Received: 28 February 2024 | Revised: 16 August 2024 | Accepted: 23 August 2024

DOI: 10.1111/jnu.13026

JOURNAL OF NURSING SCHOLARSHIP

PROFESSION AND SOCIETY

Decoding machine learning in nursing research: A scoping review of effective algorithms

Jeeyae Choi PhD, RN[1] | Hanjoo Lee MS[2] | Yeounsoo Kim-Godwin PhD, RN[1]

[1]School of Nursing, College of Health and Human Services, University of North Carolina Wilmington, Wilmington, North Carolina, USA

[2]Joint Biomedical Engineering Department, School of Medicine, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA

Correspondence
Jeeyae Choi, School of Nursing, College of Health and Human Services, University of North Carolina Wilmington, 601 South College Road, Wilmington, NC 28403-5995, USA.
Email: choij@uncw.edu

Abstract

Introduction: The rapid evolution of artificial intelligence (AI) technology has revolutionized healthcare, particularly through the integration of AI into health information systems. This transformation has significantly impacted the roles of nurses and nurse practitioners, prompting extensive research to assess the effectiveness of AI-integrated systems. This scoping review focuses on machine learning (ML) used in nursing, specifically investigating ML algorithms, model evaluation methods, areas of focus related to nursing, and the most effective ML algorithms.

Design: The scoping review followed the Preferred Reporting Items for Systematic Review and Meta-Analysis Extension for Scoping Reviews (PRISMA-ScR) guidelines.

Methods: A structured search was performed across seven databases according to PRISMA-ScR: PubMed, EMBASE, CINAHL, Web of Science, OVID, PsycINFO, and ProQuest. The quality of the final reviewed studies was assessed using the Medical Education Research Study Quality Instrument (MERSQI).

Results: Twenty-six articles published between 2019 and 2023 met the inclusion and exclusion criteria, and 46% of studies were conducted in the US. The average MERSQI score was 12.2, indicative of moderate- to high-quality studies. The most used ML algorithm was Random Forest. The four second-most used were logistic regression, least absolute shrinkage and selection operator, decision tree, and support vector machine. Most ML models were evaluated by calculating sensitivity (recall)/specificity, accuracy, receiver operating characteristic (ROC), area under the ROC (AUROC), and positive/negative prediction value (precision). Half of the studies focused on nursing staff or students and hospital readmission or emergency department visits. Only 11 articles reported the most effective ML algorithm(s).

Conclusion: The scoping review provides insights into the current status of ML research in nursing and recognition of its significance in nursing research, confirming the benefits of ML in healthcare. Recommendations include incorporating experimental designs in research studies to optimize the use of ML models across various nursing domains.

Clinical Relevance: The scoping review demonstrates substantial clinical relevance of ML applications for nurses, nurse practitioners, administrators, and researchers. The integration of ML into healthcare systems and its impact on nursing practices have important implications for patient care, resource management, and the evolution of nursing research.

KEYWORDS
artificial intelligence, machine learning, machine learning algorithms, performance validation, scoping review

INTRODUCTION

The rapid advancement of artificial intelligence (AI) technology has led to the development of sophisticated information systems, particularly within the healthcare field. The recent influx of AI-integrated health information systems is transforming the roles of nurses and nurse practitioners (NPs) and the landscape of nursing research and practice. Many advanced systems, having undergone development, evaluation, and subsequent implementation, are currently utilized across healthcare institutions. Furthermore, extensive research has been conducted regarding the impacts and efficacy of these AI-integrated systems both in the US and globally (Clancy, 2020; Pailaha, 2023; Schneidereith & Thibault, 2023; Van Bulck et al., 2023).

AI is a branch of computer science that studies and develops systems that can perform or simulate tasks that typically require human intelligence such as learning, problem-solving, and decision-making. AI employs various algorithms, from basic rule-based to more complex examples, that can "learn" from experience to improve their performance over time (Shi et al., 2023). Specific AI fields include robotics, machine learning (ML), deep learning, natural language processing, computer vision, voice recognition, and knowledge representation (Guo et al., 2020). In healthcare, ML and deep learning have been used to predict disease, aid in drug discovery, respond to disease outbreaks, and personalize healthcare. Natural language processing has been used to automate content analysis, develop clinical terminology, and confirm diagnosis (Schneidereith & Thibault, 2023). Advances in robotics have had remarkable clinical impact such as assisting coronary artery bypass grafting, aiding patients to walk upstairs, and trainable humanoids (Clancy, 2020; Van Bulck et al., 2023).

In the field of AI, ML in particular has shown great potential to improve health outcomes, reduce healthcare costs, and advance clinical research. Researchers in healthcare have used ML algorithms to identify important risk factors in patient falls and to predict depression levels in the elderly population. The massive amount of available healthcare data makes it possible to integrate ML algorithms into healthcare systems and medical devices (Bharadwaj et al., 2021). Various ML-integrated systems have been developed to aid clinicians in providing quality care and managing acute conditions.

Despite the exponential development of ML-integrated systems, clinicians including nurses and NPs, often express discomfort about their use due to lack of transparency of ML models (e.g., to predict or identify risk factor models) and unfamiliarity with validation processes (Khan et al., 2023). There is an urgent need for a comprehensive examination of ML and its application in nursing. This scoping review was conducted to synthesize and describe ML algorithms and methods of assessing performance and to present the most effective algorithms for potential utilization in nursing.

OBJECTIVES

This scoping review aims to answer the following research questions: (1) What algorithms were used to develop ML models? (2) What methods were used to evaluate model performance? (3) What areas of nursing were focused on using ML models? and (4) What algorithms were the most effective in nursing?

DESIGN

This scoping review was conducted by three investigators following the guidance of the Preferred Reporting Items for Systematic Review and Meta-Analysis Extension for Scoping Reviews (PRISMA-ScR) (Tricco et al., 2018). The quality of studies was evaluated by the Medical Education Research Study Quality Instrument (MERSQI) (Reed et al., 2007).

MATERIALS AND METHODS

Inclusion and exclusion

Nursing studies use ML algorithms to develop prediction models or models that can identify risk factors to improve nursing outcomes. Inclusion criteria were studies that (1) used specific ML algorithms; (2) evaluated ML model performance; (3) focused on areas of nursing to which ML models applied; and (4) were conducted by nursing scholars. Exclusion criteria were studies that (1) did not use ML algorithms; (2) did not evaluate ML model performance; (3) did not focus on areas of nursing; or (4) were not conducted by nursing scholars.

Search strategy

Seven databases (PubMed, EMBASE, CINAHL, Web of Science, OVID, PsycINFO, and ProQuest) were searched by title and abstract for articles published between January 2019 and December 2023. Keywords were "machine learning," "nurse," and "nurse practitioner." Although there are terms commonly associated with the domain of "machine learning," such as "deep learning" or "artificial intelligence," we opted for "machine learning" because of its precise definition and alignment with our research objectives. For instance, "deep learning" is a subset of ML (Jakhar & Kaur, 2020), whereas "artificial intelligence" encompasses the behavior of machines imitating human behavior (Mesko & Gorog, 2020; Simmons & Chappell, 1988), thereby providing a broader perspective within the scope of ML. The following limits were applied: (1) full text available; (2) English language; (3) human studies; and (4) no literature reviews, opinions, editorials, or books. Table S1 shows search queries in each database.

