation-only evaluations, or insufficient outcome reporting . Eleven studies met inclusion criteria and were included in the narrative synthesis, comprising inpatient ward, intensive care unit (ICU), pediatric ICU (PICU), neonatal ICU (NICU), and intravenous (IV) administration safety contexts. The included studies were predominantly observational cohorts, pre–post implementations, and retrospective validations embedded in clinical workflows, with no eligible randomized head-to-head trials comparing AI versus non-AI medication safety systems.

Sample sizes and units of analysis varied substantially, spanning 311 ICU albumin prescription requests in a quasi-experimental pre–post design [17], 9,342 PICU prescriptions across two implementation phases [16], 7,200 hospital admissions used for model development and simulated deployment [18], 5,017 admissions to inpatient wards with adjudicated adverse drug events (ADEs) [19], and 3,481,634 medication alerts analyzed for machine-learning–based filtering [14]. The evidence base was geographically diverse, including inpatient implementations and validations from Israel [12], France [13,18], the United States [14], Taiwan [15], the Netherlands (PICU) [16], Iran (ICU) [17], Japan (NICU) [20], and the Republic of Korea (IV infusion monitoring) [21], alongside unsupervised outlier-detection work applied to large-scale prescribing data [22]. Follow-up periods were typically defined by hospital implementation windows or retrospective extraction intervals, ranging from several months pre–post [16,17]. to multi-year retrospective cohorts [14,18,19]. Across studies, the most frequently reported clinically anchored outcome was prevention or interception of medication errors at the point of prescribing and/or pharmacist verification (primary outcome for this review), operationalized as (i) low-burden alerting with clinically validated relevance and downstream order modification, and/or (ii) reductions in measurable prescribing deviations requiring correction. In an inpatient internal medicine deployment of a probabilistic machine-learning clinical decision support system (CDSS), alerts were generated for 0.4% of all medication orders.

With 85% judged clinically valid and 43% prompting changes in subsequent medical orders, indicating meaningful interception of high-risk prescribing situations with low alert burden [12]. In a PICU pre–post evaluation of a dosing decision-support system, unintentional protocol deviations requiring adjustment fell from 0.22% of prescriptions (11/5,034) before implementation to 0.05% (2/4,308) after implementation; unjustified deviations similarly declined from 0.56% to 0.23% [16]. In an ICU pre–post study evaluating a medication decision support system for albumin prescribing, guideline adherence increased from 47.64% to 68.26% (P = 0.014), and 60.15% of alerts resulted in prescription modification; a patient-safety composite improved from 63.33% to 82.61% (P = 0.009), supporting measurable improvements in prescribing quality aligned with medication safety objectives [17].

In addition, a hospital admission prioritization tool for medication reconciliation improved yield of identifying patients with at least one unintended discrepancy, from 21% using the existing approach to 45% using the machine-learning tool in a simulated real-life evaluation, representing a reported 113% relative improvement in targeting high-risk admissions for pharmacist review [18]. A second commonly addressed outcome domain was predictive performance for identifying high-risk prescriptions or high-risk patients (often reported as area under the receiver operating characteristic curve [AUROC]. and/or precision–recall metrics). A hybrid machine-learning CDSS designed to prioritize prescription checks in hospitalized patients was totally

developed and evaluated on 10,716 patients (16,270 prescriptions), achieving an AUROC of 0.81 and an average precision of 0.75 for identifying prescriptions at high risk of medication error [13]. The admission-based medication error risk tool trained on 7,200 reconciled admissions reported recall of 0.75, precision of 0.65, F1 score of 0.70, AUROC of 0.74, and area under the precision–recall curve (AUPRC) of 0.75, consistent with moderate discriminative ability in a clinically relevant, pharmacist-facing prioritization use case [18]. For inpatient wards, a prediction model for in-hospital ADE risk at admission (5,017 admissions; 488 ADEs) achieved AUROC 0.752 with favorable calibration (calibration slope 0.997) and a Brier score of 0.056, supporting feasibility of early risk stratification as part of medication safety surveillance and targeted review [19].

