Received: 13 October 2021 | Revised: 28 February 2022 | Accepted: 11 October 2022

DOI: 10.1002/nop2.1429

REVIEW ARTICLE

NursingOpen Open Access WILEY

A systematic review of predictive models for hospital-acquired pressure injury using machine learning

You Zhou[1,2] | Xiaoxi Yang[1,2] | Shuli Ma[1,2] | Yuan Yuan[3] | Mingquan Yan[1]

[1]Department of Gastroenterology, Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, China

[2]School of Nursing, School of Public Health, Yangzhou University, Yangzhou, China

[3]Department of Nursing, Affiliated Hospital of Yangzhou University, Yangzhou University, Yangzhou, China

Correspondence
Yuan Yuan, Department of Nursing, Affiliated Hospital of Yangzhou University, No. 368 Hanjiang Middle Road, Yangzhou, Jiangsu Province, China.
Email: yuanyuan@yzu.edu.cn

Mingquan Yan, Department of Gastroenterology, Affiliated Hospital of Yangzhou University, No. 368 Hanjiang Middle Road, Yangzhou, Jiangsu Province, China.
Email: mqyan@yzu.edu.cn

Abstract

Aims and objectives: To summarize the use of machine learning (ML) for hospital-acquired pressure injury (HAPI) prediction and to systematically assess the performance and construction process of ML models to provide references for establishing high-quality ML predictive models.

Background: As an adverse event, HAPI seriously affects patient prognosis and quality of life, and causes unnecessary medical investment. At present, the performance of various scales used to predict HAPIs is still unsatisfactory. As a new statistical tool, ML has been applied to predict HAPIs. However, its performance has varied in different studies; moreover, some deficiencies in the model construction process were observed in each study.

Design: Systematic review.

Methods: Relevant articles published between 2010–2021 were identified in the PubMed, Web of Science, Scopus, Embase and CINHAL databases. Study selection was performed in accordance with the preferred reporting items for systematic reviews and meta-analysis guidelines. The quality of the included articles was assessed using the prediction model risk of bias assessment tool.

Results: Twenty-three studies out of 1793 articles were considered in this systematic review. The sample size of each study ranged from 149–75353; the prevalence of pressure injuries ranged from 0.5%–49.8%. ML showed good performance for HAPI prediction. However, some deficiencies were observed in terms of data management, data pre-processing and model validation.

Conclusions: ML, as a powerful decision-making assistance tool, is helpful for the prediction of HAPIs. However, existing studies have been insufficient in terms of data management, data pre-processing and model validation. Future studies should address these issues to establish ML models for HAPI prediction that can be widely used in clinical practice.

Relevance to Clinical Practice: This review highlights that ML is helpful in predicting HAPI; however, in the process of data management, data pre-processing and model validation, some deficiencies still need to be addressed. The ultimate goal of

You Zhou, Xiaoxi Yang, Co-first author, Contributed equally to this article.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Nursing Open* published by John Wiley & Sons Ltd.

integrating ML into HAPI prediction is to develop a practical clinical decision-making tool. A complete and rigorous model construction process should be followed in future studies to develop high-quality ML models that can be applied in clinical practice.

KEYWORDS
machine learning, pressure injury, pressure ulcer, prediction, systematic review

1 | INTRODUCTION

Hospital-acquired pressure injury (HAPI) refers to a localized damage to the skin and underlying soft tissues that occurs during hospitalization (Edsberg et al., 2016). It has been reported that in the United States, the incidence of HAPI is approximately 5%–15%, and approximately one to three million hospitalized patients are affected by it every year (Chou et al., 2013; Mervis & Phillips, 2019; Padula et al., 2015). HAPI increases the length of hospitalization and medical expenses, affects the quality of patients' lives and induces complications, such as infections, which increase mortality (Lyder et al., 2012; Reddy et al., 2006; Spilsbury et al., 2007). Currently, the HAPI prevention depends mainly on the observation and assessment by nurses. Although some risk assessment tools for pressure injury (PI), such Braden, Norton and Waterlow scales, have been widely used in clinical practice, studies have shown that their accuracy and reliability are not satisfactory (Shi et al., 2019).

As a branch of artificial intelligence, machine learning (ML) has become a new statistical method that has emerged in medical practice and is increasingly being used in diagnosis (AlJame et al., 2021; Koga et al., 2021), complications (Kambakamba et al., 2020; Mohammed et al., 2020), prognosis (Akcay et al., 2020) and recurrence (Li et al., 2021) prediction. Compared to conventional statistical models, ML can actively learn the complex relationships between data, overcome the limitations of non-linearity and maintain stability in high-dimensional datasets (Mangold et al., 2021). In addition, as medical data are surging, various types of data are included in electronic health records (EHRs). ML has an unparalleled advantage in the analysis of unstructured data (Barber et al., 2021; De Silva et al., 2021), pictures (Das et al., 2021) and other data. However, numerous ML studies have shown that several problems related to model constructions still exist. Many researchers have focused on the excellent performance of models on local datasets, but have ignored their reproducibility in other clinical environments, thus limiting further promotion of this powerful decision-making assistance tool in clinical practice (Cabitza & Campagner, 2021). A previous study reviewed the application of ML in PI management, but did not describe specific prediction tasks in detail (Jiang et al., 2021). Therefore, a systematic review is needed to summarize the application of ML for HAPI prediction and to analyse the advantages and disadvantages of the model construction process.

Impact Statement

What does this paper contribute to the global community?
1. We found that ML had good performance in HAPI prediction and can assist in HAPI prevention.
2. In the process of model construction, existing studies had deficiencies in data management, data pre-processing and model validation.
3. Future studies should follow a stricter model construction process and add more detailed descriptions for peers to learn, which could improve the reproducibility of models and help develop practical high-quality ML predictive models.

2 | THE REVIEW

2.1 | Aims

Our aim in this systematic review is to summarize the existing articles related to the use of ML for HAPI prediction, and to systematically assess the performance and construction process of ML models to provide references for the establishment of high-quality predictive models in the future.

2.2 | Design

This systematic review was performed in accordance with the preferred reporting items for systematic reviews and meta-analysis guidelines (Page et al., 2021) (Supplementary File S1).

2.3 | Search methods

The search strategy was developed with the assistance of a university librarian and was modified for different databases. To obtain higher quality articles, a comprehensive literature search was conducted in the PubMed, Web of Science, Scopus, Embase and CINHAL databases for studies reporting ML predictive tools for PI and published between 1 January 2010–14 July 2021. The search terms included in PubMed are as follows: ("Pressure Ulcer"[MeSH] OR "Pressure

Ulcer*"[tiab] OR "Pressure sore*"[tiab] OR Pressure injur*" [tiab] OR "Bedsore*"[tiab] OR "Bed sore*"[tiab] OR "Decubitus Ulcer*"[tiab] OR "Decubitus injur*"[tiab]) AND ("Machine Learning"[MeSH] OR "Machine Learning"[tiab] OR "Algorithms"[tiab] OR "neural networks"[tiab] OR "iterative learning"[tiab] OR "decision tree"[tiab] OR "support vector machine"[tiab] OR "random forest"[tiab] OR "artificial intelligence"[tiab] OR "deep learning"[tiab] OR "logistic regression"[tiab]) AND ("prognos*"[tiab] OR "predict*"[tiab] OR "scor*"[tiab] OR "valid*"[tiab]). We also identified additional relevant articles from the literature.