Selection process

A total of 191 articles were extracted from seven databases. Duplicate articles (*n* = 7) and those that did not use algorithms, develop ML models, or evaluate model performance (*n* = 125) were excluded based on review of the title and abstract, resulting in 59 potentially eligible articles. Three reviewers performed a full text review of the 59 articles and removed 33 after screening against inclusion and exclusion criteria. The same reviewers appraised the final 26 full-text articles for their quality and knowledge synthesis. The first author, a doctoral trained nurse researcher, verified the eligibility of all full-text articles. Any disagreements were discussed between the reviewers and were resolved through consensus meetings. The detailed process of identifying studies is depicted in the PRISMA-ScR flow diagram (Figure 1).

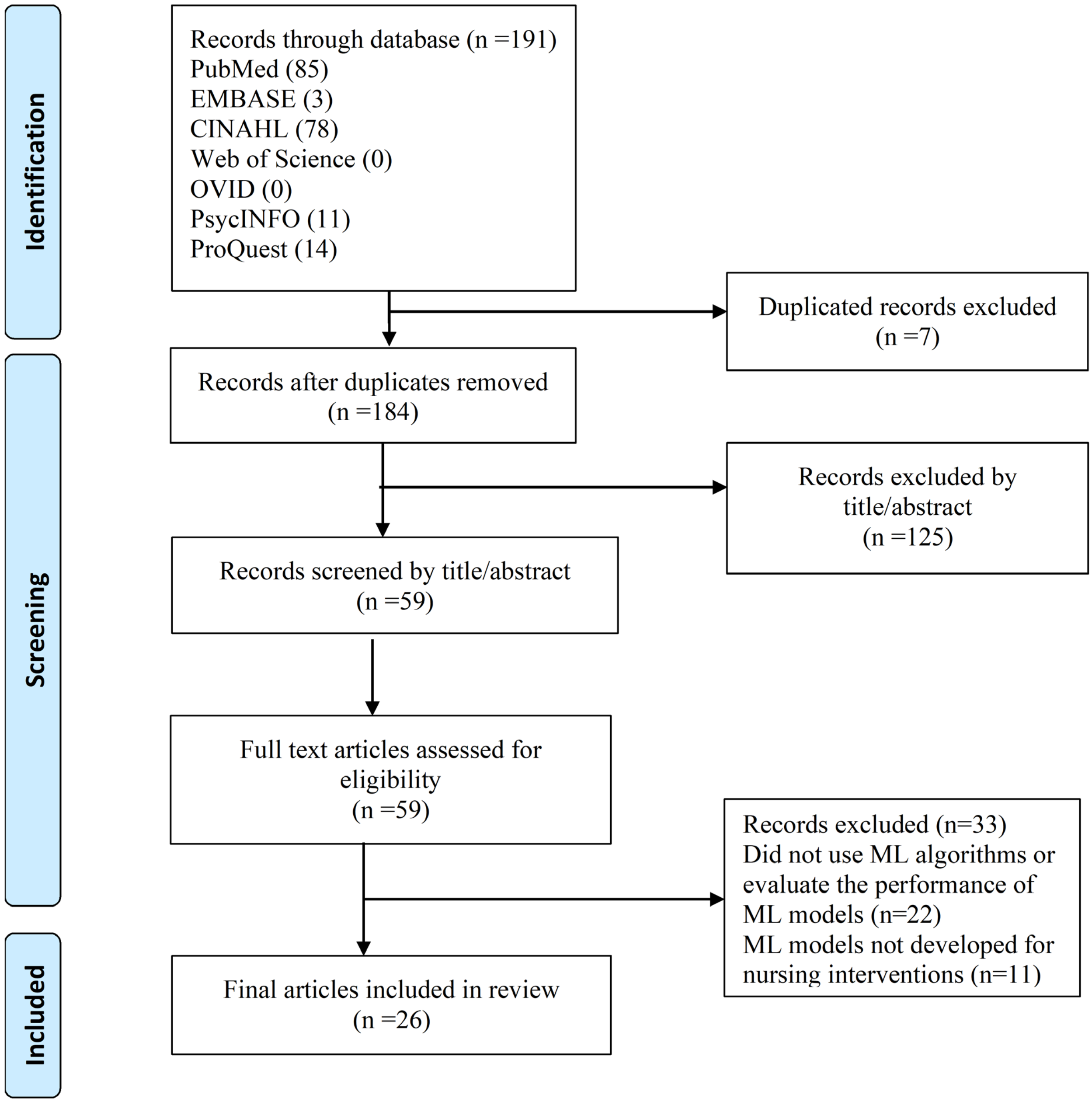

FIGURE 1 PRISMA flow diagram for study selection.

Quality assessment

Three reviewers independently evaluated the final 26 studies using the MERSQI criteria, which was developed to evaluate the quality of medical education research and has also been used in other areas of healthcare research due to its comprehensive evaluation of study quality (Cook & Reed, 2015; Ryan et al., 2022). The instrument includes six domains: (1) study design; (2) sampling; (3) types of data; (4) validity of evaluation instrument; (5) data analysis; and (6) outcomes. Each domain carries a minimum score of 1 and maximum score of 3 except for the sampling domain, which carries a minimum score of 0.5. The range of the total score is from 5 to 18. Its intra-rater reliability falls between 0.78 and 0.998. Inter-rater reliability of the MERSQI for overall scores ranges from 0.68 to 0.95 (Cook & Reed, 2015; Smith & Learman, 2017). To assess the inter-rater reliability of reviewers, the Fleiss' kappa value was measured (Belur et al., 2021). Reviewer disagreements were resolved until a consensus was reached.

Data extraction and synthesis

Three reviewers extracted data independently using a standard form containing: (1) author/year/country; (2) study design; (3) datasets (characteristics and total number); (4) study purpose; (5) algorithms/performance reported/software tool; (6) findings; and (7) MERSQI scores. The extracted data were summarized and categorized based on: (1) ML algorithm (e.g., random forest, neural network); (2) methods used to evaluate ML model performance (e.g., receiver operating characteristic [ROC], sensitivity, recall); (3) areas of nursing (e.g., patient falls or urinary tract infection); and (4) most effective ML algorithms.

Datasets were extracted instead of sample sizes because ML studies use datasets based on clinical record rather than traditional study population. Meta-analysis was not possible due to different measures such as ML model performance versus ML models as applied to various areas of nursing.

