



Review

The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: a scoping review

Jetske Graafsma , PharmD^{*.1}, Rachel M. Murphy , PharmD^{2,3},
Ewoudt M.W. van de Garde, PhD^{4,5}, Fatma Karapinar-Çarkit, PhD^{6,7}, Hieronymus J. Derijks, PhD⁸,
Rien H.L. Hoge, PhD⁹, Joanna E. Klopotoska, PhD^{2,3}, Patricia M.L.A. van den Bemt, PhD¹

¹Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, University of Groningen, Groningen, 9713GZ, The Netherlands, ²Department of Medical Informatics Amsterdam UMC, University of Amsterdam, Amsterdam, 1000GG, The Netherlands, ³Amsterdam Public Health Institute, Digital Health and Quality of Care, Amsterdam, 1105AZ, The Netherlands, ⁴Department of Pharmacy, St Antonius Hospital, Utrecht, 3430AM, The Netherlands, ⁵Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, 3584CS, The Netherlands, ⁶Department of Clinical Pharmacy and Toxicology, Maastricht University Medical Center, Maastricht, 6229HX, The Netherlands, ⁷Department of Clinical Pharmacy, CARIM, Cardiovascular Research Institute Maastricht, Maastricht University, Maastricht, 6229ER, The Netherlands, ⁸Department of Pharmacy, Jeroen Bosch Hospital, Den Bosch, 5200ME, The Netherlands, ⁹Department of Pharmacy, Wilhelmina Hospital, Assen, 9401RK, The Netherlands

*Corresponding author: Jetske Graafsma, PharmD, Department of Clinical Pharmacy and Pharmacology, University Medical Center Groningen, P.O. Box 30001—Building Code AP50, Groningen 9700 RB, The Netherlands (j.graafsma@umcg.nl)

Abstract

Objective: Current Clinical Decision Support Systems (CDSSs) generate medication alerts that are of limited clinical value, causing alert fatigue. Artificial Intelligence (AI)-based methods may help in optimizing medication alerts. Therefore, we conducted a scoping review on the current state of the use of AI to optimize medication alerts in a hospital setting. Specifically, we aimed to identify the applied AI methods used together with their performance measures and main outcome measures.

Materials and Methods: We searched Medline, Embase, and Cochrane Library database on May 25, 2023 for studies of any quantitative design, in which the use of AI-based methods was investigated to optimize medication alerts generated by CDSSs in a hospital setting. The screening process was supported by ASReview software.

Results: Out of 5625 citations screened for eligibility, 10 studies were included. Three studies (30%) reported on both statistical performance and clinical outcomes. The most often reported performance measure was positive predictive value ranging from 9% to 100%. Regarding main outcome measures, alerts optimized using AI-based methods resulted in a decreased alert burden, increased identification of inappropriate or atypical prescriptions, and enabled prediction of user responses. In only 2 studies the AI-based alerts were implemented in hospital practice, and none of the studies conducted external validation.

Discussion and Conclusion: AI-based methods can be used to optimize medication alerts in a hospital setting. However, reporting on models' development and validation should be improved, and external validation and implementation in hospital practice should be encouraged.

Key words: artificial intelligence; clinical decision support systems; medication safety; medication alerts.

Background and significance

Preventing adverse drug events (ADEs) is an essential part of medication safety efforts worldwide.¹ ADEs can lead to serious harm and even death and contribute to excess health-care costs.^{2–4} A recent retrospective cohort study in 11 hospitals in the United States showed that ADEs, despite various medication safety efforts, are still the most frequent (39.0%) type of adverse events patients encounter during their hospital stay, and are often deemed preventable (26.8%).⁵ Among the most broadly deployed tools to prevent ADEs in hospitalized patients, are clinical decision support systems (CDSSs).⁶

CDSSs generate medication alerts when the content of the electronic hospital record (EHR) of a patient matches with pre-defined rules in a knowledge database upon which CDSSs operate. The alerts are primarily shown during the medication prescribing phase. The goal of these alerts is to support healthcare providers in checking dosages, drug-drug interactions, contra-indications, duplicate therapy, drug allergies, and intolerances, and by doing so reducing the ADE risk of hospitalized patients. However, current CDSSs generate a high alert volume containing medication alerts with no or limited clinical relevance, resulting in alert fatigue and override rates as high as 96%. This alert fatigue is concerning as

it may result in missing clinically significant alerts, which compromises patient safety.^{7–11}

This is mainly caused by the fact that the current medication alerts fail to account for the complexity of care and patient specific variables.^{10,12,13} In the past years, several attempts have been made to improve appropriateness and clinical value of medication alerts. A recent scoping review of Ledger et al identified 6 types of interventions for medication alerts in hospitals: alert inactivation, alert severity reclassification, information provision, use of contextual information, threshold adjustment, and encounter suppression.¹⁴ A study by Bakker et al showed the effect of alert inactivation and severity reclassification; only drug-drug interaction alerts assessed as clinically relevant in the intensive care setting were turned on, which resulted in 12% decrease in the number of high-risk combinations.¹⁵ Selecting the alerts to be inactivated by employing a multidisciplinary committee has shown to be an effective approach.¹⁶ A way to use contextual information to optimize alerts is by designing computerized decision tree rules, to context-dependently suppress irrelevant drug-drug interaction alerts, which has been shown to result in less alerts and a higher positive predictive value (PPV).¹⁷ However, given the high number of medication alerts and high override rates, room for improvement remains. Medication alerts should be more specific, for example by targeting only high risk contraindications, or by tailoring the alerts to specific medical specialties.^{18,19}

Artificial Intelligence (AI) methods may contribute to further optimization of medication alerts generated by CDSSs, eg, by predicting physician responses, generating AI-based medication alerts, or by developing a triage system. In contrast to rigid and simple decision tree like logic upon which current CDSSs operate, AI-based methods can take into account large amounts of EHR data, recognize complex patterns, and provide individualized predictions.²⁰ In medicine, the most often used AI-based methods are machine learning (ML), deep learning (DL), and natural language processing (NLP).²¹ ML and DL can be applied to structured data, whereas NLP can be used on unstructured data. In ML data analytical algorithms are developed to extract features from data, which can be used to cluster patients characteristics or predict the probability of disease outcomes. DL is an extension of ML, which can be described as networks with a large number of layers, consequently leading to the ability to explore more complex non-linear patterns in the data. In NLP, information is extracted from unstructured data, such as clinical notes from physicians, to supplement and enrich medical data.²²

Use of these methods to create new or optimize existing medication alerts may help to reduce alert fatigue, for example by prioritizing the alerts based on appropriateness and usefulness, but also in preventing ADEs which are currently missed by the existing CDSSs.¹² Several literature reviews have already been performed on AI and CDSS in specific domains such as in oncology, dentistry, or infectious disease.^{23–25} However, none of these reviews focused on medication alerts generated by CDSS in hospitalized patients.

