International Journal of Nursing Studies 138 (2023) 104411

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/locate/ns

Diagnostic accuracy of clinical outcome prediction using nursing data in intensive care patients: A systematic review☆

Mihui Kim [a], Sangwoo Park [b], Changhwan Kim [c], Mona Choi [b,d,*]

[a] College of Nursing and Brain Korea 21 FOUR Project, Yonsei University, Seoul, Republic of Korea
[b] College of Nursing and Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, Republic of Korea
[c] School of Nursing, Johns Hopkins University, Baltimore, MD, United States of America
[d] Yonsei Evidence Based Nursing Centre of Korea, A JBI Affiliated Group, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 15 May 2022
Received in revised form 17 September 2022
Accepted 22 November 2022

Keywords:
Clinical outcome
Diagnostic accuracy
Intensive care units
Length of stay
Mortality
Nursing data
Prediction
Readmission

ABSTRACT

Background: Nursing data consist of observations of patients' conditions and information on nurses' clinical judgment based on critically ill patients' behavior and physiological signs. Nursing data in electronic health records were recently emphasized as important predictors of patients' deterioration but have not been systematically reviewed.
Objective: We conducted a systematic review of prediction models using nursing data for clinical outcomes, such as prolonged hospital stay, readmission, and mortality in intensive care patients, compared to physiological data only. In addition, the type of nursing data used in prediction model developments was investigated.
Design: A systematic review.
Methods: PubMed, CINAHL, Cochrane CENTRAL, EMBASE, IEEE Xplore Digital Library, Web of Science, and Scopus were searched. Clinical outcome prediction models using nursing data for intensive care patients were included. Clinical outcomes were prolonged hospital stay, readmission, and mortality. Data were extracted from selected studies such as study design, data source, outcome definition, sample size, predictors, reference test, model development, model performance, and evaluation. The risk of bias and applicability was assessed using the Prediction model Risk of Bias Assessment Tool checklist. Descriptive summaries were produced based on paired forest plots and summary receiver operating characteristic curves.
Results: Sixteen studies were included in the systematic review. The data types of predictors used in prediction models were categorized as physiological data, nursing data, and clinical notes. The types of nursing data consisted of nursing notes, assessments, documentation frequency, and flowsheet comments. The studies using physiological data as a reference test showed higher predictive performance in combined data or nursing data than in physiological data. The overall risk of bias indicated that most of the included studies have a high risk.
Conclusions: This study was conducted to identify and review the diagnostic accuracy of clinical outcome prediction using nursing data in intensive care patients. Most of the included studies developed models using nursing notes, and other studies used nursing assessments, documentation frequency, and flowsheet comments. Although the findings need careful interpretation due to the high risk of bias, the area under the curve scores of nursing data and combined data were higher than physiological data alone. It is necessary to establish a strategy in prediction modeling to utilize nursing data, clinical notes, and physiological data as predictors, considering the clinical context rather than physiological data alone.
Registration: The protocol for this study is registered with PROSPERO (registration number: CRD42021273319).

What is already known

- Physiological data is regarded as the strongest predictor in clinical diagnostic or prognostic models.
- Nursing data combined with physiological data are emphasized in detecting physiological deterioration in patients.
- Recently, nursing data have been emphasized as important predictors of deteriorating patient health but have not been systematically reviewed.

☆ Mona Choi, Mihui Kim, Sangwoo Park, and Changhwan Kim made equal contributions to this manuscript.

* Corresponding author at: Yonsei University College of Nursing, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea.
E-mail addresses: ystelra50@gmail.com (M. Kim), SWP875@yuhs.ac (S. Park), zchgood@naver.com (C. Kim), monachoi@yuhs.ac (M. Choi).

https://doi.org/10.1016/j.ijnurstu.2022.104411

What this paper adds

- The type of nursing data used in prediction models, including nursing notes, assessments, documentation frequency, and flowsheet comments.
- The prediction models' performance using nursing and combined data is relatively higher than solely using physiological data to discriminate clinical outcomes in intensive care patients.