RESULTS

Twenty-six articles published between January 2019 and December 2023 were included in the final review. Table S2 shows a detailed outline of each study with MERSQI scores. Twelve studies (46.1%) were conducted in the US (Hannaford et al., 2021; Hobensack et al., 2022; Huang et al., 2021; Lindberg et al., 2020; Lo et al., 2019; Park et al., 2020; Song et al., 2021; Song et al., 2022; Topaz et al., 2020; Womack et al., 2020; Yang et al., 2021; Zachariah et al., 2020). Four (15.3%) were conducted in China (Hu et al., 2021; Jin et al., 2022; Su et al., 2021; Yan et al., 2023), two (7.7%) in Canada (Havaei et al., 2021; Havaei et al., 2022), one in Australia (Conway et al., 2021), one in Japan (Nakagami et al., 2021), one in South Korea (Kim et al., 2023), one in Taiwan (Hsu et al., 2021), and one in Switzerland (Pouzols et al., 2023). Three studies were conducted by the US and Canada (Kwon et al., 2019), US and Saudi Arabia (Yakusheva et al., 2022), and Asia-Pacific countries (Dong et al., 2022). Study designs varied and no experimental studies were found. Datasets ranged from 102 to 727,676 including patients' clinical records, surveys of people in the community, home visit notes, nursing staff activities, and nursing students' academic records.

Study quality

Among the 26 studies reviewed, five scored below 12 and 21 scored 12 and above. The average MERSQI score was 12.2, ranging from 10.7 to 13.7, inferring a moderate- to high-quality study. Two studies scored 13.7: “predicting pressure injury based on nursing assessments” (Song et al., 2021); and “investigating psychological differences of nurses in Asia-Pacific countries based on survey” (Dong et al., 2022). The lowest scored study (10.7) was “predicting the level of different depression with survey datasets” (Su et al., 2021). MERSQI scores ≥12 indicate quality research (Smith & Learman, 2017). The five studies that scored <12 used datasets from one institution or didn't report instruments (e.g., software tools for data analysis), so “validity of evaluation instrument” was not scored. Average domain scores were highest (3.0) for the ‘type of data’ and ‘data analysis’ domains and were lowest (0.7) for the ‘sampling’ domain of MERSQI. There was good agreement between reviewers, with a Fleiss' kappa value of 0.42 and a 95% confidence interval (CI) between 0.30 and 0.54. Fleiss' kappa coefficient was statistically significant ($p < 0.001$).

ML algorithms

ML algorithms examined were random forest (RF), logistic regression (LR), least absolute shrinkage and selection operator (LASSO), extreme gradient boosting (XGBoost), decision tree (DT), support vector machine (SVM), Naïve Bayes, neural networks (NN), Bayesian network, linear discriminant analysis (LDA), long short-term memory (LSTM)-neural network, and others. Twenty-four (92.3%) studies used RF, 13 (50%) used LR or LR with LASSO regularization, 11 (42.3%) used DT or its variation (e.g., extra DT, J48, random tree), and 11 (42.3%) used SVM. Further, we investigated software tools used in the studies, which included R package, Python, and Weka. Ten studies (38.4%) used R package, seven (26.9%) used Python, and three (11.5%) used Weka. Two studies (7.7%) used Konstanz Information Miner (KNIME), an open-source analytics platform (https://www.knime.com/) and four (15.3%) did not report their software tool. Table 1 summarizes types of ML algorithms and Table S3 summarizes software tools used.

Evaluating performance

To evaluate developed ML models the following were calculated: Sensitivity (recall)/specificity (true/false positive rate); accuracy;

TABLE 1 Machine Learning algorithms used in studies ($N=26$).

Machine	Articles	n
Random forest	Lo et al. (2019), Kwon et al. (2019), Womack et al. (2020), Topaz et al. (2020), Lindberg et al. (2020), Hsu et al. (2021), Havaei et al. (2021 & 2022), Conway et al. (2021), Hannaford et al. (2021), Hu et al. (2021), Yakusheva et al. (2022), Yang et al. (2021), Song et al. (2021 & 2022), Nakagami et al. (2021), Su et al. (2021), Huang et al. (2021), Hobensack et al. (2022), Dong et al. (2022), Jin et al. (2022), Kim et al. (2023), Yan et al. (2023), Pouzols et al. (2023)	24 (92.3%)
Logistic Regression (LR), LR with Least Absolute Shrinkage and Selection Operator (LASSO) regularization	Kwon et al. (2019), Womack et al. (2020), Park et al. (2020), Conway et al. (2021), Hu et al. (2021), Yang et al. (2021), Song et al. (2021, 2022), Nakagami et al. (2021), Huang et al. (2021), Jin et al. (2022), Kim et al. (2023), Yan et al. (2023)	13 (50.0%)
Decision Tree, Extra Decision Tree, J48, C5.0, Random Tree, Reduced Error Pruning Tree, Gradient boosted Decision Tree, Classification tree	Topaz et al. (2020), Lindberg et al. (2020), Park et al. (2020), Zachariah et al. (2020), Hsu et al. (2021), Hannaford et al. (2021), Su et al. (2021), Hobensack et al. (2022), Jin et al. (2022), Kim et al. (2023), Yan et al. (2023)	11 (42.3%)
Support Vector Machine (SVM), Linear SVM	Womack et al. (2020), Park et al. (2020), Hsu et al. (2021), Hannaford et al. (2021), Song et al. (2021, 2022), Nakagami et al. (2021), Su et al. (2021), Hobensack et al. (2022), Jin et al. (2022), Yan et al. (2023)	11 (42.3%)
Shrinkage and Selection Operator (LASSO)	Kwon et al. (2019), Womack et al. (2020), Lindberg et al. (2020), Conway et al. (2021), Song et al. (2021, 2022), Hobensack et al. (2022), Yan et al. (2023)	8 (30.7%)
Extreme Gradient Boosting (XGBoost), Gradient boost, Adaptive boosting, Ridge regression XGBoost, Light gradient boosting Machine (GBM)	Kwon et al. (2019), Conway et al. (2021), Hannaford et al. (2021), Hu et al. (2021), Nakagami et al. (2021), Huang et al. (2021), Hobensack et al. (2022), Dong et al. (2022)	8 (30.7%)
Neural Network, K-nearest Neural Network, Recurrent Neural Network, Deep Neural Network	Zachariah et al. (2020), Hsu et al. (2021), Hannaford et al. (2021), Song et al. (2021), Su et al. (2021), Hobensack et al. (2022), Yan et al. (2023)	7 (26.9%)
Naïve Bayes, Baysesian Network	Topaz et al. (2020), Hsu et al. (2021), Hannaford et al. (2021), Hu et al. (2021), Song et al. (2022)	5 (19.2%)
Linear Discriminant Analysis (LDA)	Jin et al. (2022)	1 (3.8%)
Long Short Term Memory (LSTM) Neural Network	Pouzols et al. (2023)	1 (3.8%)

receiver operating characteristic (ROC); *F*-score (F1); positive/negative prediction value (precision); area under the ROC (AUROC); and others (e.g., Root mean square error [RMSE], Kappa, Brier Score). Twenty studies (77%) reported AUROC, nineteen studies (73.1%) reported sensitivity (recall)/specificity values (true/false-positive rate), thirteen (50.0%) reported accuracy, ten reported positive/negative prediction value (precision), seven (26.9%) reported *F*-scores, and six (23.0%) reported ROC. One study used Kappa to evaluate the predictive accuracy of their ML models, and one study used confirmatory factor analysis to calculate the predictive power of identified factors. Table 2 summarizes the evaluation methods used to measure ML model performance.