2.4 | Inclusion/exclusion criteria

Studies that met the following inclusion criteria were included in this systematic review: (a) studies using ML algorithms (including deep learning, ML and ML-based logistic regression) to build HAPI predictive models and (b) English publications. Studies that met the following exclusion criteria were excluded: (a) non-HAPI studies; (b) review, abstract, correspondence, case reports, studies not available in full text, duplicated studies and non-human studies and (c) studies that did not specify the algorithms used and lacked outcomes.

2.5 | Data abstraction and synthesis

Two reviewers independently used Excel to extract and synthesize information from the 23 identified studies. The summarized information included authors, year of publication, country, aim, type of ML model, method of model validation, sample source and size, incidence of PI, predictors and model performance. Any disagreements were resolved through a consensus by another reviewer.

2.6 | Quality appraisal

The methodological quality of the included studies was independently assessed by two reviewers using the prediction model risk of bias assessment tool (PROBAST) (Wolff et al., 2019). PROBAST was used to assess the risk of bias and the application of diagnostic and prognostic prediction model studies, which included a total of 20 questions in four domains (participants, predictors, outcome and analysis). The risk of bias for each question and domain could be answered as low, unclear or high.

3 | RESULTS

3.1 | Search outcomes

A total of 1793 studies were identified in the initial literature search. After removing duplicates, 823 studies were identified, of which 774 were excluded after the title and abstract were blindly screened by two reviewers. Studies meeting the inclusion criteria were allowed for the next round of full-text evaluation. Fifty-one studies were carefully reviewed in full text, 28 of which were exclude for the following reasons: non-HAPI studies ($n = 2$), not employing ML predictive tools ($n = 5$), non-predictive studies ($n = 10$), conference abstracts ($n = 4$), protocols ($n = 1$), reviews ($n = 3$) and non-English publications ($n = 3$). Finally, 23 studies were included in this systematic review. Disagreements were resolved through discussion. The selection procedure is summarized in Figure 1.

3.2 | Study characteristics

Table 1 presents the main characteristics of the included studies. The abovementioned 23 studies were conducted in the following countries: China ($n = 7$), the United States ($n = 10$), Japan ($n = 2$), South Korea ($n = 1$), Canada ($n = 1$), Spain ($n = 1$) and Denmark ($n = 1$), thus covering populations from different regions in North America, Asia and Europe. The main outcome was the occurrence of HAPI, and the subtypes were not limited (including PIs that occurred in different departments, surgery-related PIs and medical device-related PIs).

3.3 | Database information

In a total of 23 studies, 18 studies (Alderden et al., 2018; Alderden et al., 2021; Cai et al., 2021; Chen et al., 2018; Choi et al., 2020; Cichosz et al., 2019; Delparte et al., 2021; Deng et al., 2017; Hu et al., 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Li et al., 2019; Nakagami et al., 2021; Setoguchi et al., 2016; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Su et al., 2012) used data from EHRs, four studies (Cramer et al., 2019; Goodwin & Demner-Fushman, 2020; Sotoodeh et al., 2020; Vyas et al., 2020) used the Medical Information Mart for Intensive Care III (MIMIC III) database and one study (Raju et al., 2015) used the Military Nursing Outcomes Database (MilNOD). Most studies used structured data as input parameters to predict the HAPI, and two studies (Goodwin & Demner-Fushman, 2020; Sotoodeh et al., 2020) used unstructured nursing records. A total of 235758 patients were included in this systematic review. The sample size of the included studies varied greatly, ranging from 149–75353; the prevalence of PI ranged from 0.5%–55.6%. Moreover, ten studies (Alderden et al., 2021; Cramer et al., 2019; Hu et al., 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Nakagami et al., 2021; Setoguchi et al., 2016; Sotoodeh et al., 2020; Su et al., 2012) were based on unbalanced datasets. The source of patients in each study was also different: three studies (Cai et al., 2021; Chen et al., 2018; Su et al., 2012) focused on surgery-related PI; the populations of eleven studies (Alderden et al., 2018; Alderden et al., 2021; Choi et al., 2020; Cramer et al., 2019; Deng et al., 2017; Goodwin & Demner-Fushman, 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Sotoodeh et al., 2020; Vyas et al., 2020) were ICU patients; seven studies (Cichosz et al., 2019; Hu et al., 2020; Nakagami et al., 2021; Raju et al., 2015; Setoguchi et al., 2016; Song,

