

Artificial Intelligence-Based Clinical Decision Support Systems in Nursing Practice: Applications, Benefits, and Challenges — A Systematic Review

Idris M.^{1*}, Bamigboye T.O.², Oniyide A.A.³

¹ Department of Nursing, College of Medicine and Allied Sciences, Afe Babalola University, Ado-Ekiti, Nigeria

² Department of Nursing, Afe Babalola University, Ado-Ekiti, Nigeria

³ Neuroscience Research Unit, Department of Physiology, College of Medicine and Health Sciences, Afe Babalola University, Ado-Ekiti, Nigeria

DOI: <https://doi.org/10.51584/IJRIAS.2026.11080184>

Received: 06 September 2026; Accepted: 11 September 2026; Published: 22 September 2026

ABSTRACT

Artificial intelligence (AI)-based clinical decision support systems (CDSS) are increasingly embedded in nursing workflows, promising earlier detection of patient deterioration, more efficient documentation, and more consistent risk stratification, yet the evidence base is fragmented across clinical domains and ethical, workforce, and implementation concerns remain insufficiently synthesised from a nursing-specific perspective. This review therefore aimed to systematically synthesise current evidence on the applications, benefits, and challenges of AI-based CDSS in nursing practice, and to derive implications for nursing education, regulation, and future research. A systematic literature search conceptually following PRISMA 2020 reporting guidance was conducted across five databases (PubMed/MEDLINE, CINAHL, Scopus, Web of Science, and IEEE Xplore), supplemented by backward citation tracking of nine recent systematic or scoping reviews, covering publications largely from 2018 to 2026; findings were synthesised narratively across four application domains.

Thirty-eight unique sources met the inclusion criteria. AI-CDSS applications clustered into four domains: deterioration and sepsis early-warning systems; pressure injury and fall-risk prediction; natural language processing and generative AI for nursing documentation and care planning; and workflow, triage, and staffing optimisation. Reported benefits included earlier deterioration detection, reduced false-alarm rates, reduced documentation burden, and more standardised risk assessment, while reported challenges included algorithmic bias, alert fatigue, automation bias, deskilling risk, accountability ambiguity, integration barriers, and persistent AI-literacy gaps. AI-based CDSS therefore hold meaningful, evidence-supported potential to augment nursing practice, but current evidence is dominated by single-site, retrospective validation studies; realising this potential will require nurse involvement in design and governance, explainable and auditable algorithms, bias-mitigation standards, formal accountability frameworks, and integration of AI literacy into nursing curricula.

Keywords: artificial intelligence, clinical decision support systems, nursing practice, machine learning, nursing informatics, algorithmic bias, patient safety, systematic review

INTRODUCTION

Clinical decision support systems (CDSS) are computerised tools designed to enhance clinical decision-making by matching patient-specific information against a knowledge base or learned model to generate alerts, risk scores, or recommendations at the point of care (Musen et al., 2021). The integration of artificial

intelligence (AI) — including machine learning, deep learning, and large language models — into CDSS marks a qualitative shift from rule-based logic toward probabilistic, data-driven inference capable of detecting complex, non-linear patterns across large volumes of electronic health record data (Ikegwu et al., 2025). Nursing occupies a structurally important position in this ecosystem: nurses generate and act upon the majority of bedside physiological, behavioural, and risk-assessment data captured in the electronic health record, and are typically the first responders to algorithm-generated alerts for deterioration, sepsis, falls, and pressure injury (Yang et al., 2026). Despite this central role, nurses have historically been underrepresented in the design, validation, and governance of the AI tools deployed in their own workflows, a gap repeatedly identified in recent nursing-specific systematic reviews (Martinez-Ortigosa et al., 2023; Mikkonen et al., 2026).

Operational definitions. For this review, a clinical decision support system (CDSS) is any computerised tool that matches patient-specific information against a knowledge base or model to generate alerts, risk scores, or recommendations at the point of care, irrespective of whether that matching logic is rule-based or learned (Musen et al., 2021). A tool was classified as AI-based for the purposes of this review only where its decision logic was derived, at least in part, from a trained statistical or machine-learning model — including classical machine learning (e.g., logistic regression, random forest, gradient boosting), deep learning, or large language models — rather than from fixed, expert-authored if-then rules. This distinction matters because legacy rule-based CDSS (e.g., static early-warning scores, the Braden Scale, the STRATIFY tool) served as comparators rather than eligible interventions in this review (see Table 1), and because the probabilistic, data-driven nature of AI-based tools is itself a source of several challenges discussed in Section 3.7 (opacity, drift, and bias) that do not apply, or apply differently, to fixed-rule systems.

AI-based CDSS in nursing now span a wide range of applications: early warning systems for sepsis and deterioration (Gupta et al., 2024; Ginestra et al., 2019); supervised machine learning models for pressure injury and fall-risk prediction (Anderson et al., 2021; Song et al., 2021; Chen and Xu, 2023); natural language processing tools embedded in electronic health records for documentation (Nashwan and Abujaber, 2023); and, increasingly, generative AI and large language models for nursing care-plan drafting and patient communication (Nashwan and Bani Hani, 2023). Parallel to this technical expansion, a body of literature has emerged examining ethical, workforce, and implementation dimensions, including algorithmic bias (Panch et al., 2019; Norori et al., 2021), automation bias and over-reliance (Gaube et al., 2021), and nurses' AI literacy and readiness (Fernandes et al., 2026). This review aimed to answer four questions: (1) what are the principal applications of AI-based CDSS currently reported in nursing practice; (2) what benefits to patient safety, nursing workflow, and clinical decision-making are reported for these systems; (3) what clinical, ethical, organisational, and educational challenges constrain their safe and effective adoption in nursing; and (4) what implications follow for nursing practice, education, and policy.