Therefore, the aim of this scoping review was to provide a comprehensive overview of the current state of applying AI-based methods to optimize medication alerts generated by CDSS in a hospital setting. By synthesizing the available evidence, this review aims to inform on the potential AI offers for improvement of CDSS and to identify opportunities for future research.

Methods

Approach

The proposed scoping review was conducted in accordance with the Joanna Briggs Institute (JBI) methodology for scoping reviews²⁶ and reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR).²⁷ The protocol was registered in the Open Science Framework.²⁸

Information sources and search strategy

A preliminary search of Medline and Embase databases was conducted on May 10, 2023 before conducting this scoping review and no systematic or scoping reviews were identified on the specific topic of this review. Furthermore, at that time, no registered research protocols were found in the Open Science Framework or PROSPERO on this specific topic.

The literature search was conducted in Medline, Embase, and Cochrane Library (reviews and trials) databases on May 25, 2023 using the definite search strategy. No filters were applied with regard to publication year and country. The search was limited to the English language. The search strategy consisted of 3 segments: 1 related to AI-based methods, 1 related to pharmacotherapy, and 1 related to CDSS. The exact search strategies for each database can be found in [Supplement Appendix S1](#). Additionally, the references of relevant studies were screened to identify possible other relevant studies. Forward citation was performed in Web of Science and in Scopus to identify additional relevant studies, using the studies initially labeled relevant during the title/abstract screening phase. If the studies identified via these additional searches were not already identified via the primary search strategy, the search strategy was reviewed and adjusted if needed to be able to include all relevant studies.

Eligibility criteria

This scoping review aimed to include all studies that explore the use of AI-based methods to optimize medication alerts generated by CDSS in a hospital setting. We included a wide range of AI-based methods including supervised and unsupervised ML, NLP, and DL methods. Quantitative and mixed-method studies of any observational or interventional design (including but not limited to cohort studies, randomized controlled trials, and controlled trials) were eligible for inclusion. Scoping and systematic reviews or meta-analyses that met the inclusion criteria were not included but their references lists were screened for relevant studies. Qualitative studies, case reports, abstracts of congresses, expert opinions, editorials, and narrative reviews were excluded.

This review focusses on optimization of medication alerts generated by CDSSs. Optimization of medication alerts by applying AI can be achieved in different ways, eg, by predicting physicians responses, generating AI-based medication alerts, or developing a triage system. Regarding the type of medication alerts included, we considered alerts generated at prescribing and monitoring stages as we were interested in medication alerts for prescribers and pharmacists. Furthermore, these 2 stages hold the highest risk for ADEs.^{29,30} These alerts could warn for various risks, such as over- and underdosing, drug-drug interactions, contra-indications (including pregnancy and lactation), duplicate therapy, and drug intolerances and -allergies.

Regarding the use of AI-based methods, studies that did not report any performance measures or clinical outcomes were excluded. No restrictions were made to the type of performance measures or clinical outcomes used.

Data extraction and synthesis

The titles and abstracts of the studies generated by the search strategy were collected and uploaded into Endnote© citation manager to remove duplicates. The first reviewer (J.G.) screened the titles and abstracts using ASReview version 1.0.³¹ ASReview is a tool designed to accelerate the screening of large numbers of literature references using active learning, a type of ML, with the main principle to achieve higher accuracy with fewer training data if the algorithm can choose the data from which it learns.^{32,33} The algorithm does not choose the included studies but merely presents the studies to the researchers in order of probability of relevance based on prior knowledge. It does not influence the search, but only uses the results of the search.

ASReview requires researchers to specify relevant and irrelevant papers related to a specific research question as prior knowledge, to train its algorithm. In this study, a total of 10 relevant and 10 irrelevant studies (as assessed by J.G.) were used to train the algorithm. Based on this prior knowledge the algorithm predicted a ranking of the relevance of all papers uploaded in the tool. Thereafter the screening of the title and abstract of the studies could start. van de Schoot et al reported that the number of relevant abstracts found after

reading 10% of the abstracts ranges from 70% to 100%. Furthermore, 8% to 33% of all abstracts have to be screened to find 95% of the relevant studies.³⁰ Following these findings, it was decided that after screening at least 10% of the abstracts and titles and 50 studies were consecutively identified as irrelevant based on the ranking of ASReview, the title and abstract screening process will be terminated (ie, stopping rule).

The results using ASReview ranking algorithm were verified by a second reviewer (R.M.) by manually screening a sample of the studies included through ASReview's screening process, comprising 5% of the total number of studies and including the prior knowledge, to confirm that they would be included. Furthermore, the first and second reviewer (J.G. and J.M.) verified the results using ASReview by manually screening a random sample (comprising 5% of the total number of studies) of the studies excluded through ASReview's screening process to confirm that they would be excluded. The process of the title and abstract screening is visualized in Figure 1.

At least 80% of the findings from the first and second reviewer had to be identical to validate the abstract screening and start the full text screening. If this was not the case, the first and second author had to screen another 10%. Subsequently, discrepancies in the findings were discussed to reach consensus and the inclusion criteria were reviewed and adjusted for clarification if needed.