In NICU settings, a machine-learning medication error detection system achieved high diagnostic performance (AUROC 0.97; sensitivity 0.97; specificity 0.90; accuracy 0.91; F1 score 0.92), indicating strong discriminative performance for error detection in high-complexity neonatal prescribing and administration environments [20]. A third outcome domain frequently quantified across studies was alert burden, alert filtering potential, and/or clinically meaningful alert acceptance, reflecting the implementation tension between sensitivity for harm prevention and usability constraints such as alert fatigue. In a large-scale evaluation of medication alerts, a gradient-boosting model (LightGBM) evaluated 3,481,634 alerts while maintaining sensitivity fixed at 0.99; precision reached 0.192 and the approach was estimated to reduce alert volume by 54.1%, illustrating potential to remove low-value alerts while preserving a low false-negative rate threshold [14].

Complementary work modeling clinician response to drug-related computerized alerts reported strong discrimination for predicting physician response (AUROC 0.916) with a positive predictive value of 0.871 and specificity of 0.833 across 3,885 alerts, supporting feasibility of response-aware alert personalization as a medication safety strategy (particularly where acceptance is necessary for error interception) [15]. Notably, the inpatient deployment with a very low alert rate (0.4% of orders) simultaneously achieved high clinical validity (85%) and substantial order-change impact (43%), suggesting that prioritization and targeted alerting could mitigate alert fatigue while preserving safety impact [12]. Between-study differences that plausibly explained divergent findings were primarily attributable to (i) heterogeneity in the medication-use process targeted (prescribing/order entry vs reconciliation vs administration monitoring), (ii) differing definitions and denominators for "medication error" (protocol deviations, unintended discrepancies, ADEs, alert-triggered potential errors), and (iii) different operating points chosen for model deployment (high sensitivity thresholds vs higher precision/low-burden thresholds). For example, the PICU study evaluated a low-incidence baseline environment (protocol deviations <1%), where absolute reductions were small but clinically meaningful, whereas large-scale alert filtering work optimized precision under a fixed sensitivity (0.99), prioritizing safety at the cost of relatively low precision (0.192) in exchange for major alert-volume reduction [14,16].

Hospital admission risk tools and ADE prediction models relied on retrospective electronic health record (EHR) features and were evaluated by predictive metrics and yield-of-review rather than direct error-rate reduction, which limited direct comparability to pre–post prescribing error studies that quantified actionable deviations and modifications [18,19]. Differences in clinician workflow integration also likely contributed: systems embedded at pharmacist verification or admission medication reconciliation emphasized prioritization and yield, while prescribing-time alerting and administration monitoring emphasized real-time interception and mismatch detection [12,18,21]. Secondary outcomes were variably reported and included workflow timing, system responsiveness to changing patient status, and technology performance for administration-phase safety. In the inpatient CDSS deployment, 60% of alerts were triggered after medication dispensing due to changes in patient status.

This is highlighting a clinically relevant monitoring function beyond static order review [12]. In the ICU albumin MDSS evaluation, alert responsiveness varied by indication, implying context-dependent clinician acceptance and the need for indication-specific tuning to reduce inappropriate overrides [17]. For IV administration safety, a multimodal infusion pump monitoring approach using convolutional neural network–based deep learning demonstrated high technical performance: training/validation/test accuracies of 98.3%, 97.7%, and 98.5%, respectively, and infusion-rate estimation percentage errors of 0.22 – 2.90% in evaluation experiments, supporting feasibility of real-time mismatch detection between pump settings and actual infusion with exact states/prescriptions [21].

Finally, unsupervised outlier detection applied to medication orders demonstrated feasibility of identifying potential prescribing errors at scale (563,960 medication items), although the reporting emphasized detection capability rather than downstream clinical outcomes, limiting inference about real-world error prevention impact [22]. Overall, the synthesized evidence indicated that AI-enabled (or algorithmically enhanced) medication safety systems deployed in inpatient and perioperative-adjacent hospital settings were consistently capable of prioritizing high-risk prescriptions/patients, reducing selected categories of prescribing deviations, and/or lowering alert burden through filtering and response-aware strategies [12–22].

However, results were not directly comparable across studies due to substantial heterogeneity in target processes, definitions, denominators, deployment thresholds, and outcome measurement. These findings set up the Discussion by highlighting the trade-offs between sensitivity, precision, alert burden, and clinical integration, and by indicating where future evaluations should prioritize standardized outcome definitions, prospective impact assessment on preventable ADEs, and implementation science measures alongside predictive accuracy.