MATERIALS AND METHODS

This systematic review was conceptually guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement for transparent reporting of search, screening, and synthesis processes. As a review of previously published literature, it did not involve new human or animal subject research, and formal ethical approval was therefore not required.

Eligibility Criteria

Eligible records (a) addressed an AI-, machine learning-, deep learning-, natural language processing-, or large language model-based tool functioning as, or directly informing, a clinical decision support function; (b) reported outcomes relevant to nursing practice; and (c) were peer-reviewed primary studies, systematic, scoping, or integrative reviews, or substantive editorials/commentaries published in English. The complete PICOS (Population, Intervention, Comparator, Outcomes, Study type) framework is presented in Table 1.

Table 1. PICOS eligibility criteria used for study selection.

PICOS element	Criteria
Population	Nurses, nurse practitioners, nursing students, or patient populations for whom nursing-sensitive outcomes (falls, pressure injury, deterioration, documentation) were reported
Intervention	AI-, machine learning-, deep learning-, natural language processing-, or large language model-based tool functioning as, or directly informing, a clinical decision support function in a nursing workflow
Comparator	Standard of care, rule-based CDSS, legacy risk scales (Braden, STRATIFY, Hendrich II), or no comparator (descriptive/perception studies)
Outcomes	Nurse-facing clinical outcomes (e.g., pressure injury incidence, falls, sepsis detection time), nursing workflow metrics (documentation time, alert rates), or nurses' perceptions and AI readiness
Study types	Peer-reviewed primary studies (algorithm validation, randomised controlled trials, cohort, cross-sectional), systematic, scoping, or integrative reviews, and substantive editorials/commentaries
Language	English only
Publication years	2018–2026 (reflecting mature machine-learning CDSS deployment; 2022 onward for generative AI applications)
Exclusions	AI without a decision-support function; AI restricted to radiology/surgical robotics with no nursing interface; conference abstracts; non-peer-reviewed preprints; grey literature without identifiable peer review

Information Sources and Search Strategy

A systematic search was conducted across five bibliographic databases: PubMed/MEDLINE, CINAHL, Scopus, Web of Science, and IEEE Xplore. Boolean combinations of controlled vocabulary (MeSH/CINAHL headings) and free-text terms were applied. Core search concepts included (“artificial intelligence” OR “machine learning” OR “deep learning” OR “natural language processing” OR “large language model”) AND (“clinical decision support” OR “CDSS”) AND (“nurs*” OR “nurse”), combined with domain-specific terms including “deterioration,” “sepsis,” “pressure injury,” “fall risk,” “care plan,” and “documentation.” The search was supplemented by backward citation tracking of nine recently published systematic, scoping, and integrative reviews on AI in nursing published between 2021 and 2026. Table 2 provides the complete search strategy by database.

Table 2. Database search strategy, date ranges, and records retrieved per source.

Database	Search string (representative)	Date range	Records retrieved
PubMed/MEDLINE	(“artificial intelligence” OR “machine learning” OR “deep learning” OR “natural language processing” OR “large	2018–2026	38

	language model”) AND (“clinical decision support” OR “CDSS”) AND (“nurs*” OR “nurse”)		
CINAHL	(“artificial intelligence” OR “machine learning”) AND (“clinical decision support system” OR “CDSS”) AND (“nurs*”) AND (TI “deterioration” OR “sepsis” OR “pressure injury” OR “fall risk” OR “care plan” OR “documentation”)	2018–2026	29
Scopus	TITLE-ABS-KEY(“AI” OR “machine learning” OR “deep learning”) AND (“clinical decision support”) AND (“nursing” OR “nurse”) AND (“algorithm” OR “model” OR “prediction”)	2018–2026	28
Web of Science	TS=(“artificial intelligence” OR “NLP” OR “LLM” OR “neural network”) AND TS=(“decision support” AND “nurs*”) AND TS=(“sepsis” OR “pressure ulcer” OR “fall” OR “deterioration” OR “workflow”)	2018–2026	22
IEEE Xplore	“clinical decision support” AND (“nursing” OR “nurse”) AND (“machine learning” OR “artificial intelligence” OR “natural language processing”)	2018–2026	11
Citation tracking (9 reviews)	Backward citation tracking of included systematic, scoping, and integrative reviews (Martinez-Ortigosa et al., 2023; Mikkonen et al., 2026; Alenazi et al., 2024; Fernandes et al., 2026; El Arab et al., 2025; and four others)	2018–2026	19
Total (before deduplication)			147

Study Selection and PRISMA Flow

Titles and abstracts were screened against the eligibility criteria, followed by full-text assessment of potentially eligible records. **Figure 1** (PRISMA 2020 flow diagram) summarises the screening and selection process: 128 records were identified through database searching (PubMed/MEDLINE 38; CINAHL 29; Scopus 28; Web of Science 22; IEEE Xplore 11), plus 19 additional records via citation tracking of nine systematic/scoping reviews; after removal of duplicates, 94 records remained, of which 33 were excluded at title/abstract screening (no AI/machine-learning-based CDSS function; not a nursing-relevant domain; outside the publication-date range); 61 full-text articles were assessed for eligibility, of which 23 were excluded (no nursing-specific outcome, $n = 12$; not peer-reviewed, $n = 6$; no decision-support function, $n = 5$); 38 studies were included in the qualitative synthesis (26 primary studies and 12 reviews/editorials), comprising 16 algorithm development/validation studies, 10 nurse-perception surveys, 7 systematic/scoping reviews, and 5 integrative reviews/editorials.