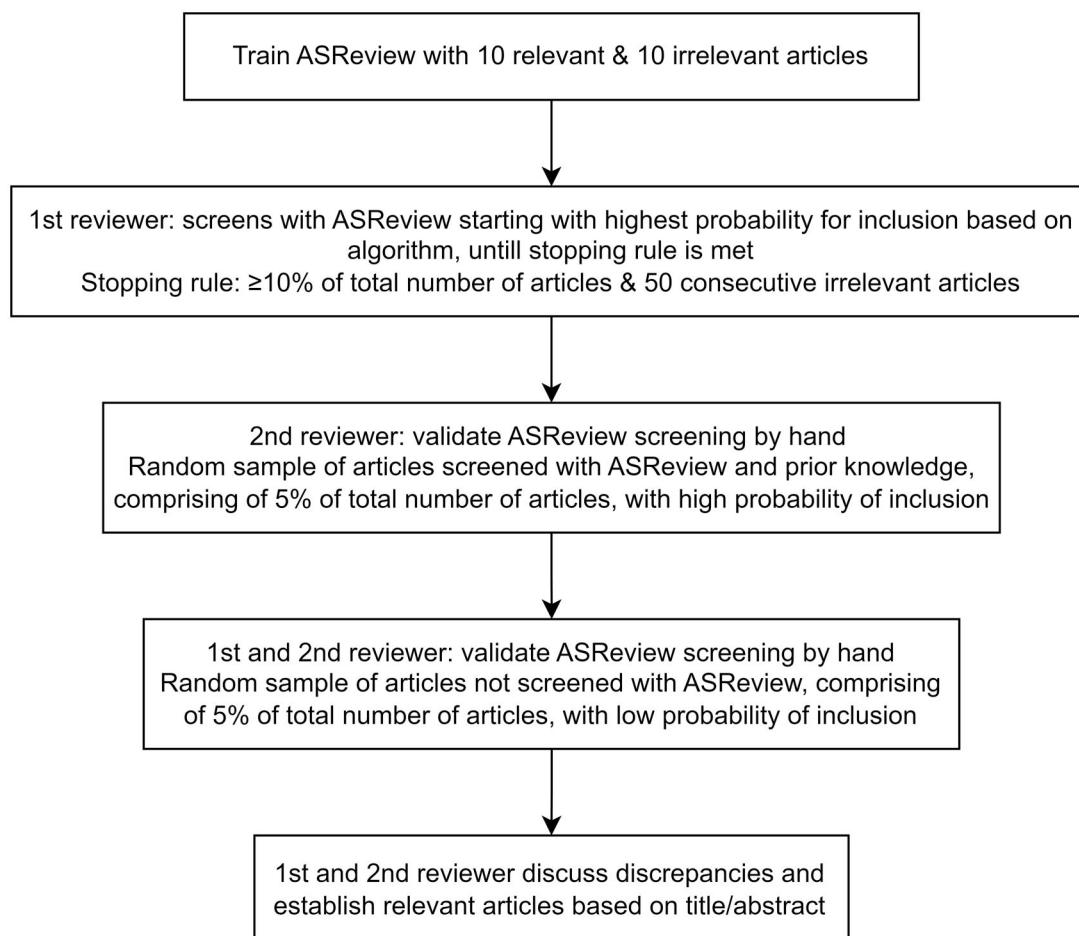


Figure 1. Process of title and abstract screening using ASReview software.

Thereafter the full text screening and data extraction could start. The full-text review was done manually. Two reviewers (J.G., R.M.) screened 10% of the studies included in the full text screening to establish information to extract from the studies. The remaining studies for full-text screening were divided equally between these 2 reviewers. In case of doubt, the 2 reviewers tried to reach consensus. If this was not possible, third reviewer (P.v.d.B.) decided. All decisions were documented.

The information extracted from the included studies consisted of basic characteristics of the studies (title, author, affiliations, year of publication, journal, and country of origin), the study aims, study setting (eg, oncology, cardiology, intensive care, pediatrics, emergency), study methods, medication alerts characteristics (eg, drug-drug interactions, dosages, allergies alerts), AI-based methods used and their statistical performance, CDSS characteristics (eg, developer, targeted at physicians, pharmacist, or other healthcare professional), clinical outcomes studied and the findings, as well as most important conclusions. The extracted data were recorded in an Excel version 2307 (Microsoft, Redmond, WA, United States).

Results

Selection of included studies

The flowchart of the selection process is shown in Figure 2. In total, 7553 citations were identified. After removal of 1928 duplicates, 5625 citations were left to be screened based on the title and abstract. After screening 10% of all citations in ASReview by the first reviewer, the stopping rule was reached since 126 subsequent irrelevant citations were found. The title and abstract screening eventually resulted in inclusion of 64 studies for full-text screening. Of these, 10 studies

met the inclusion criteria and were included for data-extraction and analysis. No additional relevant studies were identified by checking references of relevant studies and forward citation in Web of Science and Scopus.

Main characteristics of the included studies

In Table 1, a summary of the main characteristics of the 10 included studies is provided. The studies were published between 2013 and 2022. Nine studies (90%) were conducted in an academic or university hospital.^{34–42} Four studies (40%) stated the medical specialty studied.^{34,38,40,42} Segal et al included data from the internal medicine department⁴² and Lee et al from the pediatric department.⁴⁰ Beaudoin et al focused on (inappropriate) antimicrobial prescriptions, in specific piperacillin-tazobactam prescriptions.³⁴ Hogue et al included data from 7 different departments: obstetrics-gynecology and nursery, general pediatrics, surgery, oncology, specialized pediatrics, neonatal intensive care unit, and pediatric intensive care unit.³⁸ The remaining 6 studies (60%) did not explicitly state the clinical domain or patient type studied.^{35–37,39,41,43} However, Kawazoe et al did mention including data from the 8 medicines most frequently causing alerts: ursodeoxycholic acid, carvedilol, sennoside, loxoprofen, brotizolam, nifedipine, famotidine, and pravastatin.³⁵

Types of alerts optimized

Several types of medication alerts were generated by the CDSSs studied; over- and underdosing, drug-drug interactions, contra-indications (including pregnancy and lactation), duplicate therapy, and drug intolerances and -allergies. Except for the study by Kawazoe et al, where specifically dosing alerts were targeted, all others focused on optimizing the process related to the medication alerts. The study by