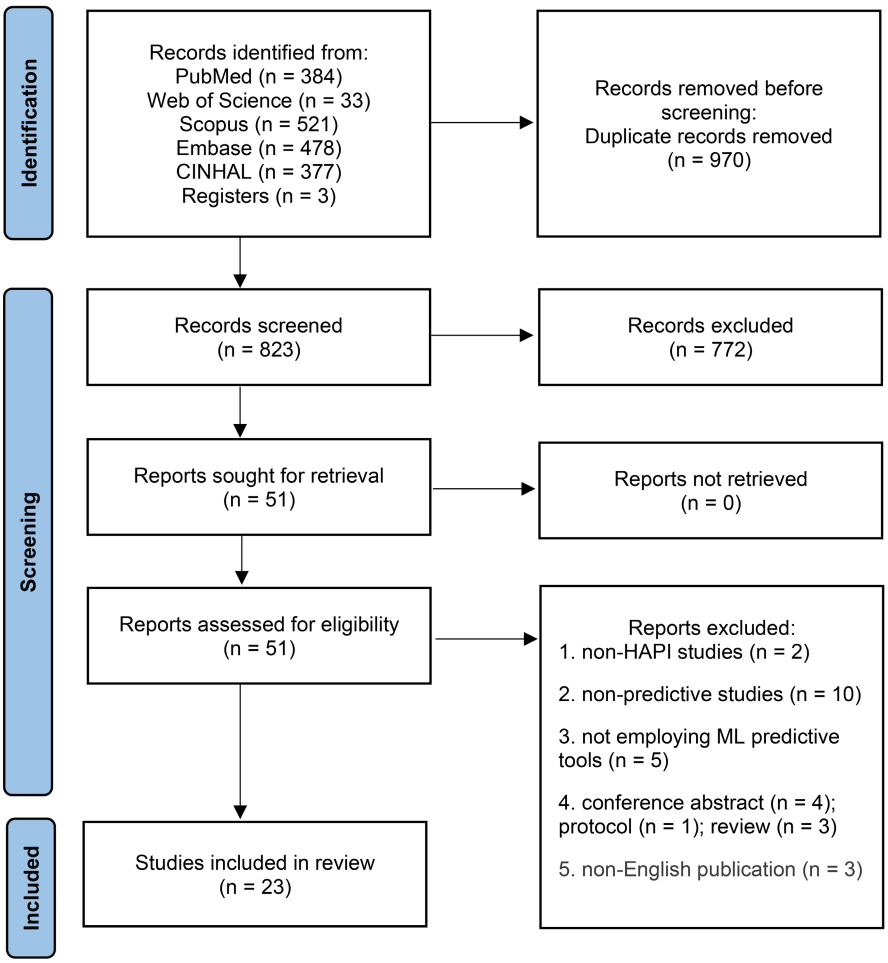

FIGURE 1 PRISMA Flow diagram followed for selecting studies

Gao, et al., 2021; Song, Kang, et al., 2021) included all hospitalized patients; one study (Delparte et al., 2021) was in a rehabilitation centre and one study (Li et al., 2019) focused on hospice care. In addition, 12 studies (Alderden et al., 2018; Alderden et al., 2021; Cichosz et al., 2019; Cramer et al., 2019; Delparte et al., 2021; Deng et al., 2017; Ladios-Martin et al., 2020; Li et al., 2019; Raju et al., 2015; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Vyas et al., 2020) reported PI risk assessment tools and three studies (Alderden et al., 2018; Alderden et al., 2021; Song, Gao, et al., 2021) reported the corresponding PI preventive measures.

3.4 | Data preparation

Eleven studies (Alderden et al., 2018; Alderden et al., 2021; Cramer et al., 2019; Goodwin & Demner-Fushman, 2020; Hu et al., 2020; Ladios-Martin et al., 2020; Raju et al., 2015; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Sotoodeh et al., 2020; Su et al., 2012) reported on data pre-processing, which includes deduplication; missing value processing through direct deletion, use of means, random forest (RF) and k-nearest neighbours for filling and multiple imputation; data standardization and natural language vectorization. Ten studies (Alderden et al., 2018; Cai et al., 2021; Chen et al., 2018; Deng et al., 2017; Hu et al., 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Song, Gao, et al., 2021; Song, Kang, et al., 2021) reported feature selection methods, including literature review, logistic regression, clinical recommendations and univariate analyses. In the ten studies with unbalanced datasets, except for the models used in some studies that were not affected by data imbalance, six studies (Alderden et al., 2021; Cramer et al., 2019; Hu et al., 2020; Ladios-Martin et al., 2020; Nakagami et al., 2021; Sotoodeh et al., 2020) reported unbalanced data processing methods, including synthetic minority oversampling technique (SMOTE), case–control, undersampling on the majority of samples and oversampling of the minority of samples.

3.5 | Model design

A total of 73 ML models were developed for HAPI prediction, and the number of models in each of the 23 studies varied from one to nine. Sixteen studies (Alderden et al., 2021; Choi et al., 2020; Cramer et al., 2019; Delparte et al., 2021; Deng et al., 2017; Goodwin & Demner-Fushman, 2020; Hu et al., 2020; Kaewprag et al., 2015; Ladios-Martin et al., 2020; Li et al., 2019; Nakagami et al., 2021; Raju et al., 2015; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Sotoodeh et al., 2020; Su et al., 2012) reported more than one ML model. Six studies (Cichosz et al., 2019; Cramer et al., 2019; Delparte

TABLE 1 Characteristics of the included studies ($n = 23$)

Author	Year	Country	Aim	ML models	Validation method	Patient source	Numbers	PI(%)
Cai	2021	China	To develop an ML-based predictive model for SRPI in patients undergoing cardiovascular surgery	XGBoost	NA	Surgery	149	24.8
Song	2021b	USA	To predict hospital-acquired and non-hospital-acquired PI using nursing assessment phenotypes	RF ANN SVM LR	5-fold cross-validation	Hospitalization	16165	38.1
Song	2021a	China	To build machine learning models for predicting PI nursing adverse event	ANN RF SVM DT	10-fold cross-validation	Hospitalization	5814	28.8
Alderden	2021	USA	To develop models predicting HAPI among surgical ICU patients using EHRs data	ANN RF AdaBoost GB LR	NA	ICU	5101	6.5
Nakagami	2021	Japan	To construct a predictive model for PI development which included feature variables that can be collected on the first day of hospitalization by nurses	LR RF SVM XGBoost	5-fold cross-validation	Hospitalization	75353	0.5
Delparte	2021	Canada	To develop a PI risk screening instrument for use during SCI rehabilitation	DT LR	NA	Rehabilitation Centre	807	22.0
Ladios Martin	2020	Spain	To build a model to detect PI risk in ICU patients	BN DT RF LR ANN SVM	NA	ICU	4277	3.2
Vyas	2020	USA	To predict PI in ICU patients using XGBoost and Braden subscales as input features	XGBoost	NA	ICU	13282	16.8
Choi	2020	Korea	To construct a risk prediction model for oral mucosal PI development in intubated patients in ICU	BN LR	cross-validation	ICU	27	55.6
Hu	2020	China	To construct inpatient PI prediction models using machine learning techniques	LR DT RF	10-fold cross-validation	Hospitalization	11838	1.4
Goodwin	2020	USA	To develop a generalizable model capable of leveraging clinical notes to predict healthcare-associated diseases 24–96 hours in advance	CANTRIP LSTM SVM LR	NA	ICU	35218	39.8
Sotoodeh	2020	USA	To predict PI in Unstructured Clinical Notes	ANN RF LR	NA	ICU	24457	3.5

TABLE 1 (Continued)

Author	Year	Country	Aim	ML models	Validation method	Patient source	Numbers	PI(%)
Cramer	2019	USA	To predict PI in the ICU using EHRs structured data	LR SVM RF GB ANN EN	5-fold cross-validation	ICU	28886	3.3
Cichosz	2019	Denmark	To investigate a new tool, Q-scale, for in-hospital prediction of PI	LR	10-fold cross-validation	Hospitalization	383	18
Li	2019	China	To develop a predictive model to explain the variables involved in the development of PIs for patients at the end of life	ANN DT SVM LR	cross-validation	Hospice	2062	49.8
Chen	2018	China	To build an ANN model for predicting surgery-related PI in cardiovascular surgical patients	ANN	NA	Surgery	149	25.2
Alderden	2018	USA	To develop a model for predicting development of PIs among surgical critical care patients.	RF	NA	ICU	6376	12.1
Kaewprag	2017	USA	To develop predictive models for PI from intensive care unit EHRs using BNs	BN	NA	ICU	7717	7.6
Deng	2017	China	To construct risk-prediction models of HAPIs in intensive care patients	DT LR	10-fold cross-validation	ICU	468	20.1
Setoguchi	2016	Japan	To develop a prediction model for PI cases that continue to occur at an acute care hospital	DT	10-fold cross-validation	Hospitalization	8286	0.6
Raju	2015	USA	To build and compare data mining models for PI prevalence	LR DT RF MARS	10-fold cross-validation	Hospitalization	1653	20.1
Kaewprag	2015	USA	To develop predictive models of PI incidence from ICU EHRs	LR BN DT RF KNN SVM	10-fold cross-validation	ICU	7717	7.6
Su	2012	China	The objective of this study is to use data mining techniques to construct the prediction model for PIs in surgical patients	MTS SVM DT LR	4-fold cross-validation	Surgery	168	4.7