For each included study, the following data were extracted: author(s) and year, country/setting, AI/machine-learning method, clinical or operational domain, sample or dataset characteristics, nurse-relevant outcomes, and reported benefits and limitations. Findings were synthesised narratively using a framework-synthesis approach, organised around four recurring application domains identified inductively during extraction, followed by a cross-cutting thematic synthesis of benefits and challenges.

Quality Appraisal

Primary algorithm-development/validation studies were appraised qualitatively for reporting of external validation, comparator/baseline performance, and prospective versus retrospective design. Survey-based studies were appraised for sampling method and response rate. Included systematic/scoping reviews were appraised for use of a recognised reporting framework (PRISMA or Joanna Briggs Institute [JBI] methodology). No study was excluded solely on quality grounds, consistent with the exploratory, practice-oriented aim of this review; methodological limitations are instead flagged throughout the synthesis. A full retrospective JBI-based critical appraisal is provided in **Supplementary Table S1**.

RESULTS

Overview of Included Studies

Thirty-eight sources met the inclusion criteria: 26 primary studies and 12 systematic, scoping, or integrative reviews/editorials. Studies originated predominantly from acute and critical care settings in the United States, Taiwan, Japan, Saudi Arabia, Qatar, and several European countries. Machine-learning methods included logistic regression, random forest, gradient boosting (XGBoost), support vector machines, long short-term memory (LSTM) networks, and GPT-4-class large language models. Tables 4-6 summarised the distribution of included studies by domain, study design, and geography, respectively; Table 3 provides a full summary of the 16 individually named studies (the remaining 22 sources are synthesised in aggregate throughout Sections 3.2–3.7, consistent with the level of individual detail reported for them in extraction).

Table 3. Characteristics and key findings of individually named included sources (n = 16 of 38; remaining sources synthesised in aggregate).

Author(s) & year	Country	AI/ML method	Domain	Sample/dataset	Key outcomes	Main findings
Gupta et al. (2024)	USA	LSTM deep learning	Deterioration/sepsis	ICU (>40,000 records, 2 health systems)	Sensitivity, specificity, false-alarm rate	4-6 h warning lead time; 3.18% false-alarm ratio
Ginestra et al. (2019)	USA	Machine learning EWS	Deterioration/sepsis	Multi-hospital (bedside nursing)	Clinician perception; alert workload	Positive likelihood ratio 13; ~10 alerts/day system-wide
Alenazi et al. (2024)	Saudi Arabia	Systematic review (multiple ML)	Deterioration/sepsis	Various (review)	Diagnostic accuracy, transparency	Improved accuracy; black-box barrier identified
Fernandes et al. (2026)	Portugal	Scoping review	Deterioration/workflow	Various (review)	Workflow efficiency, accuracy	Improved accuracy; interpretability gap noted

Anderson et al. (2021)	USA	Multiple models (ML (RF, SVM, LR)	Pressure injury	Acute care hospitals (EHR data)	AUROC vs. Braden Scale	ML outperformed Braden; earlier identification
Song et al. (2021)	USA	ML classifiers on nursing phenotypes	Pressure injury	Acute care (structured nursing documentation)	AUROC 0.74–0.99	Strong discrimination from routine nursing data
Chang et al. (2025)	Taiwan	Probability feature aggregation	Pressure injury	Admission patients (low-complexity inputs)	Early prediction accuracy	Prediction on admission without laboratory data
Chen and Xu (2023)	Taiwan	XGBoost	Fall risk	Inpatient (>53,000 records)	Accuracy 95.1%, AUROC 0.990	Outperformed STRATIFY and Hendrich II
Nashwan and Abujaber (2023)	Qatar	GPT-4-class LLM	Documentation/care planning	Simulated nursing care planning	Ethical risk catalogue, use-case mapping	Efficiency gain; privacy and bias risks catalogued
Nashwan and Bani Hani (2023)	Qatar	LLM (GPT-4-class)	Documentation/care planning	Oncology nursing (simulated)	Care-plan completeness	Draft care plans synthesised from EHR data
Rabbani et al. (2025)	UAE	Generative AI (virtual-patient simulation)	Documentation	Virtual-patient simulation (nurses)	Efficiency, usability ratings	Efficiency gain; revision needed, not verbatim use
Hernandez and Bianchi (2025)	Multi-country	AI-CDSS (multiple)	Workflow/triage	Hospital settings (review)	Efficiency, patient outcomes	Reduced complications; length-of-stay reduction noted
Panch et al. (2019)	Global	Algorithmic audit/commentary	Bias/ethics	Health systems globally	Bias mechanism taxonomy	Systematic bias from skewed training data

Norori et al. (2021)	Global	Commentary/open-science analysis	Bias/ethics	Healthcare AI broadly	Bias sources, mitigation strategies	Open-science practices proposed as mitigation
Mikkonen et al. (2026)	Finland	Systematic review	Nurse AI readiness	Nursing practice (multiple studies)	AI support for clinical decision-making	Need for nurse involvement in design/governance
El Arab et al. (2025)	Multi-country	Systematic review	AI literacy/attitudes	Nursing students and practising nurses	AI attitudes, literacy, readiness	Large literacy gap; formal curricula needed

Table 4. Distribution of included studies by application domain, primary AI/ML technique, and primary target user (n = 38).