Areas of nursing

Focal areas of nursing were patient falls, hospital admissions, infections, nursing staff and students, mental health, pressure injuries, pain, apnea, and others (e.g., capturing concerns, deep vein thrombosis). Six studies (23.0%) focused on nursing staff workload, mental health, or student graduating rates, and five (19.2%) focused on hospital readmission, emergency department visits, or length of stay (Table 3).

Effective algorithms

Eleven (42.3%) studies reported the most effective ML algorithms identified in their studies. They were RF, DT, SVM, LSTM, and XGBoost (Table 4).

DISCUSSION

This scoping review extracted and summarized 26 studies that used algorithms to build ML models for various nursing domains. The US appears to be dominant in ML research in nursing. Readily available large datasets in home health care, such as the Outcome and Assessment Information Set (OASIS) (https://www.cms.gov/) and electronic health records (EHR) systems, which are operational in 96% of acute hospitals (Health IT, 2023) in the US may be a plausible reason. For example, four studies (Hobensack et al., 2022; Lo et al., 2019; Song et al., 2022; Topaz et al., 2020) used home health care datasets and five (Huang et al., 2021; Lindberg et al., 2020; Park et al., 2020; Song et al., 2021; Zachariah et al., 2020) used electronic health records datasets.

MERSQI showed an average score of 12.2 (range: 10.7–13.7), implying moderate- to high-quality studies in the final review. Two

TABLE 2 Performance evaluation methods used in studies ($N=26$).

Machine	Articles	*n*
Area Under the Curve (AUC) Area Under the Receiver Operating Characteristic (AUROC) or ROC-AUC	Lo et al. (2019), Womack et al. (2020), Topaz et al. (2020), Zachariah et al. (2020), Hsu et al. (2021), Hannaford et al. (2021), Su et al. (2021), Hobensack et al. (2022), Jin et al. (2022), Kim et al. (2023), Pouzols et al. (2023), Lindberg et al. (2020), Park et al. (2020), Conway et al. (2021), Yakusheva et al. (2022), Yang et al. (2021), Song et al. (2021), Nakagami et al. (2021), Huang et al. (2021), Dong et al. (2022)	20 (77%)
Sensitivity (Recall) & Specificity, True/False Positive Rate	Pouzols et al. (2023), Lindberg et al. (2020), Park et al. (2020), Conway et al. (2021), Hannaford et al. (2021), Hu et al. (2021), Yakusheva et al. (2022), Yang et al. (2021), Song et al. (2021 & 2022), Nakagami et al. (2021), Su et al. (2021), Jin et al. (2022), Yan et al. (2023), Topaz et al. (2020), Hsu et al. (2021), Kim et al. (2023), Lo et al. (2019), Zachariah et al. (2020)	19 (73.1%)
Accuracy Balanced accuracy Precision accuracy	Pouzols et al. (2023), Lo et al. (2019), Womack et al. (2020), Park et al. (2020), Hsu et al. (2021), Hannaford et al. (2021), Hu et al. (2021), Yang et al. (2021), Song et al. (2021), Su et al. (2021), Jin et al. (2022), Kim et al. (2023), Yan et al. (2023)	13 (50.0%)
Positive/Negative Prediction Value (Rate), Precision, Average precision	Conway et al. (2021), Yakusheva et al. (2022), Yang et al. (2021), Su et al. (2021), Lo et al. (2019), Topaz et al. (2020), Park et al. (2020), Hsu et al. (2021), Kim et al. (2023), Yan et al. (2023)	10 (38.4%)
F-score (F1)	Womack et al. (2020), Topaz et al. (2020), Hsu et al. (2021), Yang et al. (2021), Song et al. (2021, 2022), Yan et al. (2023)	7 (26.9%)
Receiver Operating Characteristic (ROC)	Pouzols et al. (2023), Kwon et al. (2019), Lindberg et al. (2020), Hannaford et al. (2021), Hu et al. (2021), Su et al. (2021)	6 (23.0%)
Root Mean Square Error (RMSE)	Havaei et al. (2021, 2022)	2 (7.7%)
Brier Score	Jin et al. (2022), Yan et al. (2023)	2 (7.7%)
Precision-Recall Area under the Curve/Precision-Recall Curve	Huang et al. (2021), Song et al. (2022)	2 (7.7%)
Kappa	Yakusheva et al. (2022)	1 (3.8%)
False negatives	Park et al. (2020)	1 (3.8%)
Prospective Prediction Results	Su et al. (2021)	1 (3.8%)
Confirmatory factor analysis	Havaei et al. (2021)	1 (3.8%)

TABLE 3 Areas of nursing in studied ($N=26$).

Nursing care	Articles	*n*
Nursing Staff/Students: RN Strain, Workplace Safety, Students Graduating rate (BS program), Mental Health, Turnover Rate	Womack et al. (2020), Havaei et al. (2021, 2022), Hannaford et al. (2021), Dong et al. (2022), Kim et al. (2023)	6 (23.0%)
Hospital Admission: Re-admission, Emergency Dept. Visit, ICU Length of Stay	Kwon et al. (2019), Topaz et al. (2020), Yakusheva et al. (2022), Huang et al. (2021), Song et al. (2022)	5 (19.2%)
Patient Falls	Lo et al. (2019), Lindberg et al. (2020), Yang et al. (2021)	3 (11.5%)
Mental Health: Traumatic Brain Injury (TBI), Cognitive Impairment, Depression	Hsu et al. (2021), Hsu et al. (2021), Su et al. (2021)	3 (11.5%)
Pressure Injury	Song et al. (2021), Nakagami et al. (2021), Pouzols et al. (2023)	3 (11.5%)
Infections: Hospital Acquired Catheter-Associated Urinary Tract Infections (HA-UTIs), UTIs	Park et al. (2020), Zachariah et al. (2020)	2 (7.6%)
Pain	Yan et al. (2023)	1 (3.8%)
Apnea	Conway et al. (2021)	1 (3.8%)
Capturing Concerns	Hobensack et al. (2022)	1 (3.8%)
Cancer-associated Deep vein thrombosis (DVT)	Jin et al. (2022)	1 (3.8%)

factors contribute to high-quality studies. First, most studies used various ML algorithms to build models and compare their performance, which demands complicated data analysis. Second, most datasets from well-structured large healthcare institutions indicate "objective measurement." However, none of the studies used an experimental design and did not measure patient or health outcomes. These weaknesses should be further discussed and improved in ML research in nursing.

TABLE 4 Most effective ML algorithms.