1. Background

In the intensive care unit, consistent supervision using monitoring devices and timely intervention are crucial for managing the various clinical conditions among patients (Calvert et al., 2016). Electronic medical records preserve patients' information generated by monitors, medical devices, and healthcare providers (De Georgia et al., 2015), and primarily contain two types of data, namely structured or coded data and unstructured data, such as free text (Abhyankar et al., 2014). The vast availability of information stored in electronic medical records facilitates complex decision-making processes regarding patient management and enhances the accuracy of patient prognosis and disease prediction (Calvert et al., 2016; De Georgia et al., 2015; Khadanga et al., 2019; Rocker et al., 2004). The early detection of clinical conditions can facilitate the implementation of effective interventions to mitigate deteriorating patient health (Capan et al., 2017; Korach et al., 2019), such as prolonged intensive care unit stay, readmission, and mortality.

Electronic medical records show that intensive care patients often display warning signs hours before the physiological deterioration (Chan et al., 2010; Jones et al., 2011). Therefore, there is time to identify high-risk patients in a hospital setting (Jones et al., 2011). The prediction models are intended to detect patients' deterioration based on predefined criteria using information recorded in electronic medical records (Jones et al., 2011) and to provide early intervention (Collins et al., 2015; El-Rashidy et al., 2020). However, severity scoring systems, such as Acute Physiology and Chronic Health Evaluation, Sequential Organ Failure Assessment, and Simplified Acute Physiology Score, using abnormal vital signs, physiologic variables, and laboratory results do not always reflect patient condition due to the varying optimal criteria influenced by their context (Jeddah et al., 2021; Jeong, 2018; Jones et al., 2011; Marafino et al., 2018).

Nursing data comprise observations and information on nurses' clinical judgment through the interpretation of patients' behavior and physiological signs (Kang et al., 2020; Odell et al., 2009; Yu et al., 2020). Therefore, nurses document more frequently when they are concerned about the patient's condition (Collins et al., 2013; Collins and Vawdrey, 2012). Previous studies have indicated that increased free-text comments about vital signs, additional measurements of vital signs, withheld medications, and as-needed medication administrations were associated with patient mortality and cardiac arrest (Collins et al., 2013; Collins and Vawdrey, 2012; Schnock et al., 2021). These optional documentation features in nursing data with in-depth contextual information may reflect a nurse's concern about patients' deterioration (Collins et al., 2013; Kang et al., 2020) and predict the risk of future occurrence of certain events (Collins et al., 2015).

A highly accurate prognosis prediction model developed advanced machine learning approaches using feature-rich data in electronic medical records (Churpek et al., 2016; Wellner et al., 2017). In addition, accuracy in clinical prognosis predictions can improve by analyzing all data types, such as structured, unstructured, and images available in electronic medical records (Marafino et al., 2018). Moreover, several prediction models using nursing data demonstrated an increased accuracy compared to the sole use of physiological data in electronic medical records (Douw et al., 2015; Huang et al., 2021).

Although recent studies have emphasized the use of nursing data in determining patient deterioration and developing prediction models (Huang et al., 2021; Korach et al., 2020; Tran and Lee, 2018), the findings have not been systematically reviewed. Therefore, this study conducted a systematic review on prediction models using nursing data of clinical outcomes, such as prolonged intensive care unit stay, readmission, and mortality in intensive care patients, compared to solely using physiological data. In addition, the type of nursing data used in prediction model developments was investigated.

2. Materials and methods

The protocol for this study was registered with PROSPERO (registration number: CRD42021273319).

2.1. Eligibility criteria

The eligible studies were searched to evaluate the diagnostic accuracy of predicting clinical outcomes of intensive care patients using nursing data compared to physiological data. Thus, all studies that used nursing data to predict clinical outcomes in intensive care patients using machine learning techniques were included. Additionally, all studies that performed development or validation with retrospective or prospective study designs were included. The recommended Population, Index test, Reference test, and Diagnosis of interest (PIRD) structure (Campbell et al., 2015) was used to describe the eligibility criteria as follows: (1) Population was intensive care patients; (2) index test was based on prediction models using nursing data. The nursing data included both actual and modified values of nursing notes (free text type) and nursing assessments (structured or semi-structured type) written by nurses; (3) reference test refers to prediction models using data related to physiological indicators such as patients' vital signs and laboratory test results; and (4) diagnoses of interest were clinical outcomes including prolonged intensive care unit stay, readmission, and mortality. These outcomes included values of sensitivity and specificity.