Abbreviations: AdaBoost, adaptive boosting; ANN, artificial neural network; BN, Bayesian network; CANTRIP, recurrent additive network for temporal risk prediction; DT, decision tree; EHR, electronic health record; EN, elastic net; GB, gradient boosting; HAPI, hospital-acquired pressure injury; LR, logistic regression; LSTM, long short-term memory; ML, machine learning; MARS, multivariate adaptive regression splines; MTS, Mahalanobis–Taguchi system; NA, not available; PI, pressure injury; RF, random forest; SCI, spinal cord injury; SRPI, surgery-related pressure injury; SVM, support vector machine; XGBoost, extreme gradient boosting.

et al., 2021; Ladios-Martin et al., 2020; Song, Gao, et al., 2021; Vyas et al., 2020) compared ML with existing scoring scales. The most common ML models include logistic regression (LR), artificial neural networks (ANN), decision tree (DT) and RF. Thirteen studies (Choi et al., 2020; Cichosz et al., 2019; Cramer et al., 2019; Deng et al., 2017; Hu et al., 2020; Kaewprag et al., 2015; Li et al., 2019; Nakagami et al., 2021; Raju et al., 2015; Setoguchi et al., 2016; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Su et al., 2012) used cross-validation to validate the models, 10 studies (Alderden et al., 2018; Alderden et al., 2021; Cai et al., 2021; Chen et al., 2018; Delparte et al., 2021; Goodwin & Demner-Fushman, 2020; Kaewprag et al., 2017; Ladios-Martin et al., 2020; Sotoodeh et al., 2020; Vyas et al., 2020) did not report validation methods and no studies performed external validation. Ten studies (Alderden et al., 2018; Chen et al., 2018; Delparte et al., 2021; Goodwin & Demner-Fushman, 2020; Hu et al., 2020; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Sotoodeh et al., 2020; Su et al., 2012; Vyas et al., 2020) detailed the hyperparameters of the models, but most of these did not clearly report the tuning methods. Specifically, three studies (Kaewprag et al., 2017; Nakagami et al., 2021; Sotoodeh et al., 2020) used grid search to search for hyperparameters and five studies (Alderden et al., 2018; Alderden et al., 2021; Goodwin & Demner-Fushman, 2020; Song, Gao, et al., 2021; Sotoodeh et al., 2020) disclosed the source codes of their proposed models.

3.6 | Model performance

Table 2 presents the performance and predictors of the best ML model proposed in each study. The indicators used to measure the performance of ML models included the area under the receiver-operating characteristic curve (AUC), accuracy, sensitivity (SEN), specificity (SPE), positive predictive value (PPV) and negative predictive value (NPV). In the 23 studies, 17 studies (Alderden et al., 2018; Alderden et al., 2021; Cai et al., 2021; Choi et al., 2020; Cichosz et al., 2019; Delparte et al., 2021; Deng et al., 2017; Goodwin & Demner-Fushman, 2020; Hu et al., 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Nakagami et al., 2021; Raju et al., 2015; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Sotoodeh et al., 2020) reported AUC, ranging from 0.68–0.99. Specifically, a value greater than 0.9 was reported in four studies (Deng et al., 2017; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Sotoodeh et al., 2020), one study (Choi et al., 2020) reported it to be was less than 0.7. Eight studies (Chen et al., 2018; Goodwin & Demner-Fushman, 2020; Ladios-Martin et al., 2020; Li et al., 2019; Setoguchi et al., 2016; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Vyas et al., 2020) reported accuracy, ranging from 0.28–0.99; 19 studies (Cai et al., 2021; Chen et al., 2018; Choi et al., 2020; Cichosz et al., 2019; Cramer et al., 2019; Delparte et al., 2021; Deng et al., 2017; Goodwin & Demner-Fushman, 2020; Hu et al., 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Li et al., 2019; Nakagami et al., 2021; Setoguchi et al., 2016; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Su et al., 2012; Vyas et al., 2020) reported SEN and/or SPE, SEN ranged from 0.08–0.99, SPE ranged from 0.63–1.00; 15 studies (Cai et al., 2021, Chen et al., 2018, Choi et al., 2020, Cichosz et al., 2019, Cramer et al., 2019, Delparte et al., 2021, Deng et al., 2017, Goodwin & Demner-Fushman, 2020, Hu et al., 2020, Kaewprag et al., 2015, 2017, Ladios-Martin et al., 2020, Nakagami et al., 2021, Song, Gao, et al., 2021, Vyas et al., 2020) reported PPV and/or NPV, PPV ranged from 0.02–1.00, NPV ranged from 0.21–0.99. Among the 16 studies that reported multiple ML models, RF (Hu et al., 2020; Raju et al., 2015; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Sotoodeh et al., 2020), LR (Choi et al., 2020; Cramer et al., 2019; Kaewprag et al., 2015; Ladios-Martin et al., 2020) and DT (Delparte et al., 2021; Deng et al., 2017) outperformed other models in five, four and two studies respectively. Similarly, ANN (Alderden et al., 2021), support vector machine (SVM) (Li et al., 2019), gradient boosting (GB) (Alderden et al., 2021), extreme gradient boosting (XBGoost) (Nakagami et al., 2021), recurrent additive network for temporal risk prediction (CANTRIP) (Goodwin & Demner-Fushman, 2020) and Mahalanobis–Taguchi system (MTS) models (Su et al., 2012) outperformed others in one study.

3.7 | Risk of bias

Table 3 shows the risk of bias assessment of the included studies. Risk of bias was mainly present in the analysis domain. Among the 23 studies, 15 studies (Alderden et al., 2021; Cai et al., 2021; Chen et al., 2018; Choi et al., 2020; Cichosz et al., 2019; Delparte et al., 2021; Deng et al., 2017; Hu et al., 2020; Kaewprag et al., 2015, 2017; Ladios-Martin et al., 2020; Li et al., 2019; Nakagami et al., 2021; Setoguchi et al., 2016; Su et al., 2012) were judged as having a high risk of bias; six studies (Alderden et al., 2018; Cramer et al., 2019; Raju et al., 2015; Song, Gao, et al., 2021; Song, Kang, et al., 2021; Vyas et al., 2020) had a moderate risk of bias; no studies had a low risk of bias; two studies (Goodwin & Demner-Fushman, 2020; Sotoodeh et al., 2020) were not assessed for quality due to the use of unstructured data. To the best of our knowledge, PROBAST is not suitable for unstructured data, and so far, there are no appropriate tools to evaluate such studies.

4 | DISCUSSION

With the development of artificial intelligence (AI) and computer technology, ML has gradually infiltrated many disciplines. Many articles on use of ML in disease diagnosis have been published. However, to date, there have been fewer ML studies in the nursing field. To the best of our knowledge, this is the first systematic review of the application of ML to HAPI prediction. Through a review of 23 studies, we found that it is meaningful to assess the model construction process of different predictive studies, which can provide a reference for the development of high-quality predictive models in the future.