Articles	Algorithms used	Effective algorithm	Performance
Womack et al. (2020)	Extra Decision Tree, Gradient Boost, Random Forest, Logistic Regression, Space Vector Machine (SVM)	Gradient Boost	Shift hour 12: Accuracy = 0.640 F1 = 0.639 AUC = 0.638
Topaz et al. (2020)	Naïve Bayes, Decision tree (J48), Random Forest	Random Forest	Recall = 0.81 Precision = 0.83 *F*-score = 0.82 AUC = 0.76
Lindberg et al. (2020)	Decision Tree (A single classification tree), Random Forest, Adaptive Boosting	Decision Tree	Sensitivity = 0.78 Specificity = 0.78 AUROC = 0.85 Graph: ROC, True/False Positive Rate
Conway et al. (2021)	Random Forest, Logistic Regression, Shrinkage and Selection Operator (LASSO), Ridge regression Extreme Gradient Boosting (XGBoost)	Random Forest	AUROC = 0.66 Graph: Sensitivity, Specificity, Positive predictive values, Negative predictive values
Hannaford et al. (2021)	C5.0, Random Forest, Extreme Gradient Boosting (XGBoost), Neural Network, Space Vector Machine (SVM), Naïve Bayes, K-Neural Network, Logistic Regression	Random Forest	Year 5: Prediction accuracy = 100% AUC = 1 Sensitivity = 1
Song et al. (2021)	Space Vector Machine (SVM), Logistic Regression, Random Forest, Neural Network	Random Forest	AUC = 0.92 Accuracy = 0.85 Sensitivity = 0.84 Specificity = 0.85 F1 = 0.81
Nakagami et al. (2021)	Logistic Regression, Random Forest, Linear Space Vector Machine (SVM), Extreme Gradient Boosting (XGBoost)	XGBoost	Sensitivity = 0.78 Specificity = 0.74 AUC = 0.80 Positive Predictive Value = 0.016 Negative Predictive Value = 0.998
Hobensack et al. (2022)	Neural Network, Decision Tree, Random Forest, Space Vector Machine (SVM), Gradient Boosted Tree, Logistic Regression	Gradient Boosted Tree	Precision = 0.99 Recall = 0.59 *F*-score = 0.74 AUC = 0.96
Song et al. (2022) US (12.0)	Logistic Regression, Random Forest, Bayesian network, Space Vector Machine (SVM), Naïve Bayes	Random Forest SVM	Random Forest: Sensitivity = 0.927 Recall = 0.721 *F*-score = 0.811 Precision-recall curve = 0.864 SVM: Sensitivity = 0.922 Recall = 0.731 *F*-score = 0.815 Precision-recall curve = 0.821
Kim et al. (2023)	Decision Tree, Logistic Regression, Random Forest	Random Forest	Precision = 0.92 Recall = 0.91 Accuracy = 92% *F*-1Score = 0.92 AUROC (AUC-ROC) = 0.97
Pouzols et al. (2023)	Random Forest, Long short-term memory (LSTM) Neural Network	The LSTM Neural Network	Sensitivity = 0.74 Specificity = 0.82 Accuracy = 0.82 AUROC = 0.87 Graph: ROC, True/False positive rate

Abbreviations: AUC, area under the curve; AUROC (AUC-ROC), area under the receiver operating characteristic; ROC, receiver operating characteristic.

ML algorithms

Most of the studies (92.3%) might have used RF because of its well-known benefits; namely, high predictive accuracy and variable (factor) interpretability. RF is also known to identify which variables are most influential in prediction models, allowing for a better clinical understanding of the models (IBM, 2024; Moutaib et al., 2024). The next most popular algorithms used were LR, DT, and SVM. The characteristics of health datasets in our studies might drive principal investigators (PIs) to choose them given the strength of each algorithm and its suitability for those specific characteristics. For example, LR is well-suited for binary classification with high interpretability, DT is optimal for data-driven predictive models without data preprocessing requirements, and SVM consistently delivers strong performance in data-driven classification across diverse datasets (Miller et al., 2020; Nusinovici et al., 2020; Sarker et al., 2019).

However, four studies (Havaei et al., 2021; Havaei et al., 2022; Lo et al., 2019; Yakusheva et al., 2022) used only one ML algorithm to build models, which might introduce bias that could be built upon in a single ML model. Health datasets are often complex and a mix of labeled, unlabeled, and missing data. Building multiple ML models is recommended to prevent introducing confirmation bias and to facilitate identifying the most accurate and high-performing models (Engelhard et al., 2021; Rubinger et al., 2023; Vollmer et al., 2020).

Evaluation method of performance

Most articles used some combination of the following methods to evaluate ML model performance: AUROC; sensitivity (recall)/specificity (true/false-positive rate); accuracy; positive/negative prediction value (precision); and ROC. This aligns with recommended evaluation methods in the ML field (Colliot, 2023; Erickson & Kitamura, 2021; Srivastava, 2024). However, only two studies used RMSE to evaluate performance of ML models, despite it is one of the recommended evaluation methods. It implies a necessity of a formalized guideline that shows recommended ML algorithms and associated evaluation method of performances.

Areas of nursing

Eleven studies (42.3%) focused on hospital admissions or ER visits and issues affecting nurses and nursing students, such as RN workload, their mental health, or student graduation rate in BS programs. The focus on these areas of nursing may be because datasets related to these areas are often extremely large and readily available, making them suitable for ML research. Prediction models using ML algorithms of patient falls, infections (e.g., UTIs), and pressure injuries are known to be beneficial to patients, nurses, and NPs; however, only eight studies (30.8%) focused on these areas.

In fall studies, fall datasets were small in numbers ($n=530$) (Lindberg et al., 2020), or fall instances within a dataset were small (Lo et al., 2019), which can hinder the robustness of ML models due to limited data for training and validation. Other datasets were collected by a telephone survey, implying recall bias could potentially affect the performance of ML algorithms (Yang et al., 2021). This implies that fall datasets include complex variables, making it hard to apply ML. In the case of infection studies, types of UTIs vary, such as HA-CAUTI, CAUTI, or UTI. These were nursing-sensitive datasets and required more information in nursing documentation (Park et al., 2020; Zachariah et al., 2020). It implies that more information should be processed and added to the datasets before applying ML. In studies of pressure injuries, datasets were extracted from EHRs, and many datasets were not collected if they were in narratives (Pouzols et al., 2023; Song et al., 2021). In addition, it was not clear whether pressure injuries developed before hospitalization (Nakagami et al., 2021). It implies that pressure injury studies require multiple dimensions of variables to build high-performing predictive ML models.

These challenges highlight the complexity of leveraging ML for healthcare studies related to falls, infections, and pressure injuries. In summary, ensuring the reliability and effectiveness of ML approaches in improving patient outcomes requires careful data collection, preprocessing, and model development.

Effective algorithms

A little less than half of the studies identified the most effective ML algorithms and measured the performance of ML models. Among 11 studies, six (Conway et al., 2021; Hannaford et al., 2021; Kim et al., 2023; Song et al., 2021; Song et al., 2022; Topaz et al., 2020) reported that RF was the most effective algorithm compared with other algorithms applied, implying that RF can predict clinical and nursing datasets (Faiz et al., 2021; Li et al., 2020; Rahman et al., 2020; Zhu, 2024) with high accuracy. It is inferred from the main strengths of the RF algorithm, which include adaptability, scalability, and robustness (Faiz et al., 2021; Li et al., 2018; Rahman et al., 2020). RF estimates the important variables and assigns different weights for each decision tree efficiently (adaptability), works well on large datasets because it can handle a huge number of variables (scalability), and mitigates errors when the dataset has imbalances (robustness). Importantly, RF possesses all three strengths, whereas other ML algorithms do not. For example, SVM has adaptability and robustness but lacks scalability, and DT has only adaptability. It also suggests that RF is a powerful ML algorithm for healthcare professionals seeking reliable and precise insights from complex datasets.