2.2. Information sources and search strategy

A systematic literature search was conducted from the inception date of databases to November 17, 2021, using seven electronic databases, including PubMed, CINAHL, Cochrane CENTRAL, EMBASE, IEEE Xplore Digital Library, Web of Science, and Scopus, with no publication date limitation. Search terms were developed from a combination of medical subject headings (MeSH) terms, keywords, and Boolean operators in consultation with a professional medical librarian (Supplement 1). The following keyword combinations were used: (1) (Patients [Mesh] OR patient*) AND ("Intensive Care Units"[Mesh] OR "intensive care" OR "critical care"); (2) (nursing AND (unstructured OR narrative OR "free-text")) OR "Nursing Records"[Mesh] OR "nursing record*" OR "Nursing Assessment"[Mesh] OR "nursing assessment" OR "nursing note*"; (3) "ROC Curve"[Mesh] OR ROC OR "Sensitivity and Specificity"[Mesh] OR (Sensitivity AND Specificity) OR "prediction model*" OR "Machine Learning"[Mesh] OR "machine learning"; and (4) "Treatment Outcome"[Mesh] OR "Clinical outcome*" OR "Clinical Deterioration"[Mesh] OR deterioration* OR "Patient Readmission"[Mesh] OR readmission OR "Length of Stay"[Mesh] OR "length of stay" OR "Mortality"[Mesh] OR mortality OR "Death"[Mesh] OR death.

2.3. Study selection

The studies were screened independently by the two authors (M.K. and S.P.) based on the title and abstract for eligibility after discarding duplicated studies. Upon excluding studies based on the inclusion and exclusion criteria, full texts were reviewed. Furthermore, the reference lists of all eligible studies were screened to identify additional relevant studies. When required, disagreements regarding decisions were resolved by three authors (M.K., S.P., and M.C.) through discussion.

The inclusion criteria for the studies were as follows: (1) included intensive care patients; (2) conducted the development studies using machine learning techniques to predict clinical outcomes; and (3) included prolonged intensive care unit stay, readmission, or mortality for clinical outcomes. The exclusion criteria were as follows: (1) not using nursing data as predictors for prediction models; (2) not adequately describing the statistical methods of the model building process; (3) not published in English; and (4) review article or poster abstract.

2.4. Data extraction

The data extraction from the included studies was performed independently by the authors. The datasheet form was based on the critical appraisal and data extraction for systematic reviews of prediction modeling studies (CHARMS) checklist (Moons et al., 2014). Specifically, it contained fields for study design, source of data, outcome definition (outcome and follow-up duration), sample size (number of samples and number of events), predictors (type of data, type of nursing data, and timing of predictor measurement), reference test, model development (best machine learning technique), model performance (best discrimination and calibration), and model evaluation (validation).

In model performance, both discrimination and calibration properties of the models should be reported regardless of the machine learning techniques (Moons et al., 2014). Discrimination refers to how well a prediction model can discriminate those with the outcome from those without, and calibration refers to how well outcomes are predicted compared to the observed outcomes (Steyerberg, 2019). The discrimination results were categorized as nursing, physiological, and combined data based on predictor type. Physiological data refers to vital signs, laboratory test results, and calculated scores based on physiological information. Nursing data written by nurses include nursing assessment records and nursing notes. Combined data consists of physiological data, nursing data, or clinical notes recorded by nurses, physicians, and therapists.

Model evaluation is the method related to the models' predictive performance testing regarding internal and external validation, and the methods of internal validation use resampling methods, such as bootstrap, jack-knife, or cross-validation, to reduce overfitting (Moons et al., 2014).

2.5. Risk of bias and applicability

The risk of bias and clinical applicability for each included study was assessed with the Prediction model Risk Of Bias ASsessment Tool (PROBAST) checklist (Wolff et al., 2019). The authors (M.K. and S.P.) independently assessed the presence of bias and concerns regarding the applicability of the included studies. This checklist comprises 20 signaling questions in four key domains (participants, predictors, outcomes, and analysis) to assess the risk of bias. Additionally, the applicability was assessed across three domains (participants, predictors, and outcome), aligning with the review questions (Wolff et al., 2019). The risk of bias and applicability was judged as low, high, or unclear. The other author (M.C.) resolved disagreements regarding decisions through discussion, until consensus was reached.