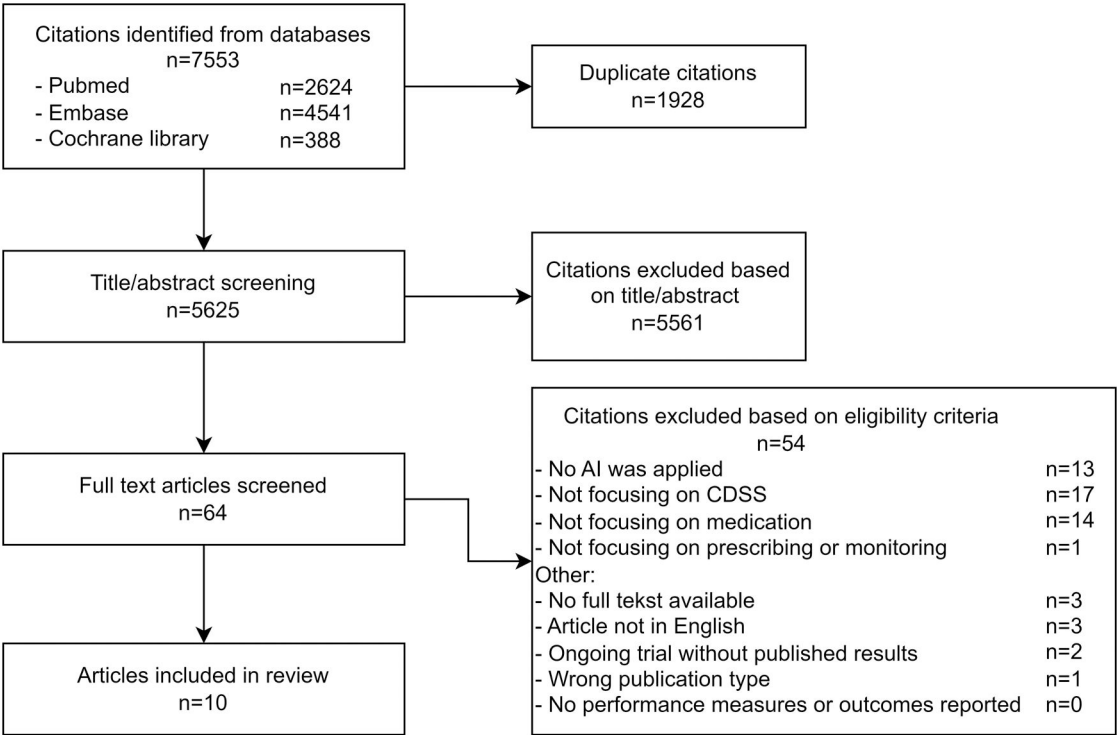


Figure 2. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram.

Table 1. Main characteristics of the included studies.

Author	Year	Country	Study design	Setting	Specialty	Task	Timeframe	Data	Dataset
Segal et al ⁴²	2019	Israel	Prospective interventional study	An 1800-bed university medical center	Internal medicine	Evaluated an existing system	July 2016–April 2018	4533 admissions and 3160 patients with 78 017 medication orders	
Schiff et al ³⁶	2017	United States	Retrospective, observational cohort study	2 university medical centers with a total of 1700 beds	Not stated	Evaluated an existing system	January 2009–December 2013	747 985 patients who had at least 1 visit during 2012–2013	
Hogue et al ³⁸	2021	Canada	Prospective, interventional study	A 500-bed tertiary care mother-and-child university hospital at CHU	7 different departments: obstetrics-gynecology and nursery, general pediatrics, surgery, oncology, specialized pediatrics, NICU ^a , PICU ^b	Developed a model	Training data: 2005–2018 Testing data: April–August 2020	Training data: not stated Testing data: 12 624 medication orders and 2114 pharmacological profiles	
Lee et al ⁴⁰	2022	Korea	Prospective interventional study	A university hospital	Pediatrics	Developed a model	Training data: January–November 2018 Testing data: December 2018	137 802 normal prescriptions and 1609 prescription errors	
Kawazoe et al ³⁵	2013	Japan	Retrospective, observational study	A university hospital with >1200 beds	Not stated	Developed a model	January 2007–December 2011	Training data: 20 000 medication orders Testing data: 10 000 medication orders	
Liu et al ⁴¹	2022	United States	Prospective, interventional study	A medium-sized university medical center	Not stated	Evaluated an existing system	January–December 2019	3 481 634 medication alerts, 8270 providers, 178 298 patients	
Poly et al ³⁷	2020	Taiwan	Retrospective, observational study	A university medical center	Not stated	Developed a model	August 2018–May 2019	Training data: 60%, validation data: 20%, testing data: 20%	
Cornly et al ⁴³	2020	France	Retrospective observational study	A 592-bed hospital	Not stated	Developed a model	January 2017–August 2018	6453 prescriptions Training data: 60%, validation data: 20%, testing data: 20%	
Balestra et al ³⁹	2021	United States	Prospective, observational study	A university medical center comprised of 3 hospitals with over 1600 beds	Not stated	Developed a model	10 July–24 July, 2017	Training data: 94 720 hospitalizations, 61 611 patients Testing data: 412 patients with 3364 prescription orders	
Beaudoin et al ³⁴	2016	Canada	Prospective interventional study	A 677-bed secondary—and tertiary—care hospital located at 2 sites	Antimicrobial prescriptions	Evaluated a previously developed model	Training data: February–November 2012 Testing data: 18 November, 2013–20 December, 2013	Training data: 2092 patients, 2584 hospitalizations, and 4430 prescriptions Testing data: 350 patients, 421 hospitalizations, and 515 prescriptions	

^a NICU, neonatal intensive care unit.^b PICU, pediatric intensive care unit.

Kawazoe et al specifically focused on optimizing dosing alerts by trying to identify appropriate dosing thresholds for the alerts.³⁵

The approach for optimizing medication alerts varied between the studies. In 4 studies, the goal was to identify and prevent prescription errors.^{36,38,40,42} Liu et al, Poly et al, and Corny et al pursued similar objectives regarding refining the relevance of alerts; focusing on predicting physician's responses^{37,41} or prioritizing prescription checks.⁴³ Balestra et al developed a model based on past recommendations, focusing on designing a model for identifying medication orders requiring an intervention.³⁹ Similarly, Beaudoin et al developed a model using past recommendations, evaluating a CDSSs consisting of a knowledge database linked to a model that extracts classification rules for alerts of inappropriate prescriptions.³⁴

CDSS used

In 2 studies a commercial system called Medaware® (Raana, Israel) was used.^{37,42} Corny et al tested the accuracy of Lumio Medication®, developed by Lumio Medical (Paris, France).⁴³ Seven studies reported developing or validating a model themselves without explicitly naming the model (Table 2).^{34,35,37-41}

Five studies reported targeting the physicians^{35-37,40,42} and 4 targeting pharmacists.^{34,38,39,43} Liu et al aimed at filtering out alerts from the user's view, without specifically mentioning who they were targeting as end-user.⁴¹

Four studies developed or studied a model to generate alerts that were normally not generated by the CDSS, which implies that the developed model is used on top of the regular CDSS.^{34-36,42} Four other studies developed a model aimed to be incorporated into the CDSS, by predicting user responses, by serving as a triage system or by filtering medication alerts generated by the regular CDSS.^{37-39,41} The 2 remaining studies developed a hybrid system linking an ML model to the regular CDSS.^{40,43}

AI-based methods used, their statistical performance, and outcomes

Nine studies reported using ML models (Table 3). Of these studies, 5 reported using supervised learning ML methods,^{34,36,37,42,43} Hogue et al reported using an unsupervised ML method,³⁸ and 3 studies did not state whether they applied supervised or unsupervised ML.^{35,39,41} One study reported developing DL models based on supervised learning.⁴⁰ None of the studies mentioned using NLP. The choice for these AI models and methods was not explicitly stated in the studies, but this generally depends on the specific requirements of the problem, the nature of the data, and the available computational resources. However, none of the included studies mentioned who chose the method of optimization nor was the field of expertise of the study team described.