In this review, 17 studies provided the AUCs of the best models. AUC, also known as the C-index, is a common indicator used to

TABLE 2 Performance and predictors of ML models

Author	Year	AUC (SD)	Accuracy (SD)	SEN(SD)	SPE(SD)	F1 score	PPV	NPV	Risk factors
Cai	2021	0.81	NA	0.08	1	NA	1	0.77	Surgery duration; Weight; Duration of the cardiopulmonary bypass procedure; Age; Disease category.
Song	2021b	Non-hospital acquired 0.92 (0.03)	0.85(0.02)	0.84(0.02)	0.85(0.02)	0.81(0.01)	NA	NA	GCS; Alb; Hb; Gait/transferring; Activity; BUN; Consciousness; Cl; SCr; Spinal cord injury.
		Hospital acquired 0.94 (0.02)	0.88(0.02)	0.87(0.03)	0.88(0.02)	0.86(0.02)	NA	NA	
Song	2021a	0.99	0.99	0.99	0.99	0.99	0.99	0.99	Age; Weight; Total intake; Total output; Temperature; SBP; GLU; Diarrhoea; Stay in bed; Restraint bands; Surgery; Norton scores; Acceptance of passive turning over; NRS 2002 scores; Diabetes; Fracture.
Alderden	2021	0.82	NA	NA	NA	0.34	NA	NA	Minimum Alb; Minimum arterial PaO_2; Surgery duration; Vasopressin infusion; Length of ICU stay prior to HAPI.
Nakagami	2021	0.80(0.02)	NA	0.78(0.03)	0.74(0.04)	NA	0.02	0.99	Difficulty in moving around; Difficulty in going up and down the stairs; Difficulty in transfer; Difficulty in standing up; Anorexia; Difficult in keeping standing position.
Delparte	2021	0.78	NA	0.93	0.63	NA	0.4	0.97	PI history; Ambulation; FIM Toileting scores; FIM Bed transfer scores.
Ladios Martin	2020	0.88	0.87	0.75	0.88	NA	0.22	0.99	Medical service; Days of oral antidiabetic agent or insulin therapy; Ability to eat; Number of red blood cell units transfused; Hb; PI present on admission; APACHE II scores.
Vyas	2020	NA	0.95	0.84	0.97	0.27	0.87	0.03	Six subscales in Braden scale: Mobility; Activity; Sensory perception; Skin moisture; Nutritional state; Friction/Shear.
Choi	2020	upper OMPI 0.82 lower OMPI 0.68	NA	upper OMPI 0.60 lower OMPI 0.85	upper OMPI 0.89 lower OMPI 0.76	NA	upper OMPI 0.23 lower OMPI 0.37	upper OMPI 0.98 lower OMPI 0.97	Bite-block or airway use; ETT holder use; Steroid use; Vasopressor use; Hct; Alb.
Hu	2020	0.86	NA	0.87	0.72	0.08	0.04	NA	Skin integrity; SBP; Expression ability; Capillary refill time; Consciousness.
Goodwin	2020	0.87	0.84	0.72	0.85	0.53	0.42	NA	NA
Sotoodeh	2020	0.95(0.01)	NA	NA	NA	0.79(0.02)	NA	NA	NA

(Continues)

TABLE 2 (Continued)

Author	Year	AUC (SD)	Accuracy (SD)	SEN(SD)	SPE(SD)	F1 score	PPV	NPV	Risk factors
Cramer	2019	NA	NA	0.49	NA	NA	0.12	NA	Stage 1 PI within first 24 hr; GCS; BUN; Arterial PaO_2; Cardiac Surg. Recovery Unit; Alb; Medical ICU; Pressure reduction device; Mechanical ventilation; Mean arterial pressure.
Cichosz	2019	0.82	NA	0.43	0.94	NA	0.72	0.92	Gender; Up and self-reliant; Limitation in activity performance; Mobility and willingness; Consciousness.
Li	2019	NA	0.79	0.81	0.79	NA	NA	NA	History of PI; Absence of cancer; Excretion; Activity/Mobility; Skin condition/Circulation.
Chen	2018	NA	0.28	0.11	0.79	0.19	0.67	0.21	Age; Disease category; Surgery duration; Perioperative corticosteroids.
Alderden	2018	Stage 1 and Higher PI: 0.79 Stage 2 and Higher PI 0.79	NA	NA	NA	NA	NA	NA	BMI; Haemoglobin; SCr; Surgery duration; Age; GLU; Lactate; Alb; GCS; Arterial SpO_2 <90%; Hypotension; Prealbumin.
Kaewprag	2017	0.830.01)	NA	0.46(0.03)	0.91(0.01)	NA	0.29(0.18)	0.95	ICD-9: 250, 403, 584, 585, 428, 785, 995, 038, 528, 482, 806, 324, 730, 290.
Deng	2017	0.93(0.88-0.97)	NA	0.86	0.82	NA	0.76	0.98	Age; Length of ICU stay; DBM; MAP; Alb; Braden scores; Mechanical ventilation; Faecal incontinence.
Setoguchi	2016	NA	0.72(0.04)	0.79(0.18)	0.72(0.04)	NA	NA	NA	Transfer activity; Surgery duration; BMI.
Raju	2015	0.83	NA	NA	NA	NA	NA	NA	Days in the hospital; Albumin; Age; Braden scores; BUN; SCr; Mobility subscale in Braden scale.
Kaewprag	2015	0.83	NA	0.16	0.99	NA	0.56	0.93	ICD-9: 344, 995, 038, 730, 785, 482, 599, 518, 112, 263; Braden scores.
Su	2012	NA	NA	0.76	0.89	0.38	NA	NA	Gender; Weight; Course; Body position during the operation; Initial body temperature; Final body temperature; Number of electronic knives used.

Abbreviations: Alb, albumin; APACHE II, Acute Physiology and Chronic Health Evaluation II; BMI, body mass index; BUN, blood urea nitrogen; Cl, chloride; DBM, diastolic blood pressure; ETT, endotracheal tube; FIM, functional independence measure; GCS, Glasgow scores; GLU, glucose; HAPI, hospital-acquired pressure injury; Hct, haematocrit; ICD-9, international classification of diseases; MAP, mean arterial pressure; NA, not available; OMPI, oral mucosal pressure injury; PI, pressure injury; SBP, systolic blood pressure; SCr, serum creatinine.