2.6. Data synthesis and statistical analysis

Data analyses were performed using Review Manager 5.4. Paired forest plots and summary receiver operating characteristic curves were used to graphically represent data synthesis (Campbell et al., 2015). Forest plots with 95% confidence intervals were created to determine the sensitivity and specificity of each study based on the true positive, false positive, false negative, and true negative. The I^2 statistic is not recommended in systematic reviews of diagnostic test accuracy as they do not account for the influence of differing threshold effects (Campbell et al., 2015; Macaskill et al., 2022). Therefore, paired forest plots and summary receiver operating characteristic curves were plotted to observe the visual assessment of variation between studies. When original research did not provide these values (i.e., two-by-two table), they were calculated using the reported values such as sensitivity, specificity, accuracy, precision, recall, sample size, and the number of events. The studies that could not be calculated using these data were excluded from the graphical representation.

This study was originally planned to conduct a meta-analysis based on the bivariate random-effects approach (Reitsma et al., 2005) to estimate the pooled results. However, due to the high heterogeneity among studies and the limited number of studies, a meta-analysis was not conducted. Instead, descriptive summaries and graphical representations were produced for the main outcomes.

3. Results

3.1. Study selection

Fig. 1 shows the comprehensive searching process and the results obtained through the search strategy in a PRISMA 2020 flow diagram.

In total, 699 records were retrieved from the initial electronic search. A total of 659 records remained after the removal of duplicates. After evaluating titles and abstracts, 611 studies were excluded by the authors. Furthermore, two records were included from a manual search using citations. Thereafter, a total of 50 full-text reviews were independently conducted by the authors, from which 34 studies were excluded based on the criteria. Reasons for exclusion were as follows: not a prediction model, not intensive care unit patients, not including clinical outcomes (prolonged hospital stay, readmission, and mortality), and not including nursing data as predictors. Thus, a total of 16 studies satisfied the eligibility criteria and were selected for the final review.

3.2. Study characteristics

Table 1 presents the characteristics of the included studies in alphabetical order. The majority of the included studies developed prediction models based on a retrospective study design. However, one study (Rojas et al., 2018) developed the prediction model using a prospective study design and validated the model by using retrospective data. The Medical Information Mart for Intensive Care database was used as a data source for most of the included studies, except for one study (Fu et al., 2021). The outcomes of interest were prolonged hospital stay (Huang et al., 2021; Weissman et al., 2018), readmission (Rojas et al., 2018), and mortality. Regarding the studies for mortality prediction, three studies (Ghassemi et al., 2014; Jo et al., 2015; Kumar et al., 2021) evaluated mortality by classifying multiple follow-up periods. Only three studies screened composite outcomes (Fu et al., 2021; Huang et al., 2021; Weissman et al., 2018), while most of the studies measured single outcomes, including readmission or mortality.

Table 2 summarizes the included studies. The data types of the predictors were physiological data, nursing data, and clinical notes recorded by nurses, physicians, and therapists. The nursing data types comprised nursing notes, nursing assessments, nursing documentation frequency, and flowsheet comments. The majority of the studies used nursing notes as predictors, while two studies used documentation frequency, flowsheet comments (Fu et al., 2021), and nursing assessments, such as Braden scale scores, Morse scores, abdominal physical exams, and cardiac rhythm assessments (Rojas et al., 2018).