The strengths of RF also indicate that it is a powerful and versatile algorithm that can be effectively applied to a wide range of complex and multifaceted nursing problems. These problems often involve multiple variables due to numerous factors such as patient demographics, medical history, and social and psychological conditions. These variables are interrelated and exhibit nonlinear relationships, making it challenging to isolate specific causes and effects.

RF's ability to handle complex data, capture nonlinear relationships, and provide valuable insights makes it a useful tool for nursing research and practice.

LIMITATIONS

This review has several limitations when interpreting the results. None of the studies adopted an experimental design or performed power analyses. In addition, only studies written in English were included in this review, which may introduce bias in the results. PRISMA-ScR was the only guideline used as the checklist items and no other review guidelines used, which may affect the study's internal validity. MERSQI is a comprehensive evaluation tool to assess study quality, but other tools could have yielded different results. The review period (January 2019 to December 2023) may have missed valuable studies published before 2019. These limitations restrict the comprehensive perspective of ML research in nursing and limit the generalizability of results.

CONCLUSION

This scoping review explored the status of ML research in nursing. The results show that nursing scholars recognize the importance of ML models, confirming the beneficial use of ML in healthcare (Jayatilake & Ganegoda, 2021; Mustafa & Rahimi Azghadi, 2021; Siddique & Chow, 2021). Nursing scholars have used ML models in various areas of nursing to prevent or reduce adverse outcomes of disease or treatment and to improve patient care. For optimal use of ML, experimental design of future research studies is strongly recommended to prove models' effectiveness in nursing.

CLINICAL RESOURCES

Key Staff: Disrupting nursing education using extended reality (XR), artificial intelligence, and machine learning: https://www.nursingworld.org/foundation/rninitiative/practice-ready-nurse-graduates/disrupting-nursing-education-with-xr-ai-and-ml/key-staff-disrupting-nursing-education-using-extended-reality-xr-artificial-intelligence-and-machine-learning/.

Artificial Intelligence at the NIH: https://datascience.nih.gov/artificial-intelligence.

MAYO Clinic Proceedings: Digital Health: https://www.mcpdigitalhealth.org/.

MIT Clinical Machine Learning Group: https://clinicalml.org/.

ACKNOWLEDGMENTS

The authors have nothing to report.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

SIGMA THETA TAU INTERNATIONAL CHAPTER

Nu Omega Chapter.

ORCID

Jeeyae Choi https://orcid.org/0000-0001-7287-6384

REFERENCES

Belur, J., Tompson, L., Thornton, A., & Simon, M. (2021). Interrater reliability in systematic review methodology: exploring variation in coder decision-making. *Sociological Methods & Research*, *50*(2), 837–865. https://doi.org/10.1177/0049124118799372

Bharadwaj, H. K., Agarwal, A., Chamola, V., Lakkaniga, N. R., Hassija, V., Guizani, M., & Sikdar, B. (2021). A review on the role of machine learning in enabling IoT based healthcare applications. *IEEE Access*, *9*, 38859–38890. https://doi.org/10.1109/access.2021.3059858

Clancy, T. R. (2020). Artificial intelligence and nursing: The future is now. *Journal of Nursing Administration*, *50*(3), 125–127. https://doi.org/10.1097/NNA.0000000000000855

Colliot, O. (2023). *Machine learning for brain disorders*. Humana Press. https://link.springer.com/book/10.1007/978-1-0716-3195-9

Conway, A., Jungquist, C. R., Chang, K., Kamboj, N., Sutherland, J., Mafeld, S., & Parotto, M. (2021). Predicting prolonged apnea during nurse-administered procedural sedation: Machine learning study. *JMIR Perioperative Medicine*, *4*, e29200. https://doi.org/10.2196/29200

Cook, D. A., & Reed, D. A. (2015). Appraising the quality of medical education research methods: The medical education research study quality instrument and the Newcastle-Ottawa scale-education. *Academic Medicine*, *90*(8), 1067–1076. https://doi.org/10.1097/ACM.0000000000000786

Dong, Y., Yeo, M. C., Tham, X. C., Danuaji, R., Nguyen, T. H., Sharma, A. K., Rn, K., Pv, M., Tai, M. L. S., Ahmad, A., Tan, B. Y. Q., Ho, R. C., Chua, M. C. H., & Sharma, V. K. (2022). Investigating psychological differences between nurses and other health care workers from the Asia-Pacific region during the early phase of COVID-19: Machine learning approach. *JMIR Nursing*, *5*, e32647. https://doi.org/10.2196/32647

Engelhard, M. M., Navar, A. M., & Pencina, M. J. (2021). Incremental benefits of machine learning-when do we need a better mousetrap? *JAMA Cardiology*, *6*(6), 621–623. https://doi.org/10.1001/jamacardio.2021.0139

Erickson, B. J., & Kitamura, F. (2021). Magician's corner: 9. Performance metrics for machine learning models. *Radiology. Artificial Intelligence*, *3*(3), e200126. https://doi.org/10.1148/ryai.2021200126

Faiz, F., Ideno, Y., Iwasaki, H., Muroi, Y., & Inoue, S. (2021). Multilabel classification of nursing activities in a realistic scenario. In *Activity and behavior computing* (pp. 269–288). Springer. https://doi.org/10.1007/978-981-15-8944-7_17

Guo, Y., Hao, Z., Zhao, S., Gong, J., & Yang, F. (2020). Artificial intelligence in health care: Bibliometric analysis. *Journal of Medical Internet Research*, *22*(7), e18228. https://doi.org/10.2196/18228

Hannaford, L., Cheng, X., & Kunes-Connell, M. (2021). Predicting nursing baccalaureate program graduates using machine learning models: A quantitative research study. *Nurse Education Today*, *99*, 104784. https://doi.org/10.1016/J.NEDT.2021.104784

Havaei, F., Ji, X. R., & Boamah, S. A. (2022). Workplace predictors of quality and safe patient care delivery among nurses using machine learning techniques. *Journal of Nursing Care Quality*, *37*, 103–109. https://doi.org/10.1097/NCQ.0000000000000600

Havaei, F., Ji, X. R., MacPhee, M., & Straight, H. (2021). Identifying the most important workplace factors in predicting nurse mental health using machine learning techniques. *BMC Nursing*, *20*(1), 216. https://doi.org/10.1186/S12912-021-00742-9

Health IT. (2023). *96% of US hospitals have EHRs, but barriers remain to interoperability, ONC says*. Becker's Healthcare. https://www.beckershospitalreview.com/ehrs/96-of-us-hospitals-have-ehrs-but-barriers-remain-to-interoperability-onc-says.html#:~:text=As%20of%202021%2C%2096%20percent,implemented%20a%20certified%20EHR%20system

Hobensack, M., Song, J., Chae, S., Kennedy, E., Zolnoori, M., Bowles, K. H., McDonald, M. V., Evans, L., & Topaz, M. (2022). Capturing concerns about patient deterioration in narrative documentation in home healthcare. *AMIA Annual Symposium Proceedings. AMIA Symposium*, *2022*, 552–559.