TABLE 3 Risk of bias assessment for the included studies

	Participants bias	Predictors bias	Outcome bias	Analysis bias	Overall bias
Cai et al., 2021	Low	Unclear	Unclear	High	High
Song, Kang, et al., 2021	Low	Unclear	Unclear	Unclear	Unclear
Song, Gao, et al., 2021	Low	Unclear	Unclear	Unclear	Unclear
Nakagami et al., 2021	Low	Unclear	Unclear	High	High
Alderden et al., 2021	Low	Unclear	Unclear	High	High
Delparte et al., 2021	Unclear	Unclear	Unclear	High	High
Ladios Martin et al., 2020	Low	Unclear	Unclear	High	High
Vyas et al., 2020	Unclear	Unclear	Unclear	Unclear	Unclear
Choi et al., 2020	Low	Unclear	Unclear	High	High
Hu et al., 2020	Low	Unclear	Unclear	High	High
Cramer et al., 2019	Low	Unclear	Unclear	Unclear	Unclear
Cichosz et al., 2019	Low	Unclear	Unclear	High	High
Li et al., 2019	Low	Unclear	High	Unclear	High
Chen et al., 2018	Low	Unclear	Unclear	High	High
Alderden et al., 2018	Low	Unclear	Unclear	Unclear	Unclear
Kaewprag et al., 2017	Low	Unclear	Unclear	High	High
Deng et al., 2017	Low	Unclear	Unclear	High	High
Setoguchi et al., 2016	Unclear	Unclear	Unclear	High	High
Raju et al., 2015	Low	Unclear	Unclear	Unclear	Unclear
Kaewprag et al., 2015	Low	Unclear	Unclear	High	High
Su et al., 2012	Unclear	Unclear	Unclear	High	High

measure the performance of predictive models, and its value ranges from 0–1. A value of 0 means that the prediction is completely inaccurate, while that of 1 indicates perfect prediction performance. Mandrekar claims that the model is considered acceptable when AUC is between 0.7–0.8, the model is considered excellent when AUC is between 0.8–0.9 and the model is considered outstanding when AUC is greater than 0.9 (Mandrekar, 2010). Based on this standard, 16 models were accepted, of which 11 were excellent, and four were outstanding. Of the four outstanding models, three were based on RF and one was based on DT. This shows that ML, especially the tree models, seems to have a higher accuracy in the prediction of the HAPI.

In this review, except for two studies that used unstructured data, all the other studies used structured data from EHRs and public databases. However, some studies have suggested that due to the existence of missing values, outliers and the curse of dimensionality, such data sources are not of high quality (Lee & Yoon, 2017). In addition, it is not clear how the records in these databases were measured and recorded and whether they were homogeneous (Gianfrancesco et al., 2018). For example, whether two variables with different names are actually the same indicator, whether the blood samples were collected within a similar time and analysed by analysers of the same brand and model, and whether clinical scores were judged by medical staff with similar clinical experience according to the same criteria. Excellent models originate from high-quality data. Currently, poor data quality is a major problem. Therefore, future studies should consider standardizing the establishment and management of databases to ensure the reliability of recorded data and lay the foundation for the establishment of high-quality predictive models. In addition, the incidence of PI showed a high degree of heterogeneity among the studies included in this systematic review, and we found that only a minority of studies reported PI risk assessment tools and the corresponding preventive measures. Future studies are suggested to disclose more information about risk assessment tools and preventive measures for high-risk patients with PI, which can make the studies more informative and valuable for nursing facilities with similar care standards and processes.

During actual model construction, data pre-processing is a critical and time-consuming task. Data pre-processing includes data non-dimensionalization, data coding and missing value filling, which account for more than 50% of the total data mining time (Chapman et al., 2000). However, in this review, only some of the studies reported the pre-processing of missing values and outliers, the coding process of category variables and some of them directly deleted patient records containing missing values. This method is not rigorous, because even for records with missing values, other features may still contain important information to predict the outcome. In addition, the significance of predictive models is to efficiently identify high-risk patients and implement intervention measures to prevent the occurrence of dangerous outcomes. In this review, the incidence of HAPIs varied among different studies. In some studies with low PI incidence and imbalanced datasets, the PPV of the model was low, which would make some low-risk patients be diagnosed as high risk, after which interventions may be applied to them, resulting in an

unnecessary waste of resources. Therefore, it is recommended that future researchers seek effective methods for data pre-processing and processing imbalanced datasets, which can improve the quality of modelling data and reduce the influence of noise on the models.

Although ML has shown excellent performance in predictive tasks, the "reproducibility crisis" has increasingly affected the promotion and application of this powerful tool in clinical practice. Medical ML studies consider only the application of ML to medicine, but also the development of practical ML tools that can be widely used in clinical practice. In this review, most studies used cross-validation for internal validation. Nearly half of the studies did not report model validation methods, and none of the studies used external validation. It is well known that an independent external validation queue is crucial for the generalizability of the model (Nieboer et al., 2016). The lack of external validation makes it impossible for the users of these models to determine whether they can show similar performance in different clinical environments, which limits the practicality of ML in the clinical field. In this systematic review, we found that different facilities have different and diverse predictors, which complicates the external validation process. Therefore, we suggest that future researchers carry out joint multi-centre studies with institutions in different regions and develop ML models based on common risk factors to establish a HAPI predictive model that is suitable for different regions with varying populations. In addition, most studies did not report the details of the model construction such as hyperparameter selection, and only five studies disclosed the source code. Because the tuning of the model mostly relies on the experience of the programmers, the lack of information also makes it difficult to reproduce these models. Therefore, it is recommended that future studies pay more attention to the description of tuning details, and even disclose the model code to peers to improve the reproducibility of the ML predictive models.

5 | LIMITATION

This study had some limitations. First, we included only English-language literature published since 2010, which may have led to potential publication bias. Second, the overall result of the studies' quality assessment is poor, which may have biased the results of this review to some extent. Finally, owing to the use of different indicators to evaluate model performance in different studies, we did not conduct a meta-analysis of a specific indicator.

6 | CONCLUSION

In conclusion, as an emerging predictive method, ML has gradually become a research hotspot for HAPI prediction and has shown great potential. However, in the process of constructing practical models that can be applied in the clinical field, especially in terms of data management, data pre-processing and model validation, many deficiencies still need to be addressed.

7 | RELEVANCE TO CLINICAL PRACTICE

Compared with other meta-analyses and system reviews of ML studies that focused on summarizing model performance and predictors, this review highlights the process of model construction. ML is helpful in predicting HAPI; however, in the processes of data management, data pre-processing and model validation, some deficiencies still exist. First, high-quality data are the source for all high-quality models. Therefore, researchers should consider whether the data source is accurate and reliable before constructing a predictive model. Second, the clinical data often contain significant noise. Scientific and rigorous pre-processing can minimize the loss of effective data. Finally, independent validation queues guarantee the model stability and generalizability. The ultimate goal of integrating ML into HAPI prediction is to develop a practical clinical decision-making tool. Following a complete and rigorous model construction process is essential in developing high-quality ML models.

AUTHOR CONTRIBUTIONS

The design of this review: You Zhou, Yuan Yuan and Mingquan Yan; Literature retrieval: You Zhou and Xiaoxi Yang; Literature screening: You Zhou and Xiaoxi Yang; Quality assessment: You Zhou and Shuli Ma; Manuscript writing: You Zhou and Xiaoxi Yang; Manuscript approval: Yuan Yuan and Mingquan Yan.

ACKNOWLEDGEMENTS

The authors would like to sincerely thank Yuan Yuan and Mingquan Yan for their help in guiding this paper.

CONFLICT OF INTEREST

None declared.