For each study, logistic regression was the highest performing machine-learning technique, followed by gradient boosting machine and support vector machine. The best discrimination value of developed models was reported as area under the curve score or accuracy. The data types in the results were categorized as physiological data, nursing data, and clinical notes. Furthermore, combined data were grouped according

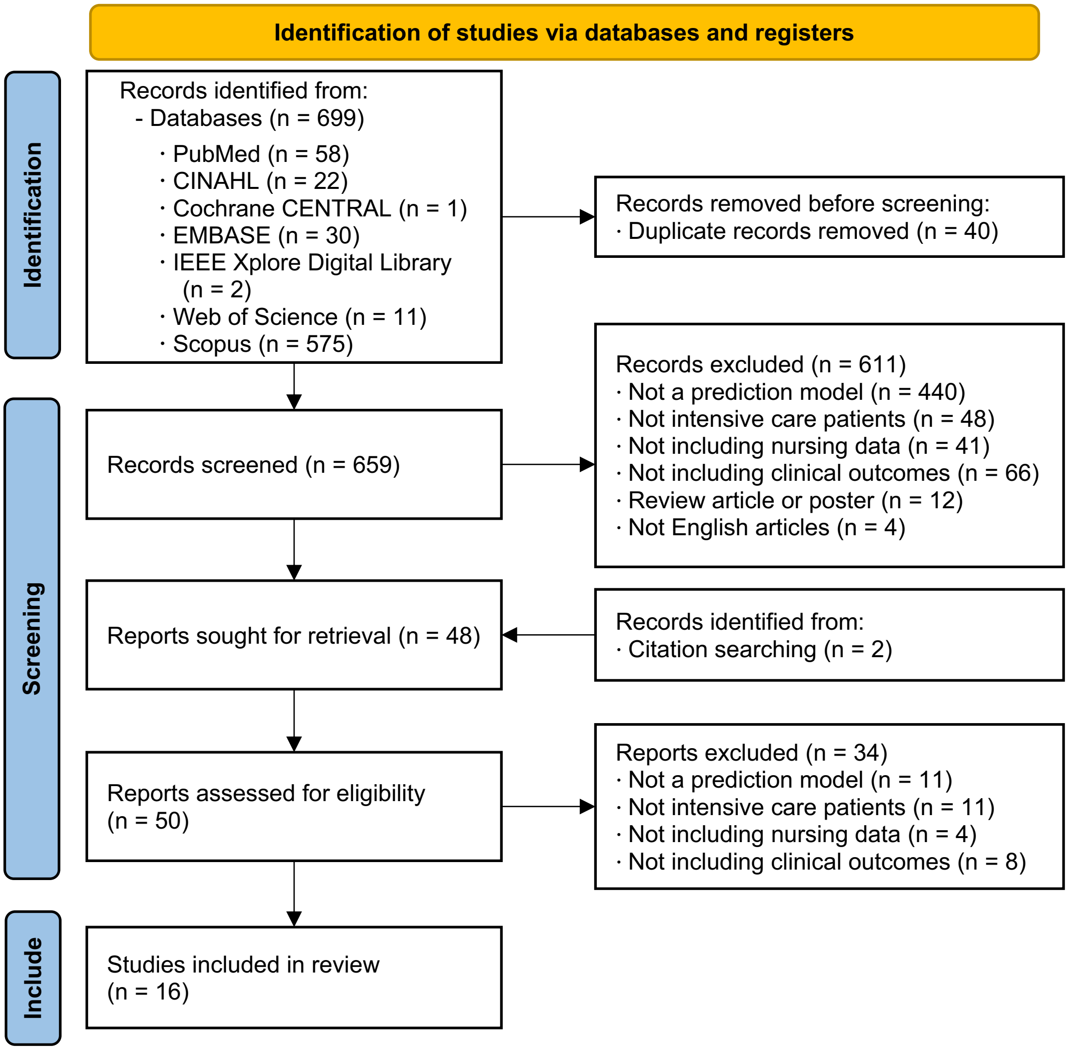

Fig. 1. PRISMA flowchart of study inclusion and exclusions.

to the data source used: (1) Combined data A included nursing data and physiological data; (2) combined data B included nursing data and clinical notes; and (3) combined data C included nursing data, physiological data, and clinical notes. The scores with the best area under the curve for physiological data, nursing data, and combined data were 0.901, 0.926, and 0.922, respectively.

The studies using physiological data as a reference test (Ghassemi et al., 2014; Hashir and Sawhney, 2020; Lehman et al., 2012; Marafino et al., 2015; Marafino et al., 2018) showed higher predictive performance in combined data or nursing data than in physiological data.