Discussion

Across the 11 included studies, artificial intelligence–enabled medication safety systems were most consistently associated with improvements in signal-to-noise at the prescribing or verification interface and with more efficient targeting of clinician attention to higher-risk situations, rather than with direct demonstrations of reduced preventable harm at the patient level. In the inpatient outlier-detection clinical decision support system (CDSS) evaluation, the alert burden remained very low (315 alerts across 78,017 medication orders; 0.4% of prescriptions), while the majority of alerts were clinically valid (85%) and frequently actionable, with 43% prompting medication discontinuation or modification within a median of 1 hour (interquartile range 0.07–4 hours) [12]. This pattern aligned with earlier evidence indicating that computerized provider order entry (CPOE) and conventional decision support interventions often reduced medication error rates but were frequently underpowered to detect reductions in adverse drug events in controlled evaluations [23,25]. The included prescription-check prioritization tool similarly emphasized discriminative capacity (area under the receiver operating characteristic curve 0.81; area under the precision–recall curve 0.75) rather than patient harm endpoints, illustrating the broader trend in which machine learning (ML) was positioned as a "screening amplifier" embedded within pharmacist workflows rather than a standalone preventive intervention [13]. The unsupervised outlier approach tested on 563,000+ prescribed medications reflected the same logic, leveraging distributional abnormality to detect potential dose and frequency anomalies when labeled "ground truth" was scarce [22].

When interpreted alongside historical implementation research, the included studies collectively suggested that effective medication safety informatics, whether classical CDSS or ML-enhanced, was most likely to have been successful when it was delivered as part of clinician workflow and produced action-oriented recommendations at the time and place of decision-making, characteristics previously associated with improved practice effects in randomized and controlled studies of decision support [24]. Evidence in pediatric and neonatal intensive care settings further contextualized how ML-oriented tools were evaluated against a backdrop of high baseline medication complexity and relatively low tolerance for alert overload.

In the pediatric intensive care unit (PICU) dosing CDSS evaluation, the incidence of protocol deviations requiring adjustment fell from 11/5,034 (0.22%) pre-implementation to 2/4,308 (0.05%) post-implementation, corresponding to a post-versus-pre relative risk (RR) of 0.21 (95% confidence interval [CI]. 0.05–0.96) for deviations requiring adjustment; unjustified deviations similarly fell from 28/5,034 (0.56%) to 10/4,308 (0.23%), RR 0.42 (95% CI 0.20–0.86) [16]. Although absolute event rates were small, these changes were consistent with the notion that pediatric intensive care medication error prevention frequently relied on narrowly specified dose-range checking integrated into routine ordering, rather than on broad-spectrum alerting. This observation was compatible with broader pediatric and neonatal intensive care evidence indicating that medication errors occurred frequently in these environments (e.g., median pediatric intensive care unit medication error rates reported as 14.6 per 100 medication orders, with neonatal intensive care unit rates ranging up to 77.9 per 100 medication orders in published quantitative good

studies), underscoring the importance of interventions that reduced risk while maintaining usability [30]. In the neonatal intensive care unit (NICU) ML model study, the algorithmic output addressed error prediction at a population level rather than offering narrow dose checks; the model achieved an area under the curve (AUC) of 0.920 (95% CI 0.876–0.970) for predicting the presence of medication errors and identified that physician- and nurse-related medication errors were common (42.2% and 57.0% of patients, respectively) in the analyzed cohort [20]. Together, the pediatric critical care findings suggested that the most credible pathway for ML in these settings was likely to have been precision screening and prioritization rather than expansive alert generation, consistent with clinically observed constraints around alarm load and the need for high value-per-alert.