ETHICAL APPROVAL

This systematic review did not require ethical approval.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

You Zhou https://orcid.org/0000-0002-1352-0211
Shuli Ma https://orcid.org/0000-0003-0382-1149

REFERENCES

Akcay, M., Etiz, D., Celik, O., & Ozen, A. (2020). Evaluation of prognosis in nasopharyngeal cancer using machine learning. *Technology in Cancer Research & Treatment*, *19*, 1533033820909829. https://doi.org/10.1177/1533033820909829

Alderden, J., Drake, K. P., Wilson, A., Dimas, J., Cummins, M. R., & Yap, T. L. (2021). Hospital acquired pressure injury prediction in surgical critical care patients. *BMC Medical Informatics and Decision Making*, *21*(1), 12. https://doi.org/10.1186/s12911-020-01371-z

Alderden, J., Pepper, G. A., Wilson, A., Whitney, J. D., Richardson, S., Butcher, R., Jo, Y., & Cummins, M. R. (2018). Predicting pressure

injury in critical care patients: A machine-learning model. *American Journal of Critical Care: An Official Publication, American Association of Critical-Care Nurses*, *27*(6), 461–468. https://doi.org/10.4037/ajcc2018525

AlJame, M., Imtiaz, A., Ahmad, I., & Mohammed, A. (2021). Deep forest model for diagnosing COVID-19 from routine blood tests. *Scientific Reports*, *11*(1), 16682. https://doi.org/10.1038/s41598-021-95957-w

Barber, E. L., Garg, R., Persenaire, C., & Simon, M. (2021). Natural language processing with machine learning to predict outcomes after ovarian cancer surgery. *Gynecologic Oncology*, *160*(1), 182–186. https://doi.org/10.1016/j.ygyno.2020.10.004

Cabitza, F., & Campagner, A. (2021). The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *International Journal of Medical Informatics*, *153*, 104510. https://doi.org/10.1016/j.ijmedinf.2021.104510

Cai, J. Y., Zha, M. L., Song, Y. P., & Chen, H. L. (2021). Predicting the development of surgery-related pressure injury using a machine learning algorithm model. *The Journal of Nursing Research: JNR*, *29*(1), e135. https://doi.org/10.1097/JNR.0000000000000411

Chapman, P., Clinton, J., Kerber, R., Kh Ab Aza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0 step-by-step data mining guide. *SPSS inc*, *9*(13), 1–73.

Chen, H. L., Yu, S. J., Xu, Y., Yu, S. Q., Zhang, J. Q., Zhao, J. Y., Liu, P., & Zhu, B. (2018). Artificial neural network: A method for prediction of surgery-related pressure injury in cardiovascular surgical patients. *Journal of Wound, Ostomy, and Continence Nursing: Official Publication of The Wound, Ostomy and Continence Nurses Society*, *45*(1), 26–30. https://doi.org/10.1097/WON.0000000000000388

Choi, B. K., Kim, M. S., & Kim, S. H. (2020). Risk prediction models for the development of oral-mucosal pressure injuries in intubated patients in intensive care units: A prospective observational study. *Journal of Tissue Viability*, *29*(4), 252–257. https://doi.org/10.1016/j.jtv.2020.06.002

Chou, R., Dana, T., Bougatsos, C., Blazina, I., Starmer, A. J., Reitel, K., & Buckley, D. I. (2013). Pressure ulcer risk assessment and prevention: A systematic comparative effectiveness review. *Annals of Internal Medicine*, *159*(1), 28–38. https://doi.org/10.7326/0003-4819-159-1-201307020-00006

Cichosz, S. L., Voelsang, A. B., Tarnow, L., Hasenkam, J. M., & Fleischer, J. (2019). Prediction of in-hospital pressure ulcer development. *Advances in Wound Care*, *8*(1), 1–6. https://doi.org/10.1089/wound.2018.0803

Cramer, E. M., Seneviratne, M. G., Sharifi, H., Ozturk, A., & Hernandez-Boussard, T. (2019). Predicting the incidence of pressure ulcers in the intensive care unit using machine learning. *EGEMS (Washington, DC)*, *7*(1), 49. https://doi.org/10.5334/egems.307

Das, A., Connell, M., & Khetarpal, S. (2021). Digital image analysis of ultrasound images using machine learning to diagnose pediatric nonalcoholic fatty liver disease. *Clinical Imaging*, *77*, 62–68. https://doi.org/10.1016/j.clinimag.2021.02.038

De Silva, K., Mathews, N., Teede, H., Forbes, A., Jönsson, D., Demmer, R. T., & Enticott, J. (2021). Clinical notes as prognostic markers of mortality associated with diabetes mellitus following critical care: A retrospective cohort analysis using machine learning and unstructured big data. *Computers in Biology and Medicine*, *132*, 104305. https://doi.org/10.1016/j.compbiomed.2021.104305

Delparte, J. J., Flett, H. M., Scovil, C. Y., & Burns, A. S. (2021). Development of the spinal cord injury pressure sore onset risk screening (SCI-PreSORS) instrument: A pressure injury risk decision tree for spinal cord injury rehabilitation. *Spinal Cord*, *59*(2), 123–131. https://doi.org/10.1038/s41393-020-0510-y

Deng, X., Yu, T., & Hu, A. (2017). Predicting the risk for hospital-acquired pressure ulcers in critical care patients. *Critical Care Nurse*, *37*(4), e1–e11. https://doi.org/10.4037/ccn2017548

Edsberg, L. E., Black, J. M., Goldberg, M., McNichol, L., Moore, L., & Sieggreen, M. (2016). Revised national pressure ulcer advisory panel pressure injury staging system: Revised pressure injury staging system. *Journal of Wound, Ostomy, and Continence Nursing: Official Publication of The Wound, Ostomy and Continence Nurses Society*, *43*(6), 585–597. https://doi.org/10.1097/WON.0000000000000281

Gianfrancesco, M. A., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *JAMA Internal Medicine*, *178*(11), 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763

Goodwin, T. R., & Demner-Fushman, D. (2020). A customizable deep learning model for nosocomial risk prediction from critical care notes with indirect supervision. *Journal of the American Medical Informatics Association: JAMIA*, *27*(4), 567–576. https://doi.org/10.1093/jamia/ocaa004

Hu, Y.-H., Lee, Y.-L., Kang, M.-F., & Lee, P.-J. (2020). Constructing inpatient pressure injury prediction models using machine learning techniques. *Computers, Informatics, Nursing: CIN*, *38*(8), 415–423. https://doi.org/10.1097/CIN.0000000000000604

Jiang, M., Ma, Y., Guo, S., Jin, L., Lv, L., Han, L., & An, N. (2021). Using machine learning technologies in pressure injury management: Systematic review. *JMIR Medical Informatics*, *9*(3), e25704. https://doi.org/10.2196/25704

Kaewprag, P., Newton, C., Vermillion, B., Hyun, S., Huang, K., & Machiraju, R. (2015). predictive modeling for pressure ulcers from intensive care unit electronic health records. *AMIA Joint Summits on Translational Science*, *2015*, 82–86.