Only four studies (De Silva et al., 2021; Marafino et al., 2015; Marafino et al., 2018; Weissman et al., 2018) reported whether calibration was performed and was reported as calibration plots, Hosmer–Lemeshow test, and confusion matrix. All studies conducted internal validation using K-fold validation, while bootstrap was additionally performed for one study (Marafino et al., 2018). External validation was only performed in one study (Rojas et al., 2018).

3.3. Graphical representation

Nine of the 16 studies were included in the graphical presentation. The remaining seven were excluded due to insufficient data. Fig. 2 displays paired forest plots of the sensitivity and specificity for each predictor type: physiological, nursing, and combined data.

Only two studies evaluated physiological data, and its sensitivity varied from 64% to 100%, and the specificity varied from 11% to 67%. In nursing data (five studies), the sensitivity varied from 45% to 86% and the specificity from 70% to 91%. The observed heterogeneity was high. In four studies, data source and type of nursing data used, were the Medical Information Mart for Intensive Care database and nursing notes (sensitivity 63%–86%, specificity 70%–91%). One study used nursing documentation frequency and flowsheet comments (Fu et al., 2021). The clinical outcomes of the three studies were mortality (sensitivity 72%–86%, specificity 70%–91%), and the two studies were composite outcomes of mortality and prolonged intensive care unit stay or cardiac arrest (sensitivity 45%–63%, specificity 77%–87%; Fu et al., 2021; Huang et al., 2021).

Regarding combined data, the sensitivity estimates ranged from 28% to 95% (combined data A, 28%–79%; combined data B, 66%–94%; combined data C, 92%–95%). In addition, the specificity estimates ranged from 40% to 95% (combined data A, 72%–95%; combined data B, 40%–89%; combined data C, 56%–88%). Due to the heterogeneity among these studies being high, the pooled values were not produced. The heterogeneity was explored by outcomes and type of nursing data. In combined data A, two studies used nursing notes (sensitivity 76%–79%, specificity 72%–84%), and one used nursing assessments (Rojas et al., 2018). Regarding the outcomes in combined data B, two studies included mortality (sensitivity 86%–87%, specificity 79%–89%; De Silva et al., 2021; Ghassemi et al., 2014), and two studies included composite

Table 1
Study characteristics of the included studies ($N = 16$).

Author (year)	Study design	Source of data (period)	Outcomes	Follow-up duration	Sample size (events)	Unit
De Silva et al. (2021)	Retrospective	MIMIC-III (2001–2012)	1-year mortality	365 days	1841 (619); 5942 (2275)	Documents
Fu et al. (2021)	Retrospective	Intensive care units in the northeast region of the US (2016–2019)	Mortality, cardiac arrest, or rapid response team calls	Intensive care unit stays	6720 (161)	Admissions
Ghassemi et al. (2014)	Retrospective	MIMIC-II (2001–2008)	In-hospital and 30-/365-day mortality	In-hospital, 30-, 365-day	19,308 (not reported)	Patients
Hashir and Sawhney (2020)	Retrospective	MIMIC-III (2001–2015)	Mortality	Hospital stays	38,597 (4439)	Patients
Huang et al. (2021)	Retrospective	MIMIC-III (2008–2012)	Intensive care unit stay ≥ 7 days or mortality	2–7 days during intensive care unit stay	6521 (2341)	Admissions
Jo et al. (2015)	Retrospective	MIMIC-II (2001–2008)	1/7/30/180/365-day mortality	1-, 7-, 30-, 180-, 365-day	8808 (not reported)	Patients
Khine et al. (2019)	Retrospective	MIMIC-III (2001–2012)	30-day mortality	30 days	483,485 (145,046)	Samples
Kumar et al. (2021)	Retrospective	MIMIC-III (2001–2012)	7/30/180/365-day mortality	7-, 30-, 180-, 365-day	2346 (258/736/253/653)	Patients
Lehman et al. (2012)	Retrospective	MIMIC-II (2001–2008)	In-hospital mortality	Hospital stays	14,739 (2154)	Patients
Marafino et al. (2015)	Retrospective	MIMIC-II (2001–2008)	In-hospital mortality	Intensive care unit stays	25,826 (2099)	Patients
Marafino et al. (2018)	Retrospective	20 intensive care units at 3 hospitals (2001–2017)	In-hospital mortality	Hospital stays	101,196 (10,505)	Patients
Rojas et al. (2018)	Development: prospective Validation: retrospective	Development: University of Chicago Medical Center (2008–2016) Validation: MIMIC-III (2001–2012)	Unplanned readmission	Any time after intensive care unit discharge	Development: 24,885 (2834) Validation: 42,303 (3458)	Admissions
Tran and Lee (2018)	Retrospective	MIMIC-III (2001–2012)	30-day mortality	30 days	27,477 (3097)	Patients
Waudby-Smith et al. (2018)	Retrospective	MIMIC-III (2001–2012)	30-day mortality	30 days	27,477 (3029)	Patients
Weissman et al. (2018)	Retrospective	MIMIC-III (2001–2012)	Intensive care unit stay ≥ 7 days or mortality	2–7 days during intensive care unit stay	25,947 (5054)	Admissions
Zalewski et al. (2017)	Retrospective	MIMIC-II (2001–2008)	In-hospital mortality	Hospital stays	17,274 (not reported)	Patients