The admission-focused ML studies reinforced that resource allocation was a central mechanism by which AI-assisted systems may have improved medication safety outcomes. The hospital admission medication reconciliation prioritization tool improved the yield of identifying patients with at least one unintended discrepancy from 23/110 (21%) under the conventional approach to 49/110 (45%) under the ML tool, RR 2.13 (95% CI 1.40–3.24) for discrepancy identification under fixed review capacity [18]. This approach complemented established medication reconciliation evidence, where pharmacist-led programs were associated with reductions in post-discharge healthcare utilization; in a systematic review and meta-analysis of pharmacist-led medication reconciliation at hospital transitions, pooled estimates showed reductions in adverse drug event–related hospital revisits (RR 0.33, 95% CI 0.20–0.53), emergency department visits (RR 0.72, 95% CI 0.57–0.92), and readmissions (RR 0.81, 95% CI 0.70–0.95) [31].

Interpreted together, these findings suggested that ML-based prioritization could have functioned as an upstream "triage layer" that increased the efficiency of pharmacist-led reconciliation programs, potentially allowing the same staffing to cover more high-risk admissions or to increase the intensity of review among those most likely to harbor clinically relevant discrepancies. A parallel admission-time model aimed at predicting inpatient adverse drug events also illustrated the practical challenge of low event prevalence for operational deployment: despite acceptable the discrimination (gradient boosting machine AUC 0.747, 95% CI 0.735–0.759; AUC–precision–recall 0.134, 95% CI 0.131–0.137), low positive predictive performance remained likely when adverse drug event incidence was low, implying that the principal benefit could have been in reliably identifying low-risk patients for whom intensive review may not have been needed [19]. Thus, the included admission-time studies collectively implied that the principal comparative advantage of ML may have been targeting rather than replacing clinical judgment, and that evaluation frameworks emphasizing precision–recall were better aligned with clinical operations than receiver-operating curve metrics alone. Alert fatigue and clinician response behavior emerged as a critical interpretive lens for the included studies that focused on CDSS alert filtering and physician response prediction.

The large alert-log study demonstrated that, with sensitivity fixed at 0.99, a gradient boosting approach achieved precision 0.192 and could have reduced alert volume by 54.1%, suggesting that substantial reductions in alert exposure could have been achieved without materially increasing false negatives beyond a pre-specified threshold [14]. Complementarily, physician response prediction models trained on disease medication–related CDSS alerts achieved strong discrimination (artificial neural network AUROC 0.94, accuracy 0.85, sensitivity 0.87, specificity 0.83) [15]. These findings were interpretable against a longstanding body of literature documenting high override rates for drug safety alerts in CPOE systems (overrides reported in the range of 49% to 96% across included studies in a major review), indicating that high-volume, low-specificity alerting could have produced error-prone conditions rather than safety gains [28]. The included ML filtering findings therefore suggested that AI would have been most beneficial when it reduced exposure to low-value alerts while preserving detection of high-risk events with high sensitivity.

However, the same mechanism also implied a risk: if model drift, documentation changes, or policy changes shifted alert characteristics over time, ML-based suppression strategies could have inadvertently concealed clinically important warnings, reinforcing the importance of local monitoring and ongoing model governance as part of routine CDSS maintenance. The included interventional and monitoring studies in administration-stage medication safety suggested that AI systems may have contributed most directly when matched to well-defined, high-risk failure modes. The infusion monitoring system used a multimodal sensing

and deep learning to detect mismatches between infusion pump settings, the actual infusion state, and prescribing instructions in real time; it achieved training/validation/test accuracies of 98.3%, 97.7%, and 98.5% and infusion-rate estimation errors in the range of 0.22–2.90%, indicating strong technical performance under evaluation conditions [21]. This result was consonant with prior evidence on smart infusion pumps, where systematic review findings indicated that smart pumps intercepted wrong-rate, wrong-dose, and pump-setting errors but were limited by compliance issues and high override rates for "soft limits," suggesting that technical capability alone was insufficient without high-fidelity clinical integration [27].

In contrast to infusion-stage systems, barcode-supported administration technologies provided a well-established benchmark for measurable reductions in administration errors. In a before-and-after clinical trial of barcode-enabled electronic medication administration records (eMAR), nontiming administration errors fell from 11.5% to 6.8% (relative reduction 41.4%, P<0.001), and potential adverse drug events (excluding timing-related) fell from 3.1% to 1.6% (relative reduction 50.8%, P<0.001), while transcription errors were eliminated (from 6.1% to 0%) [26]. When these comparisons were made, the included AI infusion monitoring evaluation appeared promising but comparatively early in translational maturity, because it emphasized device-level accuracy metrics rather than demonstrable reductions in medication errors or adverse events in routine care pathways.