Only 2 of the included studies reported implementing the developed AI-based model in daily practice (Table 3).^{40,42} Segal et al integrated Medaware into an existing EHR system in 1 university hospital in Israel. Initially, the system operated in a "silent mode" for several months.⁴¹ During this period analyses and monitoring were performed. Once the performance level was acceptable, the system switched to live mode in a single internal medicine department and the physicians started receiving alerts in the EHR system and could respond to them. The model by Lee et al was deployed in 1 university

hospital in Korea, targeting pediatric outpatients and a selected number of medications.⁴⁰

Seven studies reported outcomes, with the most often studied outcome (5 studies, 50%) being alert burden compared to the regular CDSS (Table 4).^{35,40-43} In these studies the alert burden was decreased by 14%-90%. Seven studies report statistical performance measures, in which a high variation is shown.^{34,37-41,43} The most often used performance measure was PPV broadly ranging from 9% to 100%. Lower PPV's were reported in the studies of Hogue et al, Liu et al, and Balestra et al^{38,39,41} ranging from 9% to 49%, and higher PPV's were reported in the studies of Lee et al, Poly et al, Corny et al, and Beaudoin et al ranging from 73% to 100%.^{34,37,40,43} Furthermore, the highest sensitivity and specificity of included studies were reported in the studies from Hogue et al (looking at the pharmacological profiles) and Poly et al ranging from 57% to 100%.^{37,38} Only 4 studies reported both statistical performance and outcomes.^{34,40,41,43}

Discussion

Main findings

This scoping review provides a comprehensive overview of currently available evidence on the use of AI-based methods to optimize medication alerts generated by CDSS in the hospital setting. A decrease in alert burden was the most often studied outcome, varying from 14% to 90% in the included studies. Seven studies reported PPV as a performance measure, which broadly ranged from 9% to 100%. The highest PPV's were reported in the studies of Lee et al, Poly et al, Corny et al, and Beaudoin et al ranging from 73% to 100%. These results indicate that AI-based methods have the potential to optimize medication alerts, but at the same time the results also show that there is substantial room for further improvement in application of these methods for this goal and reporting about such applications. Also, given the lack of external validation, the generalizability of the models and potential for implementation in hospital practice is limited.

To the best of our knowledge, this is the first review on the use of AI-based methods to optimize medication alerts generated by CDSS in the hospital setting. Several studies have been performed on AI in decision support systems in healthcare, but studies on medication alerts specifically are lacking. Furthermore, this review demonstrates that AI represents a novel approach to optimizing medication alerts. A recent scoping review of Ledger et al identified 6 types of interventions for optimization of medication alerts in hospitals; alert inactivation, alert severity reclassification, information provision, use of contextual information, threshold adjustment, and encounter suppression.¹⁴ AI is particularly well-suited for conducting these interventions, as is also shown by the studies included in this review.

The developed AI-based models have shown to decrease the alert burden and help identify more inappropriate or atypical prescriptions, compared to the regular CDSSs, subsequently leading to a decreased alert fatigue.¹⁵ In general, when more alerts are shown to physicians, they are less likely to intervene on these alerts.⁴⁴ However, for AI-based medication alerts to be effective and safe, the models generating these alerts must have an optimal balance between sensitivity and specificity.⁴⁵ The higher the specificity of the models, the less false alerts will be shown, leading to less alert fatigue. The higher the sensitivity, the better the models' ability to

Table 2. Type of alerts or process optimized and CDSS used.

Author	Type of alerts targeted	Prediction of	Name of CDSS	End-user	Type of model
Segal et al ⁴²	NA	Prescription errors	Medaware system	Physician	Model is used on top of regular CDSS
Schiff et al ³⁶	NA	Prescription errors	Medaware system	Physician	Model is used on top of regular CDSS
Hogue et al ³⁸	NA	Atypical drug orders and pharmacological profiles	Not stated	Pharmacist	Model is incorporated in CDSS
Lee et al ⁴⁰	NA	Prescription errors	Not stated	Physician	Hybrid system linking model to regular CDSS
Kawazoe et al ³⁵	Dosing alerts	NA	Not stated	Physician	Model is used on top of regular CDSS
Liu et al ⁴¹	NA	User responses	Not stated	Not stated	Model is incorporated in CDSS
Poly et al ³⁷	NA	User responses	Not stated	Physician	Model is incorporated in CDSS
Corny et al ⁴³	NA	Prescription errors	Lumio Medication System	Pharmacist	Hybrid system linking model to regular CDSS
Balestra et al ³⁹	NA	Medication orders requiring interventions	Not stated	Pharmacist	Model is incorporated in CDSS
Beaudoin et al ³⁴	NA	Prescription errors	Not stated	Pharmacist	Model is used on top of regular CDSS

Abbreviation: NA = not applicable.

Table 3. AI-based methods used, implementation, and validation.

Author	AI-based method used	SL ^a or UL ^b	Implemented in practice	Validation
Segal et al ⁴²	ML	SL	Yes	Internal
Schiff et al ³⁶	ML	SL	No	Not performed
Hogue et al ³⁸	ML: GANomaly	UL	No	Internal
Lee et al ⁴⁰	DL: deep NN ^c (eg, autoencoder)	SL	Yes	Internal
Kawazoe et al ³⁵	ML: bagging, CART ^d , RF ^e	Not stated	No	Internal
Liu et al ⁴¹	ML: GBT ^f (eg, LightGBM), NN ^c (eg, autoencoder), RF ^e , SVM ^g	Not stated	No	Internal
Poly et al ³⁷	ML: GBT ^f , NN ^c (eg, autoencoder), RF ^e , NB ^h	SL	No	Internal
Corny et al ⁴³	ML	SL	No	Internal
Balestra et al ³⁹	ML: GBT ^f (eg, LightGBM)	Not stated	No	Internal
Beaudoin et al ³⁴	ML	SL	No	Internal

^a SL, supervised learning.

^b UL, unsupervised learning.

^c NN, neural network.

^d CART, classification and regression trees.

^e RF, random forest.

^f GBT, gradient boosted trees.

^g SVM, support vector machine.

^h NB, naive Bayes.

produce alerts which warn about events with high probability of patient harm. None of the studies showed a sensitivity and specificity higher than 90%. This shows that optimizing medication alerts in the vast array of all prescriptions and medication alerts is a challenging task, also for AI-based methods.