Abbreviation: MIMIC, Medical Information Mart for Intensive Care.

outcomes (sensitivity 66%–94%, specificity 40%–79%; Huang et al., 2021; Weissman et al., 2018).

Fig. 3 presents the summary receiver operating characteristic curves using a single point (sensitivity and specificity) according to predictor data types. In nursing data and combined data A, the scattered points showed more variability in estimated sensitivity than specificity across studies. Combined data B and C's plots showed more variability in estimated specificity than sensitivity across studies.

3.4. Risk of bias and applicability of studies

The PROBAST checklist was used for the quality evaluation regarding the risk of bias and applicability concerns of the included studies (Supplement 2). The overall risk of bias was high among most of the included studies. Meanwhile, most of the studies demonstrated low concern regarding applicability, except for three studies (De Silva et al., 2021; Jo et al., 2015; Kumar et al., 2021) which showed unclear concern.

4. Discussion

4.1. Summary of evidence

This study was conducted to identify and review the diagnostic accuracy of clinical outcome prediction models using nursing data in intensive care patients. This study focused on comparing prediction models using nursing data with prediction models using other data types. In addition, the types of nursing data used in prediction models were investigated. A total of 16 studies identified clinical outcome prediction models using nursing data in intensive care patients. The prediction models from the included studies were developed and validated within the last ten years. Recent advances in machine learning techniques have increased the availability of free text data used in prediction models to enhance model accuracy (Marafino et al., 2018), while several studies focused on modeling various unstructured text data.

Nursing assessments combined with vital signs and laboratory test results in electronic medical records are emphasized in detecting the physiological deterioration among patients in a hospital setting (Capan et al., 2017; Dykes et al., 2009). However, the use of nursing assessments has been used less in predicting clinical outcomes, despite their clinical importance. Most of the studies included in this study developed models using nursing notes to capture the important features. The remaining studies used nursing assessments, documentation frequency, and flowsheet comments (Fu et al., 2021; Rojas et al., 2018). Additionally, it is necessary to utilize various predictors that reflect the clinical context among the vast amount of nursing data that reflect patients' conditions.

The studies using physiological data as a reference test showed higher predictive performance in combined data or nursing data than in physiological data only. The scores with the best area under the curve were estimated to range from 72% to 93% (physiological data, 72%–90%; nursing data, 72%–93%; combined data, 76%–92%). In addition, the estimated sensitivity and specificity varied, based on data type. However, this evidence was weak due to the high risk of bias in the included studies.

Table 2
Summary of the included studies.