Application domain	Studies (n)	Primary AI/ML technique(s)	Primary target user(s)	Description
Deterioration and sepsis early-warning systems	10	Deep learning (LSTM); ML ensembles	Bedside/ICU nurses; rapid-response teams	Studies focused on early warning for clinical deterioration and sepsis
Pressure injury and fall risk	11	Classical ML (random forest, XGBoost, logistic regression)	Bedside nurses; wound-care/skin-integrity specialists	Studies focused on pressure injury and fall-risk prediction
NLP, documentation, and generative AI	9	NLP; generative AI/LLMs (GPT-4-class)	Bedside nurses; nursing-documentation/care-planning staff	Studies on documentation, NLP, and generative AI care planning
Workflow, triage, and staffing	5	ML-based workflow/triage models	Charge nurses; nurse managers/schedulers	Studies on workflow, triage, and staffing optimisation
Cross-cutting (bias, ethics, literacy)	3	Cross-cutting (multiple ML/LLM types)	All nursing staff; nurse educators/leaders/informaticists	Cross-cutting bias, ethics, and literacy studies

Note. Study counts reflect each study's primary application domain; some cross-cutting reviews are counted in the domain they primarily address. AI/ML technique and target user reflect the modal (most common) values reported across studies in that domain; individual studies within a domain sometimes used more than one technique or served more than one user group (see Table 3 for study-level detail).

Table 5. Distribution of included studies by study design (n = 38).

Study design	n (%)
Algorithm development/validation (retrospective)	16 (42%)
Nurse perception/survey	10 (26%)
Systematic/scoping review	7 (18%)
Integrative review/editorial	5 (13%)

Table 6. Geographic distribution of included primary studies.

Country/region	Studies (n)
United States	10
Taiwan	3
Qatar	2
Saudi Arabia	2
Portugal	1
Finland	1
Multiple/international	8
Not specified/simulated	11

Note. “Not specified/simulated” includes virtual-patient environments and multi-country reviews without a single host country.

Domain 1: Deterioration and Sepsis Early-Warning Systems

Sepsis and patient-deterioration prediction represent the most mature application of AI-CDSS relevant to nursing surveillance. Gupta et al. (2024) developed SepsisAI, an LSTM-based deep-learning model trained on more than 40,000 ICU patient records from two US health systems, which issued deterioration warnings at a median of six hours and alerts at a median of four hours before clinical sepsis onset, with a false-alarm ratio of 3.18% — substantially lower than legacy rule-based sepsis alerts.

Ginestra et al. (2019) prospectively evaluated clinician (including bedside nursing) perceptions of a machine-learning-based early warning system (EWS 2.0) for severe sepsis and septic shock deployed across a multi-hospital US health system. The model achieved a positive likelihood ratio of 13 for severe sepsis/septic shock, with the alert threshold calibrated to produce approximately ten alerts per day across the system. Clinician acceptance was shaped less by raw model accuracy than by perceived workload burden, alert transparency, and trust — an early empirical demonstration that the human-factors dimension of AI-CDSS deployment is as consequential as algorithmic performance.

Domain 2: Pressure Injury and Fall-Risk Prediction

Anderson et al. (2021) demonstrated that machine-learning models leveraging electronic health record and nursing-flowsheet data could predict hospital-acquired pressure injury risk with performance exceeding traditional scales such as the Braden Scale. Song et al. (2021) extended this work using nursing-assessment “phenotypes” derived from structured nursing documentation, reporting that machine-learning classifiers achieved area-under-the-receiver-operating-characteristic-curve (AUROC) values in the 0.74–0.99 range, enabling earlier identification of at-risk patients than manual risk-scale completion alone.

For fall risk, Chen and Xu (2023) applied an XGBoost model to more than 53,000 inpatient records from a Taiwanese hospital, achieving a cross-validated accuracy of 95.1% and AUROC of 0.990, substantially outperforming conventional manual tools such as the STRATIFY scale and the Hendrich II Fall Risk Model. The comparative performance of AI models across these domains is summarised in Table 7.

Table 7. Predictive performance metrics for key AI-CDSS models identified in this review.

Model/study	Algorithm	Setting	Sample	LR +	AUROC	Accuracy	Alert metric	Key advantage/design
Sepsis detection (Sepsis AI)	LSTM	ICU	>40,000	—	~96.8%	—	3.18% false-alarm ratio	4–6 h lead time (multi-site, prospective)
Sepsis/deterioration (EWS 2.0)	ML ensemble	Multi-hospital	Large	13.0	—	—	~10 alerts/day (calibrated)	Clinician-trust study (prospective perception)
Pressure injury (Anderson et al.)	RF/XGBoost/LR	Acute care	Hospital EHR	—	AUROC range	—	Superior to Braden Scale	Earlier identification (retrospective)
Pressure injury (Song et al.)	ML classifiers	Acute care	Nursing flowcharts	—	0.74–0.99	—	Routine nursing data sufficient	Phenotype-based (retrospective)
Early pressure injury (Chang et al.)	Probability feature aggregation	Admission	Admission data	—	Improved vs. baseline	—	No laboratory data needed	Low-complexity input (retrospective)
Fall risk (Chen and Xu)	XGBoost	Inpatient, Taiwan	53,000 +	—	0.990	95.1%	Outperforms STRATIFY/Hendrich II	High specificity (retrospective)

AUROC, area under the receiver operating characteristic curve; LR+, positive likelihood ratio; EWS, early warning system; —, not reported.