Hsu, S. D., Chao, E., Chen, S. J., Hueng, D. Y., Lan, H. Y., & Chiang, H. H. (2021). Machine learning algorithms to predict in-hospital mortality in patients with traumatic brain injury. *Journal of Personalized Medicine*, *11*(11), 1144. https://doi.org/10.3390/JPM11111144

Hu, M., Shu, X., Yu, G., Wu, X., Välimäki, M., & Feng, H. (2021). A risk prediction model based on machine learning for cognitive impairment among Chinese community-dwelling elderly people with normal cognition: Development and validation study. *Journal of Medical Internet Research*, *23*(2), e20298. https://doi.org/10.2196/20298

Huang, K., Gray, T. F., Romero-Brufau, S., Tulsky, J. A., & Lindvall, C. (2021). Using nursing notes to improve clinical outcome prediction in intensive care patients: a retrospective cohort study. *Journal of the American Medical Informatics Association: JAMIA*, *28*, 1660–1666. https://doi.org/10.1093/JAMIA/OCAB051

IBM. (2024). *What is random forest?* IBM. https://www.ibm.com/topics/random-forest

Jakhar, D., & Kaur, I. (2020). Artificial intelligence, machine learning and deep learning: Definitions and differences. *Clinical and Experimental Dermatology*, *45*(1), 131–132. https://doi.org/10.1111/ced.14029

Jayatilake, S. M. D. A. C., & Ganegoda, G. U. (2021). Involvement of machine learning tools in healthcare decision making. *Journal of Healthcare Engineering*, *2021*, 6679512. https://doi.org/10.1155/2021/6679512

Jin, S., Qin, D., Liang, B. S., Zhang, L. C., Wei, X. X., Wang, Y. J., Zhuang, B., Zhang, T., Yang, Z. P., Cao, Y. W., Jin, S. L., Yang, P., Jiang, B., Rao, B. Q., Shi, H. P., & Lu, Q. (2022). Machine learning predicts cancer-associated deep vein thrombosis using clinically available variables. *International Journal of Medical Informatics*, *161*, 104733. https://doi.org/10.1016/J.IJMEDINF.2022.104733

Khan, B., Fatima, H., Qureshi, A., Kumar, S., Hanan, A., Hussain, J., & Abdullah, S. (2023). Drawbacks of artificial intelligence and their potential solutions in the healthcare sector. *Biomedical Materials & Devices*, *1*, 731–738. https://doi.org/10.1007/s44174-023-00063-2

Kim, S. K., Kim, E. J., Kim, H. K., Song, S. S., Park, B. N., & Jo, K. W. (2023). Development of a nurse turnover prediction model in Korea using machine learning. *HealthCare*, *11*(11), 1583. https://doi.org/10.3390/HEALTHCARE11111583

Kwon, J. Y., Karim, M. E., Topaz, M., & Currie, L. M. (2019). Nurses "seeing forest for the trees" in the age of machine learning: Using nursing knowledge to improve relevance and performance. *CIN. Computers, Informatics, Nursing*, *37*(4), 203–212. https://doi.org/10.1097/CIN.0000000000000508

Li, J., Tian, Y., Zhu, Y., Zhou, T., Li, J., Ding, K., & Li, J. (2020). A multicenter random forest model for effective prognosis prediction in collaborative clinical research network. *Artificial Intelligence in Medicine*, *103*, 101814. https://doi.org/10.1016/j.artmed.2020.101814

Li, X., Cui, L., Tao, S., Chen, J., Zhang, X., & Zhang, G. Q. (2018). HyCLASSS: A hybrid classifier for automatic sleep stage scoring. *IEEE Journal of Biomedical and Health Informatics*, *22*(2), 375–385. https://doi.org/10.1109/JBHI.2017.2668993

Lindberg, D. S., Prosperi, M., Bjarnadottir, R. I., Thomas, J., Crane, M., Chen, Z., Shear, K., Solberg, L. M., Snigurska, U. A., Wu, Y., Xia, Y., & Lucero, R. J. (2020). Identification of important factors in an inpatient fall risk prediction model to improve the quality of care using EHR and electronic administrative data: A machine-learning approach. *International Journal of Medical Informatics*, *143*, 104272. https://doi.org/10.1016/J.IJMEDINF.2020.104272

Lo, Y., Lynch, S. F., Urbanowicz, R. J., Olson, R. S., Ritter, A. Z., Whitehouse, C. R., O'Connor, M., Keim, S. K., McDonald, M., Moore, J. H., & Bowlesc, K. H. (2019). Using machine learning on home health care sssessments to predict fall risk. *Studies in Health Technology and Informatics*, *264*, 684–688. https://doi.org/10.3233/SHTI190310

Mesko, B., & Gorog, M. (2020). A short guide for medical professionals in the era of artificial intelligence. *NPJ Digital Medicine*, *3*, 126. https://doi.org/10.1038/s41746-020-00333-z

Miller, C. H., Sacchet, M. D., & Gotlib, I. H. (2020). Support vector machines and affective science. *Emotion Review*, *12*(4), 297–308. https://doi.org/10.1177/1754073920930784

Moutaib, M., Fattah, M., Farhaoui, Y., Aghoutane, B., & El Bekkali, M. (2024). Fetal electrocardiogram prediction using machine learning: A random forest-based approach. *Indonesian Journal of Electrical Engineering and Computer Science*, *33*(2), 1076. https://doi.org/10.11591/ijeecs.v33.i2.pp1076-1083

Mustafa, A., & Rahimi Azghadi, M. (2021). Automated machine learning for healthcare and clinical notes analysis. *Computers*, *10*(2), 24. https://doi.org/10.3390/computers10020024

Nakagami, G., Yokota, S., Kitamura, A., Takahashi, T., Morita, K., Noguchi, H., Ohe, K., & Sanada, H. (2021). Supervised machine learning-based prediction for in-hospital pressure injury development using electronic health records: A retrospective observational cohort study in a university hospital in Japan. *International Journal of Nursing Studies*, *119*, 103932. https://doi.org/10.1016/J.IJNURSTU.2021.103932

Nusinovici, S., Tham, Y. C., Chak Yan, M. Y., Wei Ting, D. S., Li, J., Sabanayagam, C., Wong, T. Y., & Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of Clinical Epidemiology*, *122*, 56–69. https://doi.org/10.1016/j.jclinepi.2020.03.002

Pailaha, A. D. (2023). The impact and issues of artificial intelligence in nursing science and healthcare settings. *SAGE Open Nursing*, *9*, 23779608231196847. https://doi.org/10.1177/23779608231196847

Park, J. I., Bliss, D. Z., Chi, C. L., Delaney, C. W., & Westra, B. L. (2020). Knowledge discovery with machine learning for hospital-acquired catheter-associated urinary tract infections. *CIN: Computers, Informatics, Nursing*, *38*(1), 28–35. https://doi.org/10.1097/CIN.0000000000000562

Pouzols, S., Despraz, J., Mabire, C., & Raisaro, J. L. (2023). Development of a predictive model for hospital-acquired pressure injuries. *CIN: Computers, Informatics, Nursing*, *41*, 884–891. https://doi.org/10.1097/CIN.0000000000001029

Rahman, A., Nahid, N., Hassan, I., & Ahad, M. A. R. (2020). Nurse care activity recognition Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers.