Despite the promising results showing a decreased alert burden and high PPV values, in most studies the scope is limited as the study was conducted in a single hospital^{36,37,39–43} and/or with a focus on specific medication or departments.^{40,42} Furthermore, in only 2 studies, the optimized alerts or alert process were implemented in practice. Although not explicitly stated in the included studies, lack of alignment between the needs from hospital practice with AI-based efforts, and insufficient gains in terms of clinical outcomes may explain low uptake in hospital practice.^{46,47} Also,

before an AI-based model can be adopted in hospital practice, extensive validation has to be performed, internally and externally. However, in most studies, the AI algorithm was developed using a limited dataset of only 1 hospital and none of the studies mention conducting external validation. Successful implementation may be achieved when more focus is given on the process of implementation in practice, rather than focusing solely on statistical performance measures.^{43,45} Also, the integration of the AI-based models into EHR systems to generate optimized medication alerts is challenging due to interoperability issues that often arise with such integrations.^{46,48–50}

None of the included studies mentioned information about model development and validation. This limits transparency and replicability of the research and makes it difficult to

Table 4. Results, outcomes, and statistical performance measures reported.

Author	Results	Outcomes	Performance measures
Segal et al ⁴²	Number of alerts as % of prescriptions: 37.10 for legacy CDS compared to 0.40 for Medaware. Additionally, the Medaware system performed surveillance on medication after prescribing, where the legacy CDS did not, comprising 60% of alerts.	% of alerts Clinically relevant Caused change in practice 5.30 Post prescribing surveillance ^a 0	Not reported
Schiff et al ³⁶	See clinical outcomes	Total of clinically valid alerts was 76.2% (96/126) of which 25% had less value, 18.8% had medium clinical value, and 56.2% had high clinical value.	Not reported
Hogue et al ³⁸	See performance measures	Not reported	A: identifying atypical medication orders with AI B: identifying pharmacological profiles with AI A PPV (%) 35 NPV (%) 96 Sensitivity (%) 26 Specificity (%) 97 AUROC 0.80 AUPR 0.25 F1 0.30 Prescription error detection PPV (%) 73 Sensitivity (%) 81 F1 0.76 B PPV (%) 49 NPV (%) 93 Sensitivity (%) 75 Specificity (%) 82 AUROC 0.88 AUPR 0.60 F1 0.59
Lee et al ⁴⁰	Number of alerts per month was reduced from 1613 (legacy CDSS) to 158 alerts (hybrid system)	Not reported	
Kawazoe et al ³⁵	Correlation coefficient = 0.80-0.95 at its highest for RF. ^b The results indicate that when predicted thresholds for medication alerts are applied, in 4 drugs a reduction in alert volume will occur and in 4 drugs an increase will occur. In total, the predicted thresholds will reduce the alerts by half using the static threshold.	Not reported	
Liu et al ⁴¹	See clinical outcomes and performance measures	The proportion of alerts predicted as being non successful and thereby filtered out was 25 for LR, 14 for SVM, 31 for NN, 39 for RF, and 54 for GBT.	Predicting user responses to medication alerts LR SVM NN RF Probability threshold 0.02 0.04 0.01 0.15 PPV ^c (%) 12 10 13 15 F1 ^c 0.21 0.19 0.23 0.25 Predicting physicians response to CDSS alerts NN RF NB GBT PPV (%) 88 87 83 95 NPV (%) 82 85 84 67 Sensitivity (%) 87 88 87 79 Specificity (%) 83 82 79 90 Accuracy (%) 89 86 84 83 F1 0.87 0.87 0.85 0.86 SVM 100 0.54 57 100 58 0.73
Poly et al ³⁷	See performance measures	Not reported	
Cornry et al ⁴³	The algorithm outperformed classic systems in its capacity to both detect patients with a medication error and to limit the number of false alerts. Of the remaining 26% prescription orders that required pharmacist intervention (false negatives) that were not intercepted by the algorithm, none were life-threatening.	Not reported	Identifying prescribing errors CDSS PPV (%) 54 Sensitivity (%) 69 AUROC (95% CI) 0.65 (0.61-0.69) AUPR (95% CI) 0.56 (0.50-0.62) F1 0.61 Lumio Medication 74 74 0.81 (0.78-0.84) 0.75 (0.70-0.80) 0.74

(continued)

Table 4. (continued)

Author	Results	Outcomes	Performance measures
Balestra et al ³⁹	See performance measures	Not reported	Identifying orders requiring intervention PPV (%) 9 Sensitivity (%) 99 Specificity (%) 37 Accuracy (%) 41 AUROC 0.908 AUPR 0.439
Beaudoin et al ³⁴	Reviewed prescriptions Prescriptions triggering alerts Rate of actionable alerts (%) The 17 prescriptions identified by the learning module included the 5 prescriptions that were missed by the knowledge base of the baseline system.	B+L 270 B 43 L 105 17 16 16 Whereas the baseline system should be valued by its ability to identify a high proportion of inappropriate prescriptions (high recall) followed by its ability to trigger few false alerts (high precision), the learning module is valued by its ability to discover clinically relevant rules that complement those of the baseline system and to trigger actionable alerts.	Identifying confirmed inappropriate prescriptions B+L 74 (68-79) Sensitivity 96 (92-98) Accuracy 79 (74-83) Baseline Learning 82 (77-87) 62 (52-71) 94 (90-97) 31 (25-38) 85 (81-89) 51 (46-56)

Abbreviations: PPV = positive predictive value (=precision), NPV = negative predictive value, Sensitivity = recall, AUROC = area under the receiver-operator curve, AUPR = area under the precision-recall, F1 = a metric in that balances precision (or PPV) and recall (or sensitivity), providing a value between 0 and 1, Correlation coefficient = coefficient between predicted doses and the actual doses, Probability threshold = a predefined value used in a classification model to decide the predicted class of an observation, with the model assigning the observation to one class if its predicted probability surpasses the threshold and to the other class otherwise. RF = random forest, LR = linear regression, SVM = support vector machine, NN = neural network, GBT = gradient boosted trees, NB = naive Bayes, B = baseline, L = learning.

^a Post prescribing surveillance = monitoring changing in the patient record to identify potential ADEs.

^b RF performed better than the BAG and CART algorithm, all correlation coefficients are shown in figures in the original manuscript.

^c Sensitivity value was set to ≥99%.

assess the risk of bias and potential usefulness of the prediction model. This reflects a wider trend in reporting of clinical prediction model studies which has already been recognized for several years.⁵¹ Additionally, none of the included studies reported applying reporting guidelines such as Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) Statement,⁵² which may explain why the relevant information about models development and validation was often missing. Adhering to the TRIPOD statement would ensure that enough information is provided for the reader to fully understand how the model was developed and validated. This transparency would support further research in the use of AI to improve CDSSs.