Domain 3: NLP, Documentation, and Generative AI Care Planning

Documentation has long been identified as one of the largest non-direct-care time burdens for nurses, and AI-enabled voice-recognition and predictive-text tools embedded in electronic health records have been reported to reduce time spent on routine charting and to improve standardisation of clinical terminology.

Nashwan and Abujaber (2023) examined the application of GPT-4-class large language models to nursing care planning, proposing use cases in patient-data assessment and interpretation, family communication support, and identification of gaps in existing care plans, while explicitly cataloguing ethical risks including data privacy, fairness and bias in generated recommendations, and the appropriate boundary between human and AI roles. A companion analysis applied this reasoning to oncology nursing care planning (Nashwan and Bani Hani, 2023). Importantly, subsequent bias-focused research found that large-language-model-generated nursing care plans varied systematically in length and in the frequency of physiological versus psychological care terms according to simulated patients' sociodemographic characteristics — despite identical underlying clinical scenarios.

Domain 4: Workflow, Triage, and Staffing Optimisation

A smaller but consistent body of evidence describes AI-CDSS applications supporting nursing workflow and resource allocation, including predictive workload classification, automated triage prioritisation, and care-coordination support (Hernandez and Bianchi, 2025). Reviewers caution that automated insights must be balanced against professional judgement to avoid over-reliance, and that operational efficiency gains have been more consistently documented than gains in patient-level clinical outcomes.

Cross-Cutting Synthesis: Benefits of AI-Based CDSS in Nursing

Across the four domains, five recurring categories of benefit were identified. Because improved technical performance does not automatically translate into safer or better care, Table 8 classifies each benefit by outcome type — technical (model-performance metrics such as AUROC or accuracy), process (workflow or workload measures such as documentation time or alert volume), clinician-facing (measures of clinician perception, trust, or behaviour), or patient (measures of patient-level harm or benefit) — alongside the underlying evidence and its strength.

Table 8. Cross-cutting synthesis of reported benefits from AI-CDSS in nursing, classified by outcome type.

Benefit category	Domain	Outcome type	Key sources	Evidence summary	Evidence strength
Earlier detection of deterioration	Deterioration/sepsis	Technical + process (lead-time/AUROC; no direct patient-outcome data reported)	Gupta et al. (2024) ; Ginestra et al. (2019) ; Alenazi et al. (2024)	AUROC >0.85; 4–6 h lead time over standard care	Strong (multiple prospective/retrospective studies)
Reduced false-positive alert rates	Deterioration/sepsis	Process (alert-burden/workload proxy, not a	Gupta et al. (2024) ; Ginestra	3.18% false-alarm ratio; ~10	Moderate (2 primary studies)

		patient outcome)	et al. (2019)	actionable alerts/day	
Improved pressure-injury risk prediction vs. Braden Scale	Pressure injury	Technical (predictive accuracy vs. legacy scale; pressure-injury incidence not reported)	Anderson et al. (2021) ; Song et al. (2021) ; Chang et al. (2025)	AUROC 0.74–0.99; earlier identification at admission	Strong (3 primary + 1 review)
Improved fall-risk prediction	Fall risk	Technical (predictive accuracy; fall-incidence reduction not reported)	Chen and Xu (2023)	Accuracy 95.1%, AUROC 0.990; outperforms STRATIFY and Hendrich II	Moderate (single-site, large n)
Reduced documentation burden	Documentation/NLP	Process (clinician time/workload)	Nashwan and Abujaber (2023); Rabbani et al. (2025)	Shorter charting time; standardised terminology	Emerging (simulation-based, limited prospective data)
Standardised, objective risk assessment	Multiple domains	Process/technical (assessment consistency; not validated against injury/fall incidence)	Song et al. (2021) ; Chen and Xu (2023)	Removes subjective variability of manual scales	Moderate (retrospective validation)
Operational/workforce efficiency	Workflow/triage	Process, with preliminary patient-level signal (length-of-stay reduction reported at review level only)	Hernandez and Bianchi (2025); Muñoz et al. (2026)	Reduced nursing workload; shorter ICU length of stay	Emerging (review-level, limited primary data)

Outcome type: Technical (model-performance metrics only, e.g., AUROC, accuracy); Process (workflow/workload measures, e.g., documentation time, alert volume); Clinician-facing (clinician perception, trust, or behaviour); Patient (patient-level harm/benefit measures). Most benefits in this table are technical or process outcomes; direct, prospectively measured patient-outcome evidence (e.g., confirmed reductions in pressure-injury or fall incidence, or in sepsis-related mortality) was not reported for any domain in the included literature and should not be inferred from predictive-accuracy metrics alone. Evidence strength: Strong (multiple independent primary studies); Moderate (single high-quality primary study or review); Emerging (simulation/preliminary evidence only).

Cross-Cutting Synthesis: Challenges and Barriers

Six recurring categories of challenge were identified across the included literature. Table 9 presents a structured summary with severity ratings and proposed mitigations.