Reed, D. A., Cook, D. A., Beckman, T. J., Levine, R. B., Kern, D. E., & Wright, S. M. (2007). Association between funding and quality of published medical education research. *JAMA*, *298*(9), 1002–1009. https://doi.org/10.1001/jama.298.9.1002

Rubinger, L., Gazendam, A., Ekhtiari, S., & Bhandari, M. (2023). Machine learning and artificial intelligence in research and healthcare. *Injury*, *54*(Suppl 3), S69–S73. https://doi.org/10.1016/j.injury.2022.01.046

Ryan, G. V., Callaghan, S., Rafferty, A., Higgins, M. F., Mangina, E., & McAuliffe, F. (2022). Learning outcomes of immersive technologies in health care student education: Systematic review of the literature. *Journal of Medical Internet Research*, *24*(2), e30082. https://doi.org/10.2196/30082

Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2019). BehavDT: A behavioral decision tree learning to build user-centric context-aware predictive model. *Mobile Networks and Applications*, *25*(3), 1151–1161. https://doi.org/10.1007/s11036-019-01443-z

Schneidereith, T. A., & Thibault, J. (2023). The basics of artificial intelligence in nursing: Fundamentals and recommendations for educators. *Journal of Nursing Education*, *62*(12), 716–720. https://doi.org/10.3928/01484834-20231006-03

Shi, J., Wei, S., Gao, Y., Mei, F., Tian, J., Zhao, Y., & Li, Z. (2023). Global output on artificial intelligence in the field of nursing: A bibliometric analysis and science mapping. *Journal of Nursing Scholarship*, *55*(4), 853–863. https://doi.org/10.1111/jnu.12852

Siddique, S., & Chow, J. C. L. (2021). Machine learning in healthcare communication. *Encyclopedia*, *1*(1), 220–239. https://doi.org/10.3390/encyclopedia1010021

Simmons, A., & Chappell, S. (1988). Artificial intelligence -definition and practice. *IEEE Journal of Oceanic Engineering*, *13*(2), 14–42.

Smith, R. P., & Learman, L. A. (2017). A plea for MERSQI: The medical education research study quality instrument. *Obstetrics and Gynecology*, *130*(4), 686–690. https://doi.org/10.1097/AOG.0000000000002091

Song, J., Hobensack, M., Bowles, K. H., McDonald, M. V., Cato, K., Rossetti, S. C., Chae, S., Kennedy, E., Barrón, Y., Sridharan, S., & Topaz, M. (2022). Clinical notes: An untapped opportunity for improving risk prediction for hospitalization and emergency department visit during home health care. *Journal of Biomedical Informatics*, *128*, 104039. https://doi.org/10.1016/J.JBI.2022.104039

Song, W., Kang, M. J., Zhang, L., Jung, W., Song, J., Bates, D. W., & Dykes, P. C. (2021). Predicting pressure injury using nursing assessment phenotypes and machine learning methods. *Journal of the American Medical Informatics Association: JAMIA*, *28*(4), 759–765. https://doi.org/10.1093/JAMIA/OCAA336

Srivastava, T. (2024). *12 important model evaluation metrics for machine learning everyone should know (updated 2023)*. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluation-error-metrics/#:~:text=Accuracy%2C%20confusion%20matrix%2C%20log-,the%20most%20popular%20evaluation%20metrics

Su, D., Zhang, X., He, K., & Chen, Y. (2021). Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. *Journal of Affective Disorders*, *282*, 289–298. https://doi.org/10.1016/J.JAD.2020.12.160

Topaz, M., Woo, K., Ryvicker, M., Zolnoori, M., & Cato, K. (2020). Home healthcare clinical notes predict patient hospitalization and emergency department visits. *Nursing Research*, *69*(6), 448–454. https://doi.org/10.1097/NNR.0000000000000470

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., … Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine*, *169*(7), 467–473. https://doi.org/10.7326/M18-0850

Van Bulck, L., Couturier, R., & Moons, P. (2023). Applications of artificial intelligence for nursing: Has a new era arrived? *European Journal of Cardiovascular Nursing*, *22*(3), e19–e20. https://doi.org/10.1093/eurjcn/zvac097

Vollmer, S., Mateen, B. A., Bohner, G., Kiraly, F. J., Ghani, R., Jonsson, P., Cumbers, S., Jonas, A., McAllister, K. S. L., Myles, P., Granger, D., Birse, M., Branson, R., Moons, K. G. M., Collins, G. S., Ioannidis, J. P. A., Holmes, C., & Hemingway, H. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*, *368*, l6927. https://doi.org/10.1136/bmj.l6927

Womack, D. M., Hribar, M. R., Steege, L. M., Vuckovic, N. H., Eldredge, D. H., & Gorman, P. N. (2020). Registered nurse strain detection using ambient data: An exploratory study of underutilized operational data streams in the hospital workplace. *Applied Clinical Informatics*, *11*(4), 598–605. https://doi.org/10.1055/S-0040-1715829

Yakusheva, O., Bang, J. T., Hughes, R. G., Bobay, K. L., Costa, L., & Weiss, M. E. (2022). Nonlinear association of nurse staffing and readmissions uncovered in machine learning analysis. *Health Services Research*, *57*, 311–321. https://doi.org/10.1111/1475-6773.13695

Yan, Z., Liu, M., Wang, X., Wang, J., Wang, Z., Liu, J., Wu, S., & Luan, X. (2023). Construction and validation of machine learning algorithms to predict chronic post-surgical pain among patients undergoing total knee arthroplasty. *Pain Management Nursing*, *24*(6), 627–633. https://doi.org/10.1016/J.PMN.2023.04.008

Yang, R., Plasek, J. M., Cummins, M. R., & Sward, K. A. (2021). Predicting falls among community-dwelling older adults: A demonstration of applied machine learning. *CIN. Computers, Informatics, Nursing*, *39*(5), 273–280. https://doi.org/10.1097/CIN.0000000000000688

Zachariah, P., Sanabria, E., Liu, J., Cohen, B., Yao, D., & Larson, E. (2020). Novel strategies for predicting healthcare-associated infections at admission: Implications for nursing care. *Nursing Research*, *69*(5), 399–403. https://doi.org/10.1097/NNR.0000000000000449

Zhu, E. (2024). Explore the application of machine learning algorithms in predicting stroke. *Highlights in Science, Engineering and Technology*, *92*, 269–274.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1.

Table S2.

Table S3.

How to cite this article: Choi, J., Lee, H. & Kim-Godwin, Y. (2025). Decoding machine learning in nursing research: A scoping review of effective algorithms. *Journal of Nursing Scholarship*, *57*, 119–129. https://doi.org/10.1111/jnu.13026