Strengths, limitations, and future perspectives

Strengths of this scoping review include the comprehensive search in different electronic databases and the screening of the references of relevant studies together with forward citation. Furthermore, the title and abstract screening was validated by a second screener, to maintain consistency in the inclusion of studies. Additionally, the methodology and execution of this scoping review was structurally assessed since it was conducted in accordance with the JBI methodology for scoping review and the PRISMA-ScR.^{26,27}

Limitations include the restriction to only include peer-reviewed studies. Non peer-reviewed studies may provide a more recent overview of the current state of AI-based methods for optimization of medication alerts generated by CDSS. On the other hand, novel methods for CDSSs in clinical settings must be assessed via rigorously peer-reviewed studies. Another limitation is the lack of critical appraisal of the included studies, but this is in accordance with the guidelines on scoping reviews we used.^{26,27} Also, given the high heterogeneity of studies included in terms of setting, methods and outcomes studies, a critical appraisal is of limited value. Furthermore, the utilization of ASReview may have resulted in missing relevant studies. Nonetheless, the inherent techniques of ASReview and the applied methodologies as described have mitigated this likelihood to a minimum. Additionally, these potentially missed studies were expected to be identified by forward citation. Finally, important contributions from Non-English speaking countries may have been missed.

Future studies on AI-based methods for optimization of medication alerts generated by CDSS in hospital settings should include larger datasets, to be able to extensively validate the models internally and externally. Prior to and during the development of CDSS using AI-based methods, it is essential to ensure sufficient support and collaboration with healthcare professionals, to facilitate trust, clinical value, and the implementation of the models in hospital practice. Moreover, such studies should use reporting guidelines, such as the upcoming ML focused TRIPOD (TRIPOD-ML) to enable critical appraisal of the results presented.⁵³ Once the models are developed and validated, adequate attention must be given to their implementation into hospital practice.

Conclusions

This scoping review provides an overview of the use of AI-based methods for optimizing medication alerts generated by CDSS in hospital setting. The then studies included show that AI has the capacity to adequately reduce alert burden and identify inappropriate prescriptions, but the datasets used

were relatively small and the models lack formal validation. Most have not been implemented in hospital practice. Future studies should focus on validation and implementation of AI-based methods, and use reporting guidelines such as the TRIPOD Statement to report their work.

Acknowledgments

We would like to express our gratitude to Ms. K. Sijsma from the central medical library at the University Medical Center Groningen for her assistance in developing the literature search strategy.

Author contributions

The initial project idea was conceived, managed, and supervised by Patricia M.L.A. van den Bemt. Jetske Graafsma and Rachel M. Murphy conducted the searches, screened the references, and extracted the data. Jetske Graafsma analyzed the extracted data and drafted the first version of the manuscript. All authors were involved in the writing of the manuscript.

Supplementary material

[Supplementary material](#) is available at *Journal of the American Medical Informatics Association* online.

Funding

This work was supported by the Dutch national Medicines Coordination Center (Landelijk Coördinatiecentrum Geneesmiddelen; LCG). The funders had no role in the study design, data collection and analysis, or preparation of the manuscript.

Conflicts of interest

R.H. is a CMO at GastonMedical which is a software company that develops medical decision support systems. The other authors report no competing interest.

Data availability

No new data were generated in support of this research. All relevant data are presented in the article.

References

1. Bates DW, Larizgoitia I, Prasopa-Plaizier N, Jha AK; Research Priority Setting Working Group of the WHO World Alliance for Patient Safety. Global priorities for patient safety research. *BMJ*. 2009;338:b1775.
2. Leendertse AJ, Egberts ACG, Stoker LJ, van den Bemt PMLA; HARM Study Group. Frequency of and risk factors for preventable medication-related hospital admissions in the Netherlands. *Arch Intern Med*. 2008;168(17):1890-1896.
3. Lghoul-Oulad Saïd F, Hek K, Flinterman LE, et al. Prevalence and incidence rate of hospital admissions related to medication between 2008 and 2013 in The Netherlands. *Pharmacoepidemiol Drug Saf*. 2020;29(12):1659-1668.
4. Laatikainen O, Miettinen J, Sneek S, Lehtiniemi H, Tenhunen O, Turpeinen M. The prevalence of medication-related adverse events