Table 9. Cross-cutting synthesis of challenges identified in the included literature.

Challenge	Severity	Key sources	Evidence/manifestation	Proposed mitigation
Algorithmic bias and data representativeness	High	Panch et al. (2019); Norori et al. (2021); LLM care-plan simulation study	Skewed training data; LLM-generated care plans varied by patient sociodemographics	Participatory model development; inclusive data standards; bias auditing by sociodemographic group
Alert fatigue and automation bias	High	Ginestra et al. (2019); Gaube et al. (2021); Alenazi et al. (2024)	Over-reliance suppresses critical appraisal; disengagement from excessive false alarms	Calibrate alert thresholds with nursing input; explainability tools; regular audits
Deskilling/erosion of clinical judgement	Moderate–high	Mikkonen et al. (2026); multiple reviews	Risk of atrophied assessment skills when AI outputs are treated as authoritative	Position AI as advisory; maintain manual assessment competencies; simulation training
Accountability and transparency (“black box”)	High	Alenazi et al. (2024); Fernandes et al. (2026); Alderden et al. (2024)	Unclear responsibility when AI contributes to an adverse outcome; opaque model reasoning	Mandatory explainability; clear accountability frameworks in policy; governance committees
Validation gap (retrospective to clinical deployment)	High	Pressure-injury ICU scoping review; multiple domain reviews	Only 2/14 pressure-injury/ICU studies reported deployment outcomes; single-site generalisability risk	Require prospective, multi-site validation before procurement; nurse-informaticist oversight post-deployment
Privacy, cybersecurity, and regulatory ambiguity	Moderate	Nashwan and Abujaber (2023); Nashwan and Bani Hani (2023)	LLMs and AI-CDSS process sensitive unstructured patient data, creating both privacy exposure and a broader cybersecurity attack surface (e.g., data exfiltration, model manipulation) as AI tools are integrated with EHR infrastructure; regulatory	Data de-identification; GDPR/HIPAA-compliant architectures; cybersecurity controls (access management, audit logging, penetration testing) for AI-EHR

			frameworks lag the technology	integration; institutional AI governance policy
AI-literacy gap in the nursing workforce	High	El Arab et al. (2025); Chatzichristos et al. (2025); multiple surveys	Majority of nurses report self-directed AI learning; limited formal training	Integrate AI literacy into pre-licensure curricula; structured continuing professional development; simulation-based AI appraisal exercises

Severity: High = cited in ≥ 4 included sources; Moderate = cited in 2–3 sources.

DISCUSSION

This systematic review synthesised evidence on AI-based CDSS across four nursing-relevant domains — deterioration/sepsis early warning, pressure-injury and fall prediction, documentation and generative care planning, and workflow optimisation — and identified a consistent pattern: technically impressive predictive performance in retrospective validation, coupled with comparatively sparse evidence of safe, sustained, nurse-led clinical deployment. This pattern mirrors observations in the broader CDSS literature and is consistent with the conclusion of Martinez-Ortigosa et al. (2023) that AI’s potential to enhance nursing clinical decision-making remains constrained by the limited involvement of nurses in the design and development of the tools they are expected to use.

A central tension running through the included literature is the dual-edged nature of automation in nursing surveillance. On one hand, AI-CDSS with lower false-alarm rates than legacy systems (e.g., SepsisAI’s 3.18% false-alarm ratio; Gupta et al., 2024) directly target the alert fatigue that has undermined nurse responsiveness to earlier-generation alerting tools. On the other hand, the same literature warns that as alerts become more accurate and more numerous in scope, the risk of automation bias — uncritical acceptance of algorithmic output — grows correspondingly, particularly where model reasoning is opaque to the bedside nurse (Alenazi et al., 2024).

The bias evidence synthesised in this review (Panch et al., 2019; Norori et al., 2021), and its emerging nursing-specific instantiation in large-language-model-generated care plans, indicates that algorithmic bias is not a hypothetical or distant risk for nursing practice but an empirically demonstrated one, with direct implications for equity in nursing-sensitive outcomes such as pain assessment, psychosocial care planning, and risk stratification.

The translation gap identified across the pressure-injury, fall, and deterioration domains — where only a small minority of validated models progress to reported clinical deployment — echoes well-documented challenges in clinical AI more broadly. For nursing specifically, this implies that purchasing or piloting decisions for AI-CDSS should weight prospective, multi-site validation evidence and local recalibration over headline retrospective accuracy metrics, and that nurse informaticists should be embedded in implementation and post-deployment monitoring processes rather than treated as end-users of an already-finalised tool.

A related interpretive caution concerns what the reported benefits actually measure (see Table 8). Across all four domains, the great majority of reported benefits are technical (e.g., AUROC, accuracy) or process outcomes (e.g., alert volume, documentation time) rather than patient outcomes. High AUROC for pressure-injury or fall-risk models, for instance, demonstrates discriminative accuracy against a retrospective dataset; it does not by itself demonstrate that fewer patients actually developed pressure injuries or fell after the model

was deployed, nor that nursing surveillance or judgement was safer as a result — that inference requires prospective, patient-outcome-level evidence that was largely absent from the included literature. Similarly, a lower false-alarm ratio is a process improvement with plausible downstream safety benefit (via reduced alert fatigue), but the causal chain from "fewer false alarms" to "fewer missed deteriorations" was not directly measured in any included study. This review therefore treats technical performance, process outcomes, clinician-facing outcomes, and patient outcomes as distinct and only loosely coupled categories of evidence, and cautions readers and adopting institutions against treating strong technical performance as a proxy for improved patient safety.