The systematic review title addressed perioperative and inpatient medication errors, yet the included evidence base remained weighted toward inpatient prescribing, alert optimization, pediatric dosing support, and admission-time reconciliation or harm prediction, with limited direct evaluation in perioperative medication preparation and administration workflows. This imbalance was important because the perioperative medication-use process carried distinct risks and time pressures; in prospective observational work, medication errors and/or adverse drug events were observed in 5.3% of perioperative medication administrations (95% CI 4.5–6.0), with 79.3% deemed preventable and a substantial proportion classified as serious, indicating a high-risk environment where rapid, ergonomic safety defenses would have been needed [29]. The included the ICU medication decision support intervention demonstrated how targeted systems could have improved guideline adherence and altered prescribing behavior in a high-acuity context (guideline adherence increased from 47.64% to 68.26%, and 60.15% of alerts led to prescription modification) [17], but the generalizability of these effects to anesthesia medication preparation and administration remained uncertain. The cross-cutting methodological signal across the included studies was that model performance was strongly shaped by outcome definition and by event prevalence. For example, admission-time adverse drug event prediction achieved AUCs near 0.75 but exhibited low AUC–precision–recall values (approximately 0.13–0.14) despite statistically significant improvements over logistic regression.

This is demonstrating that "good discrimination" could coexist with limited operational value if the downstream workflow required high positive predictive value to justify interruption or manual review [19]. Consequently, the included evidence suggested that the most plausible near-term safety impact of AI systems may have been achieved when algorithms were used to (i) prevent egregious dosing and protocol deviations in pediatrics [16], (ii) provide low-burden, high-validity outlier alerts that triggered order modification [12], and (iii) improve the yield of scarce pharmacist reconciliation resources through risk-based selection [18], rather than when they attempted to comprehensively replace conventional medication safety infrastructure. This systematic review had several limitations that may have influenced the completeness and interpretability of the synthesized evidence.

The search strategy primarily relied on PubMed and English-language publication constraints, which may have excluded relevant evaluations indexed exclusively in other biomedical or engineering databases or published in non-English journals. The heterogeneity of study designs, target settings (inpatient wards, pediatric intensive care units, neonatal intensive care units, and intensive care units), and outcome definitions (protocol deviations, unintended discrepancies, alert filtering performance, guideline adherence, adverse drug event prediction, and device-monitoring accuracy) precluded meaningful quantitative pooling and increased the risk that narrative synthesis overstated coherence across fundamentally different interventions. Additionally, most included evaluations were single-center or system-specific, which may have limited external validity and amplified the influence of local workflows, prescribing culture, and the baseline safety infrastructure. Finally,

because the review did not include protocol registration, there was an increased risk that methodological decisions (e.g., subgroup emphasis) were influenced by the available evidence base rather than strictly prespecified analytic intent. Despite these limitations, the review had several strengths. The included evidence was restricted to clinical trial or cohort-type evaluations embedded in real hospital care, thereby prioritizing translational relevance over purely simulated modeling studies and enabling interpretation in terms of real implementation constraints (alert burden, workflow integration, and actionable decision-making).

The synthesis incorporated multiple medication-use stages (prescribing/order verification, admission medication reconciliation, administration monitoring) and emphasized both clinically anchored outcomes (order modification, discrepancy yield, protocol deviation reduction) and implementation-relevant metrics (precision under fixed sensitivity, area under the precision–recall curve), which strengthened interpretability for safety engineering and hospital governance. Moreover, the external comparisons contextualized the AI findings within the broader evolution of medication safety interventions, CPOE/CDSS, barcode eMAR, and smart pump infrastructures, highlighting where AI-enhanced systems appeared to extend existing approaches versus where they remained earlier-stage proofs of capability.