Kaewprag, P., Newton, C., Vermillion, B., Hyun, S., Huang, K., & Machiraju, R. (2017). Predictive models for pressure ulcers from intensive care unit electronic health records using Bayesian networks. *BMC Medical Informatics and Decision Making*, *17*(Suppl. 2), 65. https://doi.org/10.1186/s12911-017-0471-z

Kambakamba, P., Mannil, M., Herrera, P. E., Müller, P. C., Kuemmerli, C., Linecker, M., von Spiczak, J., Hüllner, M. W., Raptis, D. A., Petrowsky, H., Clavien, P.-A., & Alkadhi, H. (2020). The potential of machine learning to predict postoperative pancreatic fistula based on preoperative, non-contrast-enhanced CT: A proof-of-principle study. *Surgery*, *167*(2), 448–454. https://doi.org/10.1016/j.surg.2019.09.019

Koga, S., Ikeda, A., & Dickson, D. W. (2021). Deep learning-based model for diagnosing Alzheimer's disease and tauopathies. *Neuropathology and Applied Neurobiology*, *48*(1), e12759. https://doi.org/10.1111/nan.12759

Ladios-Martin, M., Fernández-De-maya, J., Ballesta-López, F. J., Belso-Garzas, A., Mas-Asencio, M., & Cabañero-Martínez, M. J. (2020). Predictive modeling of pressure injury risk in patients admitted to an intensive care unit. *American Journal of Critical Care: An Official Publication, American Association of Critical-Care Nurses*, *29*(4), e70–e80. https://doi.org/10.4037/ajcc2020237

Lee, C. H., & Yoon, H.-J. (2017). Medical big data: Promise and challenges. *Kidney Research and Clinical Practice*, *36*(1), 3–11. https://doi.org/10.23876/j.krcp.2017.36.1.3

Li, H. L., Lin, S. W., & Hwang, Y. T. (2019). Using nursing information and data mining to explore the factors that predict pressure injuries for patients at the end of life. *Computers, Informatics, Nursing: CIN*, *37*(3), 133–141. https://doi.org/10.1097/CIN.0000000000000489

Li, X., Yang, L., Yuan, Z., Lou, J., Fan, Y., Shi, A., Huang, J., Zhao, M., & Wu, Y. (2021). Multi-institutional development and external validation of machine learning-based models to predict relapse risk of pancreatic ductal adenocarcinoma after radical resection. *Journal of Translational Medicine*, *19*(1), 281. https://doi.org/10.1186/s12967-021-02955-7

Lyder, C. H., Wang, Y., Metersky, M., Curry, M., Kliman, R., Verzier, N. R., & Hunt, D. R. (2012). Hospital-acquired pressure ulcers: Results

from the national medicare patient safety monitoring system study. *Journal of the American Geriatrics Society*, *60*(9), 1603–1608. https://doi.org/10.1111/j.1532-5415.2012.04106.x

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer*, *5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Mangold, C., Zoretic, S., Thallapureddy, K., Moreira, A., Chorath, K., & Moreira, A. (2021). Machine learning models for predicting neonatal mortality: A systematic review. *Neonatology*, *118*(4), 394–405. https://doi.org/10.1159/000516891

Mervis, J. S., & Phillips, T. J. (2019). Pressure ulcers: Pathophysiology, epidemiology, risk factors, and presentation. *Journal of the American Academy of Dermatology*, *81*(4), 881–890. https://doi.org/10.1016/j.jaad.2018.12.069

Mohammed, A., Podila, P. S. B., Davis, R. L., Ataga, K. I., Hankins, J. S., & Kamaleswaran, R. (2020). Using machine learning to predict early onset acute organ failure in critically ill intensive care unit patients with sickle cell disease: Retrospective study. *Journal of Medical Internet Research*, *22*(5), e14693. https://doi.org/10.2196/14693

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

Nieboer, D., van der Ploeg, T., & Steyerberg, E. W. (2016). Assessing discriminative performance at external validation of clinical prediction models. *PloS one*, *11*(2), e0148820. https://doi.org/10.1371/journal.pone.0148820

Padula, W. V., Wald, H. M., & Meltzer, D. O. (2015). Pressure ulcer prevention and management. *Annals of Internal Medicine*, *163*(2), 148–149. https://doi.org/10.7326/L15-5110

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Raju, D., Su, X., Patrician, P. A., Loan, L. A., & McCarthy, M. S. (2015). Exploring factors associated with pressure ulcers: A data mining approach. *International Journal of Nursing Studies*, *52*(1), 102–111. https://doi.org/10.1016/j.ijnurstu.2014.08.002

Reddy, M., Gill, S. S., & Rochon, P. A. (2006). Preventing pressure ulcers: A systematic review. *JAMA*, *296*(8), 974–984. https://doi.org/10.1001/jama.296.8.974

Setoguchi, Y., Ghaibeh, A. A., Mitani, K., Abe, Y., Hashimoto, I., & Moriguchi, H. (2016). Predictability of pressure ulcers based on operation duration, transfer activity, and body mass index through the use of an alternating decision tree. *The Journal of Medical Investigation: JMI*, *63*(3-4), 248–255. https://doi.org/10.2152/jmi.63.248

Shi, C., Dumville, J. C., & Cullum, N. (2019). Evaluating the development and validation of empirically-derived prognostic models for pressure ulcer risk assessment: A systematic review. *International Journal of Nursing Studies*, *89*, 88–103. https://doi.org/10.1016/j.ijnurstu.2018.08.005

Song, J., Gao, Y., Yin, P., Li, Y., Li, Y., Zhang, J., Su, Q., Fu, X., & Pi, H. (2021). The random forest model has the best accuracy among the four pressure ulcer prediction models using machine learning algorithms. *Risk Management and Healthcare Policy*, *14*, 1175–1187. https://doi.org/10.2147/RMHP.S297838

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

Sotoodeh, M., Gero, Z. H., Zhang, W., Hertzberg, V. S., & Ho, J. C. (2020). Pressure ulcer injury in unstructured clinical notes: Detection and interpretation. *Annual Symposium Proceedings. AMIA Symposium*, *2020*, 1160–1169.

Spilsbury, K., Nelson, A., Cullum, N., Iglesias, C., Nixon, J., & Mason, S. (2007). Pressure ulcers and their treatment and effects on quality of life: Hospital inpatient perspectives. *Journal of Advanced Nursing*, *57*(5), 494–504. https://doi.org/10.1111/j.1365-2648.2006.04140.x

Su, C. T., Wang, P. C., Chen, Y. C., & Chen, L. F. (2012). Data mining techniques for assisting the diagnosis of pressure ulcer development in surgical patients. *Journal of Medical Systems*, *36*(4), 2387–2399. https://doi.org/10.1007/s10916-011-9706-1

Vyas, K., Samadani, A., Milosevic, M., Ostadabbas, S., & Parvaneh, S. (2020). Additional value of augmenting current subscales in braden scale with advanced machine learning technique for pressure injury risk assessment. *International Conference on Bioinformatics and Biomedicine*, *2020*(BIBM2020), 2993–2995.

Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. *Annals of Internal Medicine*, *170*(1), 51–58. https://doi.org/10.7326/M18-1376

SUPPORTING INFORMATION

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

How to cite this article: Zhou, Y., Yang, X., Ma, S., Yuan, Y., & Yan, M. (2023). A systematic review of predictive models for hospital-acquired pressure injury using machine learning. *Nursing Open*, *10*, 1234–1246. https://doi.org/10.1002/nop2.1429