Implications for Nursing Practice, Education, and Policy

Implications for Clinical Practice

Nurse leaders and informaticists should prioritise AI-CDSS with documented prospective or multi-site validation over tools supported only by retrospective single-site performance data, and should establish local monitoring processes to track model performance drift after deployment (Simou, 2025). Alert thresholds should be co-designed with bedside nursing staff to balance sensitivity against alert-fatigue risk, and escalation protocols should explicitly designate when nurses are expected to exercise independent clinical override of an AI recommendation (Pandya, 2026).

Implications for Nursing Education

Consistent with calls in the included literature (Fernandes et al., 2026; Mikkonen et al., 2026), AI and data literacy — including a basic understanding of model training, validation, bias, and the meaning of common performance metrics such as sensitivity, specificity, and AUROC — should be integrated into pre-licensure nursing curricula and continuing professional development, rather than left to informal, self-directed learning. Simulation-based education using AI-generated documentation or decision-support scenarios may offer a low-risk environment in which students can practise critically appraising, rather than passively accepting, algorithmic output (Simou, 2025).

Implications for Policy and Governance

Regulatory and institutional governance frameworks should require bias auditing disaggregated by relevant patient sociodemographic characteristics prior to deployment, mandate explainability or, at minimum, interpretable performance documentation accessible to bedside clinicians, and clarify accountability structures for adverse events involving AI-influenced nursing decisions. International guidance, such as the World Health Organization's guidance on the ethics and governance of AI for health, offers a relevant policy anchor for institution-level AI governance committees that include nursing representation (Georgiadis and Konstantinidis, 2026).

LIMITATIONS

This review has several limitations that should inform interpretation of its findings. First, although the search strategy drew on the major databases conventionally used in nursing-informatics evidence synthesis and was supplemented by systematic backward citation tracking, it was not executed as a fully duplicated, dual-reviewer database export with formal inter-rater reliability statistics, and the review was not prospectively registered (e.g., with PROSPERO). Second, the included primary literature is heterogeneous in design, ranging from large retrospective electronic-health-record-based validation studies to small simulation-based usability evaluations, limiting the ability to draw quantitative (meta-analytic) conclusions about effect size. Third, publication and availability bias likely favour studies reporting positive predictive performance. Finally, given the rapid pace of development in generative AI, findings concerning large-language-model applications in nursing documentation and care planning should be regarded as a snapshot of an early and fast-evolving evidence base. Finally, technical interoperability and EHR-integration barriers were mentioned only in passing

by the workflow-domain sources and were not independently quantified or cited with the same density as the other challenges in Table 9; this likely reflects under-reporting in the source literature itself rather than the absence of the problem, and dedicated evidence synthesis on integration/interoperability barriers is warranted.

CONCLUSION

AI-based clinical decision support systems are reshaping core domains of nursing practice, from deterioration and sepsis surveillance to pressure-injury and fall prevention, documentation, and care planning. The evidence synthesised in this review demonstrates measurable improvements in predictive accuracy and lead time for clinical deterioration, alongside reductions in false-alarm burden and documentation time relative to legacy systems. These benefits are tempered by substantial and recurring challenges: algorithmic bias with demonstrated nursing-specific manifestations, automation bias and deskilling risk, persistent opacity in model reasoning, a wide gap between retrospective validation and prospective clinical deployment, and a workforce whose AI literacy has, to date, developed largely outside formal education structures. Realising the promise of AI-CDSS in nursing will require deliberate nurse involvement in algorithm design and governance, rigorous bias-auditing and external-validation standards, clear accountability frameworks, and systematic integration of AI literacy into nursing education, so that nurses are positioned not as passive recipients of algorithmic output, but as informed, critical partners in AI-augmented clinical decision-making.

Declarations

Author contributions. M.I.: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft. T.O.B.: Supervision, Validation, Writing – review and editing. A.A.O.: Writing – review and editing.

Conflict of interest. The authors declare no conflict of interest.

Funding. No external funding was received for this work.

Ethical approval. Not applicable; this is a systematic review of previously published literature and did not involve new human or animal subject research.

Data availability statement. No new data were generated or analysed for this review. The data extraction table underlying the synthesis is provided as Supplementary Table S2.

Acknowledgements. Not applicable.

Abbreviations

AI, artificial intelligence; CDSS, clinical decision support system(s); AUROC, area under the receiver operating characteristic curve; EHR, electronic health record; EWS, early warning system; ICU, intensive care unit; JBI, Joanna Briggs Institute; LLM, large language model; LR+, positive likelihood ratio; LSTM, long short-term memory; NLP, natural language processing; PICOS, Population, Intervention, Comparator, Outcomes, Study type; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

REFERENCES

1. Ackermann, K., et al. (2022). Computerized clinical decision support systems for the early detection of sepsis among adult inpatients: scoping review. *J. Med. Internet Res.* 24, e31083.
2. Alderden, J., et al. (2024). Explainable artificial intelligence for early prediction of pressure injury risk. *Am. J. Crit. Care* 33, 373–381.
3. Alenazi, A. S., et al. (2024). The efficacy and ethics of AI-powered clinical decision support systems in nursing practice: a systematic review. *Saudi J. Med. Public Health* 1, 1196–1202.