Overall, the evidence synthesized in this review indicated that AI-enabled medication safety systems were most consistently associated with improvements in targeted detection and prioritization, low-burden outlier alerting with high clinical validity and frequent order modification [12], improved prioritization of high-risk prescriptions [13,22], reduced pediatric protocol deviations [16], enhanced yield of medication reconciliation activities (RR 2.13, 95% CI 1.40–3.24) [18], and improved discrimination for adverse drug event prediction with explicit limitations under low event prevalence (AUC 0.747, 95% CI 0.735–0.759; AUC–precision–recall 0.134, 95% CI 0.131–0.137) [19]. The clearest comparative lesson from external literature was that technology-mediated medication safety gains historically occurred when systems were integrated into workflow and produced measurable reductions in high-frequency errors (e.g., barcode eMAR relative reductions of 41.4% in nontiming administration errors and 50.8% in potential adverse drug events) [26], while persistent challenges such as alert overriding (49–96%) limited the impact of poorly tuned the drug safety alerts [28]. For Saudi Arabia, the national medication error reporting evidence demonstrated that reported medication errors were concentrated in the prescribing stage (84.8%), that only 5.8% reached patients, and that work overload and lack of experience were commonly associated contributing factors (31.6% and 22.7%, respectively) [32]. In perioperative practice, a Saudi survey indicated that 69% of anesthesia clinicians had experienced anesthetic drug errors at least once, with haste and workload prominent contributors (each 60.3%) and fear of medicolegal consequences acting as a major barrier to reporting (77.7%) [33]. These patterns suggested that Saudi implementation priorities could have emphasized AI-supported optimization of prescribing surveillance and reconciliation targeting in parallel with perioperative human-factors interventions (standardized syringe labeling, double-check systems, and technology-assisted verification), with explicit attention to nonpunitive safety culture and measurement strategies that linked AI deployment to reductions in clinically meaningful errors and preventable harm.

Conclusions

The available hospital-based evidence indicated that artificial intelligence–enabled medication safety systems generally supported prevention of medication errors in perioperative-adjacent and inpatient care by improving identification and prioritization of high-risk prescriptions or patients, strengthening prescribing quality and guideline adherence in implementation evaluations, and reducing the operational burden of medication safety surveillance through more selective, response-aware alerting. The most consistent benefits were observed when these tools were tightly integrated into routine clinical workflows, such as order entry, pharmacist verification, medication reconciliation, or real-time administration monitoring, and configured to deliver low-burden outputs with high clinical relevance. At the same time, direct comparison of effect sizes across studies was constrained by heterogeneity in settings, definitions of medication error, outcome denominators, and evaluation designs, and relatively few reports assessed preventable adverse drug events or sustained patient-level outcomes over longer follow-up.

References:

1. *Hodkinson A, Tyler N, Ashcroft DM, et al. Preventable medication harm across health care settings: a systematic review and meta-analysis. BMC Med. 2020;18(1):313.*
2. *Elliott RA, Camacho E, Jankovic D, et al. Economic analysis of the prevalence and clinical and economic burden of medication error in England. BMJ Qual Saf. 2021;30(2):96-105.*
3. *Ye J. Patient safety of perioperative medication through the lens of digital health and artificial intelligence. JMIR Perioper Med. 2023;6:e34453.*
4. *Graafsma J, Murphy RM, van de Garde EMW, et al. The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: a scoping review. J Am Med Inform Assoc. 2024;31(6):1411-1422.*
5. *Chen CY, Chen YL, Scholl J, et al. Ability of machine-learning based clinical decision support system to reduce alert fatigue, wrong-drug errors, and alert users about look alike, sound alike medication. Comput Methods Programs Biomed. 2024;243:107869.*
6. *Raven K, Zhang J, Vagliano I, et al. Determinants influencing medication-related alert handling in the electronic health record: a systematic review and meta-analysis. Int J Med Inform. 2025;203:106011.*
7. *Golder S, Xu D, O'Connor K, et al. Adverse drug event detection using natural language processing and machine learning: a scoping review. Drug Saf. 2025;48(4):321-337.*
8. *Patel M, Nanji KC. Artificial intelligence in perioperative medication-related clinical decision support. Anesthesiol Clin. 2025;43(3):587-602.*
9. *Ikeda K, Koike M, Yamazaki S, et al. Primary anesthesia provider characteristics and risk factors for intraoperative medication errors: a retrospective cohort study. BMC Anesthesiol. 2025;26(1):44.*
10. *AlAmri HM, Tobaiqy M, Ali A. A qualitative study of prescribing errors among clozapine prescribers. Int J Clin Pharm. 2024;46(3):894-902.*
11. *Nanji KC, Patel A, Shaikh S, et al. Evaluation of perioperative medication errors and adverse drug events. Anesthesiology. 2016;124(1):25-34.*
12. *Corny J, Rajkumar A, Martin O, et al. Machine learning-based clinical decision support to identify prescriptions with high risk of medication error. J Am Med Inform Assoc. 2020;27(11):1688-1694.*
13. *Poly TN, Islam MM, Yang HC, et al. Machine learning approach to reduce alert fatigue using a disease medication-related clinical decision support system: model development and validation. JMIR Med Inform. 2020;8(11):e19489.*
14. *Abdo A, Gallay L, Vallecillo T, et al. A machine learning-based clinical predictive tool to identify patients at high risk of medication errors. Sci Rep. 2024;14(1):32022.*
15. *Mor Naor G, Tocut M, Moalem M, et al. Screening for medication errors and adverse events using outlier detection screening algorithms in an inpatient setting. J Med Syst. 2022;46(12):88.*
16. *Hashemi F, van Gelder TG, Bollen CW, et al. The effect of a decision support system on the incidence of prescription errors in a PICU. J Clin Pharm Ther. 2022;47(3):330-344.*
17. *Dashti N, Hosseinjani H, Eslami S, et al. Effect of a medication decision support system on albumin prescribing: a pre-post study in intensive care units. J Res Pharm Pract. 2025;14(2):39-49.*
18. *Chan K, Hartman J, Nowacki AS, et al. Detecting medication administration errors with artificial intelligence-enabled wearable cameras. NPJ Digit Med. 2024;7(1):226.*
19. *Amici LD, van Pelt M, Mylott L, et al. Clinical decision support as a prevention tool for medication errors in the operating room: a retrospective cross-sectional study. Anesth Analg. 2024;139(4):832-839.*
20. *Yang X, Bian J, Gong Y, et al. MADEx: a system for detecting medications, adverse drug events, and their relations from clinical notes. Drug Saf. 2019;42(1):123-133.*
21. *Chapman AB, Peterson KS, Alba PR, et al. Detecting adverse drug events with rapidly trained classification models. Drug Saf. 2019;42(1):147-156.*
22. *van der Sijs H, Aarts J, Vulto A, Berg M. Overriding of drug safety alerts in computerized physician order entry. J Am Med Inform Assoc. 2006;13(2):138-147.*
23. *Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ. 2005;330(7494):765.*
24. *Kaushal R, Bates DW, Landrigan C, et al. Medication errors and adverse drug events in pediatric inpatients. Arch Intern Med. 2003;163(12):1409-1416.*

25. *van Rosse F, Maat B, Rademaker CMA, et al. The effect of computerized physician order entry on medication prescription errors and clinical outcome in pediatric and intensive care: a systematic review. Pediatrics. 2009;123(4):1184-1190.*

26. *Poon EG, Keohane CA, Yoon CS, et al. Effect of bar-code technology on the safety of medication administration. N Engl J Med. 2010;362(18):1698-1707.*

27. *Ohashi K, Dalleur O, Dykes PC, Bates DW. Benefits and risks of using smart pumps to reduce medication error rates: a systematic review. Drug Saf. 2014;37(12):1011-1020.*

28. *Mekonnen AB, McLachlan AJ, Brien JE. Effectiveness of pharmacist-led medication reconciliation programmes on clinical outcomes at hospital transitions: a systematic review and meta-analysis. BMJ Open. 2016;6(2):e010003.*

29. *Alghamdi SA, Keers RN, Sutherland A, Ashcroft DM. Prevalence and nature of medication errors and preventable adverse drug events in paediatric and neonatal intensive care settings: a systematic review. Drug Saf. 2019;42(12):1423-1436.*

30. *Alshammari TM, Alenzi KA, Alatawi Y, et al. Current situation of medication errors in Saudi Arabia: a nationwide observational study. J Patient Saf. 2022;18(2):e448-e453.*