- in inpatients—a systematic review and meta-analysis. *Eur J Clin Pharmacol*. 2017;73(12):1539-1549.
5. Bates DW, Levine DM, Salmasian H, et al. The safety of inpatient health care. *N Engl J Med*. 2023;388(2):142-153.
 6. Warlé-van Herwaarden MF, Kramers C, Sturkenboom MC, van den Bemt PMLA, De Smet PAGM; Dutch HARM-Wrestling Task Force. Targeting outpatient drug safety: recommendations of the Dutch HARM-Wrestling Task Force. *Drug Saf*. 2012;35(3):245-259.
 7. Eppenga WL, Derijks HJ, Conemans JMH, Hermens WAJJ, Wensing M, De Smet PAGM. Comparison of a basic and an advanced pharmacotherapy-related clinical decision support system in a hospital care setting in The Netherlands. *J Am Med Inform Assoc*. 2012;19(1):66-71.
 8. Reese TJ, Kawamoto K, Fiol GD, et al. When an alert is not an alert: a pilot study to characterize behavior and cognition associated with medication alerts. *AMIA Annu Symp Proc*. 2018;2018:1488-1497.
 9. de Wit HAJM, Mestres Gonzalvo C, Cardenas J, et al. Evaluation of clinical rules in a standalone pharmacy based clinical decision support system for hospitalized and nursing home patients. *Int J Med Inform*. 2015;84(6):396-405.
 10. Helmons PJ, Suijkerbuijk BO, Nannan Panday PV, Kosterink JGW. Drug-drug interaction checking assisted by clinical decision support: a return on investment analysis. *J Am Med Inform Assoc*. 2015;22(4):764-772.
 11. Poly TN, Islam MM, Yang H, Li YJ. Appropriateness of overridden alerts in computerized physician order entry: systematic review. *JMIR Med Inform*. 2020;8(7):e15653.
 12. 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.
 13. Wasylewicz ATM, Scheepers-Hoeks AMJW. Clinical decision support systems. In: Kubben P, Dumontier M, Dekker A, eds. *Fundamentals of Clinical Data Science*. The Author(s); 2019:153-169.
 14. Ledger TS, Brooke-Cowden K, Coiera E. Post-implementation optimization of medication alerts in hospital computerized provider order entry systems: a scoping review. *J Am Med Inform Assoc*. 2023;30(12):2064-2071.
 15. Bakker T, Klopotoska JE, Dongelmans DA, et al.; SIMPLIFY Study Group. The effect of computerised decision support alerts tailored to intensive care on the administration of high-risk drug combinations, and their monitoring: a cluster randomised stepped-wedge trial. *Lancet*. 2024;403(10425):439-449.
 16. Van Dort BA, Zheng WY, Sundar V, Baysari MT. Optimizing clinical decision support alerts in electronic medical records: a systematic review of reported strategies adopted by hospitals. *J Am Med Inform Assoc*. 2021;28(1):177-183.
 17. Wasylewicz ATM, van de Burgt BWM, Manten T, et al. Contextualized drug-drug interaction management improves clinical utility compared with basic drug-drug interaction management in hospitalized patients. *Clin Pharmacol Ther*. 2022;112(2):382-390.
 18. Sutton RT, Pincok 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;3:17. eCollection 2020.
 19. van der Sijs H, Aarts J, van Gelder T, Berg M, Vulto A. Turning off frequently overridden drug alerts: limited opportunities for doing it safely. *J Am Med Inform Assoc*. 2008;15(4):439-448.
 20. Liu X, Cruz Rivera S, Moher D, Calvert MJ, Denniston AK; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med*. 2020;26(9):1364-1374.
 21. Bohr A, Memarzadeh K. Chapter 2 - The rise of artificial intelligence in healthcare applications. In: Bohr A, Memarzadeh K, eds. *Artificial Intelligence in Healthcare*. Academic Press; 2020:25-60.
 22. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2(4):230-243.
 23. Oehring R, Ramasetti N, Ng S, et al. Use and accuracy of decision support systems using artificial intelligence for tumor diseases: a systematic review and meta-analysis. *Front Oncol*. 2023;13:1224347.
 24. Machoy ME, Szyszka-Sommerfeld L, Vegh A, Gedrange T, Woźniak K. The ways of using machine learning in dentistry. *Adv Clin Exp Med*. 2020;29(3):375-384.
 25. Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. *Clin Microbiol Infect*. 2020;26(5):584-595.
 26. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc*. 2015;13(3):141-146.
 27. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-473.
 28. Graafsma J, Murphy RM, Klopotoska JE, et al. The use of artificial intelligence to improve medication-related clinical decision support systems: a scoping review protocol. *Open Science Framework*. 2023.
 29. Kanjanarat P, Winterstein AG, Johns TE, Hatton RC, Gonzalez-Rothi R, Segal R. Nature of preventable adverse drug events in hospitals: a literature review. *Am J Health Syst Pharm*. 2003;160(17):1750-1759.
 30. Thomsen LA, Winterstein AG, Søndergaard B, Haugbølle LS, Melander A. Systematic review of the incidence and characteristics of preventable adverse drug events in ambulatory care. *Ann Pharmacother*. 2007;41(9):1411-1426.
 31. ASReview LAB Developers. 2024. ASReview LAB—a tool for AI-assisted systematic review (v1.5). Zenodo.
 32. van de Schoot R, de Bruin J, Schram R, et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat Mach Intell*. 2021;3(2):125-133.
 33. Wang Z, Nayfeh T, Tetzlaff J, O'Brien P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. *PLoS One*. 2020;15(1):e0227742.
 34. Beaudoin M, Kabanza F, Nault V, Valiquette L. Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. *Artif Intell Med*. 2016;68:29-36.
 35. Kawazoe Y, Miyo K, Kurahashi I, Sakurai R, Ohe K. Prediction-based threshold for medication alert. *Stud Health Technol Inform*. 2013;192:229-233.
 36. Schiff GD, Volk LA, Volodarskaya M, et al. Screening for medication errors using an outlier detection system. *J Am Med Inform Assoc*. 2017;24(2):281-287.
 37. Poly TN, Islam MM, Muhtar MS, Yang H, Nguyen PAA, Li YJ. 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.
 38. Hogue S-C, Chen F, Brassard G, et al. Pharmacists' perceptions of a machine learning model for the identification of atypical medication orders. *J Am Med Inform Assoc*. 2021;28(8):1712-1718.
 39. Balestra M, Chen J, Iturrate E, Aphinyanaphongs Y, Nov O. Predicting inpatient pharmacy order interventions using provider action data. *JAMIA Open*. 2021;4(3):ooab083.
 40. Lee S, Shin J, Kim HS, et al. Hybrid method incorporating a rule-based approach and deep learning for prescription error prediction. *Drug Saf*. 2022;45(1):27-35.
 41. Liu S, Kawamoto K, Del Fiol G, et al. The potential for leveraging machine learning to filter medication alerts. *J Am Med Inform Assoc*. 2022;29(5):891-899.
 42. Segal G, Segev A, Brom A, Lifshitz Y, Wasserstrum Y, Zimlichman E. Reducing drug prescription errors and adverse drug events by application of a probabilistic, machine-learning based clinical

- decision support system in an inpatient setting. *J Am Med Inform Assoc.* 2019;26(12):1560-1565.
43. Corny J, Rajkumar A, Martin O, et al. A machine learning-based clinical decision support system to identify prescriptions with a high risk of medication error. *J Am Med Inform Assoc.* 2020;27(11):1688-1694.
 44. Ancker JS, Edwards A, Nosal S, Hauser D, Mauer E, Kaushal R, with the HITEC Investigators. Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Med Inform Decis Mak.* 2017;17(1):36-38.
 45. 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.
 46. Chalasani SH, Syed J, Ramesh M, Patil V, Pramod Kumar TM. Artificial intelligence in the field of pharmacy practice: a literature review. *Explor Res Clin Soc Pharm.* 2023;12:100346.
 47. Sharma M, Savage C, Nair M, Larsson I, Svedberg P, Nygren JM. Artificial intelligence applications in health care practice: scoping review. *J Med Internet Res.* 2022;24(10):e40238.
 48. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195.
 49. Belenguer L. AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI Ethics.* 2022;2(4):771-787.
 50. Parliament E, Directorate-General For Parliamentary RS, Lekadir K, Quaglio G, Tselioudis Garmendia A, Gallin C. *Artificial Intelligence in Healthcare—Applications, Risks, and Ethical and Societal Impacts.* European Parliament; 2022.
 51. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med.* 2015;162(1):55-63.
 52. Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. *Lancet.* 2019;393(10181):1577-1579.
 53. Bouwmeester W, Zuithoff NPA, Mallett S, et al. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med.* 2012;9(5):1-12.