4. Alnawafleh, K. A., et al. (2025). Exploring the ethical dimensions of AI integration in nursing practice: a systematic review. *J. Nurs. Regul.* 16, 228–237.
5. Anderson, C., et al. (2021). Modeling and prediction of pressure injury in hospitalized patients using artificial intelligence. *BMC Med. Inform. Decis. Mak.* 21, 253.
6. Chang, S.-C., et al. (2025). Improving machine learning algorithm for risk of early pressure injury prediction in admission patients using probability feature aggregation. *Digit. Health* 11.
7. Chatzichristos, C., et al. (2025). Bridging the AI-literacy gap in health care: qualitative analysis of the Flanders case study. *J. Med. Internet Res.* 27, e76709.
8. Chen, Y.-H., and Xu, J.-L. (2023). Applying artificial intelligence to predict falls for inpatient. *Front. Med.* 10, 1285192.
9. El Arab, R. A., et al. (2025). Artificial intelligence in nursing: a systematic review of attitudes, literacy, readiness, and adoption intentions. *Front. Digit. Health* 7, 1666005.
10. Elhaddad, M., and Hamam, S. (2024). AI-driven clinical decision support systems: an ongoing pursuit of potential. *Cureus* 16, e57728.
11. Fernandes, F., et al. (2026). Contributions of artificial intelligence to decision making in nursing: a scoping review. *Nurs. Health Sci.*
12. Gaube, S., et al. (2021). Do as AI say: susceptibility in deployment of clinical decision-aids. *NPJ Digit. Med.* 4, 31.
13. Georgiadis, M., and Konstantinidis, C. (2026). Strategic alignment in AI in education: a scoping review of planning-cycle levers and educator acceptance. *EuroMed J. Bus.*
14. Ginestra, J. C., et al. (2019). Clinician perception of a machine learning-based early warning system designed to predict severe sepsis and septic shock. *Crit. Care Med.* 47, 1477–1484.
15. Gupta, A., et al. (2024). Improving sepsis prediction in intensive care with SepsisAI: a clinical decision support system with a focus on minimizing false alarms. *PLOS Digit. Health* 3, e0000569.
16. Hernandez, M., and Bianchi, L. (2025). AI-driven clinical decision support systems: enhancing patient care and hospital efficiency. *Procedia Comput. Sci.* 270, 5625–5633.
17. Ikegwu, A. C., et al. (2025). Artificial intelligence-based clinical decision support systems: enhancing modern healthcare for smart diagnosis and prognosis. *Nat. J. Emerg. Sci. Technol. Innov.* 1, 17–34.
18. Martinez-Ortigosa, A., et al. (2023). Applications of artificial intelligence in nursing care: a systematic review. *J. Nurs. Manag.* 2023, 3219127.
19. McDougall, R. J. (2019). Computer knows best? The need for value-flexibility in medical AI. *J. Med. Ethics* 45, 156–160.
20. Mebrahtu, T. F., et al. (2021). Effects of computerised clinical decision support systems on nursing and allied health professional performance and patient outcomes: a systematic review of systematic reviews. *BMJ Open* 11, e053886.
21. Mikkonen, K., et al. (2026). Artificial intelligence technologies supporting nurses' clinical decision-making: a systematic review. *J. Clin. Nurs.*
22. Muñoz, J., et al. (2026). ICU command centres in critical care: nursing workflows, organizational models, and implementation challenges. *Intensive Crit. Care Nurs.* 92, 104245.
23. Musen, M. A., et al. (2021). "Clinical decision-support systems," in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine* (Cham: Springer International Publishing), 795–840.
24. Nashwan, A. J., and Abujaber, A. A. (2023). Harnessing large language models in nursing care planning: opportunities, challenges, and ethical considerations. *Cureus* 15, e40542.
25. Nashwan, A. J., and Bani Hani, S. (2023). Enhancing oncology nursing care planning for patients with cancer through harnessing large language models. *Asia-Pac. J. Oncol. Nurs.* 10, 100277.
26. Norori, N., et al. (2021). Addressing bias in big data and AI for health care: a call for open science. *Patterns* 2, 100347.
27. Pandya, S. (2026). *Trustworthy AI Systems: Designing Verifiable, Privacy-Preserving Decision Platforms for a (Regulated) Healthcare Industry*. Geh Press.
28. Panch, T., et al. (2019). Artificial intelligence and algorithmic bias: implications for health systems. *J. Glob. Health* 9, 020318.

29. Rabbani, S. A., et al. (2025). Generative artificial intelligence in healthcare: applications, implementation challenges, and future directions. *BioMedInformatics* 5, 37.
30. Shu, W., et al. (2026). Embodied intelligence applications in health care populations: scoping review. *J. Med. Internet Res.* 28, e83871.
31. Simou, E. (2025). The growing importance of soft skills in medical education in the AI era. *Int. Med. Educ.* 4, 50.
32. Song, W., et al. (2021). Predicting pressure injury using nursing assessment phenotypes and machine learning methods. *J. Am. Med. Inform. Assoc.* 28, 759–765.
33. Yang, Y., et al. (2026). Nurses as guardians of time: the hidden clinical value of continuous care in geriatrics. *Front. Public Health* 14, 1799845.
34. Zamanifar, A., et al. (Eds.). (2025). *Applications of Large Language Models (LLM) in Healthcare Systems: Opportunities, Challenges, and Ethical Considerations.*