31. *Aldossary S, Alshahrani A, Alqahtani A, et al. Medication errors and adverse drug reactions in a Saudi hospital: the impact of computerized provider order entry system. Glob J Qual Saf Healthc. 2021;4(2):60-70.*

32. *Catchpole KR, Neyens DM, Abernathy JH 3rd, et al. Rethinking anesthesia medication "errors": the OR-SMART patient safety learning laboratory. J Patient Saf. 2025;21(7):503-509.*

33. *Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71.*

Table 1. Characteristics and key findings of the studies included in the review on Artificial Intelligence in Medication Safety Systems for the Prevention of Perioperative and Inpatient Medication Errors

Study Reference	Study Design	Population	Intervention / Exposure	Disease / Condition	Main Outcomes
[12]. Segal et al., 2019	Implementation cohort	Adult inpatients (hospital wards)	Probabilistic ML outlier-detection CDSS	Inpatient medication safety	Alerts 0.4% of orders; 85% clinically valid; 43% led to order change; median response 1 h.
[13]. Corny et al., 2020	Development/validation cohort	Hospitalized adults (multiple units)	ML-CDSS to prioritize high-risk prescriptions	Medication error risk (prescribing)	AUROC 0.81; AUPRC 0.75 for high-risk prescription detection.
[14]. Liu et al., 2022	Retrospective cohort	Hospital medication alert logs	ML filtering of medication alerts (GBM)	Alert fatigue / safety alerts	At sensitivity 0.99: precision 0.192; estimated 54.1% alert-volume reduction.
[15]. Poly et al., 2020	Development/validation cohort	Clinician-handled CDSS alerts	ANN model predicting alert acceptance	Alert fatigue / decision support	AUROC 0.94; accuracy 0.85; sensitivity 0.87; specificity 0.83 for response prediction.
[16]. Hashemi et al., 2022	Pre–post cohort	Pediatric ICU (PICU) patients	Dosing CDSS integrated in prescribing	PICU prescribing errors	Dose deviations needing adjustment RR 0.21 (95% CI 0.05–0.96); unjustified deviations RR 0.42 (95% CI 0.20–0.86).
[17]. Dashti et al., 2025	Pre–post cohort	ICU patients prescribed albumin	Medication decision support system (MDSS)	ICU prescribing quality (albumin)	Guideline adherence 47.6%→68.3%; 60.2% alerts modified orders; safety composite 63.3%→82.6%.
[18]. Abdo et al., 2024	Development + simulation cohort	Hospital admissions at intake	ML tool to prioritize high-risk admissions	Medication errors / reconciliation	AUROC 0.74; discrepancy yield 45% vs 21% (RR 2.13, 95% CI 1.40–3.24) in evaluation subset.
[19]. Langenberger et al., 2023	Development/validation cohort	Adult ward admissions	ML prediction of in-hospital ADE risk	Adverse drug events	AUROC 0.752; Brier 0.056; calibration slope 0.997 at admission.
[20]. Yalçın et al., 2023	Development/validation cohort	Neonatal ICU (NICU) patients	ML-based medication error detection	NICU medication errors	AUROC 0.97; sensitivity 0.97; specificity 0.90; accuracy 0.91; F1 0.92.

[21]. Hwang et al., 2021	Technical validation study	IV infusion administrations	Multimodal deep-learning pump monitoring	IV administration safety	Accuracy 98.5%; infusion-rate error 0.22–2.90% in evaluation experiments.
[22]. Dos Santos et al., 2019	Retrospective data-mining study	Large prescribing dataset	Unsupervised outlier detection (DDC-Outlier)	Prescribing anomaly detection	Identified outlier prescriptions for review at scale; downstream clinical impact not quantified.

Abbreviations: AI, artificial intelligence; ANN, artificial neural network; AUPRC, area under precision–recall curve; AUROC, area under receiver operating characteristic curve; CDSS, clinical decision support system; CI, confidence interval; GBM, gradient boosting machine; ICU, intensive care unit; IV, intravenous; ML, machine learning; MDSS, medication decision support system; NICU, neonatal intensive care unit; PICU, pediatric intensive care unit; RR, relative risk.

.

Medicina Katastrof
(ISSN:2070-1004) (E-ISSN:2686-7966)
Scopus Index(2025)