Author (year)	Type of data	Type of nursing data	Timing of predictor measurement	Reference test	Best machine-learning technique	Best discrimination (area under the curve scores)	Calibration	Validation
De Silva et al. (2021)	Nursing data, clinical notes	Nursing notes	Intensive care unit admission period	–	Logistic regression	Nursing data: 0.893 Combined data B[b]: 0.922	Confusion matrix	K-fold validation
Fu et al. (2021)	Nursing data	Nursing documentation frequency, flowsheet comments	From 12 to 36 h prior to discharge	–	Logistic regression	Nursing data: 0.718	–	K-fold validation
Ghassemi et al. (2014)	Physiological data, nursing data, clinical notes	Nursing notes	Within 24 h after admission	Physiological data	Support vector machine	Physiological data: 0.901 (in-hospital), 0.745 (30-day), 0.901 (365-day) Combined data B[b]: 0.944 (in-hospital), 0.783 (30-day), 0.901 (365-day) Combined data C[b]: 0.776 (in-hospital), 0.755 (30-day), 0.813 (365-day)	–	K-fold validation
Hashir and Sawhney (2020)	Physiological data, nursing data, clinical notes	Nursing notes	Within 48 h after admission	Physiological data	Neural networks	Physiological data: 0.877 Combined data B[b]: 0.888 Combined data C[b]: 0.902	–	K-fold validation
Huang et al. (2021)	Nursing data, clinical notes[a]	Nursing notes	Within 48 h after admission	–	Gradient boosting machine	Nursing data: 0.826 Combined data B[b]: 0.839	–	K-fold validation
Jo et al. (2015)	Nursing data	Nursing notes	Reference time within a given time frame	–	Support vector machine	Nursing data: 0.733 (1-day), 0.778 (7-day), 0.782 (30-day), 0.788 (180-day), 0.790 (365-day)	–	K-fold validation
Khine et al. (2019)	Physiological data, nursing data	Nursing notes	Intensive care unit admission period	–	Neural networks	Physiological data: 72%[c] Nursing data: 78%[c] Combined data A[b]: 82%[c]	–	K-fold validation
Kumar et al. (2021)	Nursing data	Nursing notes	Intensive care unit admission period	–	Extreme gradient boost	Nursing data: 0.899 (7-day), 0.926 (30-day), 0.678 (180-day), 0.607 (365-day)	–	K-fold validation
Lehman et al. (2012)	Physiological data, nursing data	Nursing notes	Within 24 h after admission	Physiological data	Logistic regression	Physiological data: 0.72 Nursing data: 0.78 Combined data A[b]: 0.82	–	K-fold validation
Marafino et al. (2015)	Physiological data, nursing data	Nursing notes	Within 24 h after admission	Physiological data	Logistic regression	Physiological data: 0.791 Combined data A[b]: 0.889	Hosmer–Lemeshow test	K-fold validation
Marafino et al. (2018)	Physiological data, nursing data, clinical notes	Nursing notes	Within 24 h after admission	Physiological data	Logistic regression	Physiological data: 0.831 Combined data C[b]: 0.922	Calibration plot, Hosmer–Lemeshow test	K-fold validation, bootstrap
Rojas et al. (2018)	Physiological data, nursing data	Nursing assessments: Braden scale score, Morse score, abdominal physical exam, cardiac rhythm	Intensive care unit admission period	–	Gradient boosting machine	Combined data A[b]: 0.76	–	K-fold validation, external validation
Tran and Lee (2018)	Physiological data, nursing data	Nursing notes	Intensive care unit admission period	–	Random forest	Combined data A[b]: 0.827	–	K-fold validation
Waudby-Smith et al. (2018)	Physiological data, nursing data	Nursing notes	Intensive care unit admission period	–	Logistic regression	Nursing data: 0.809 Combined data A[b]: 0.819	–	K-fold validation
Weissman et al. (2018)	Physiological data, nursing data, clinical notes	Nursing notes	Within 48 h after admission	–	Gradient boosting machine	Combined data B[b]: 0.83 Combined data C[b]: 0.89	Calibration plot	K-fold validation
Zalewski et al. (2017)	Physiological data, nursing data	Nursing notes	Within 24 h after admission	–	Logistic regression	Combined data A[b]: 0.80		K-fold validation

[a] Clinical notes are defined as those written by physicians and other clinicians.
[b] Combined data A includes nursing data and physiological data, B includes nursing data and clinical notes, and C includes nursing data, physiological data, and clinical notes.
[c] Accuracy.

A. Physiological data

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ghassemi et al., 2014	625	4041	2	489	1.00 [0.99, 1.00]	0.11 [0.10, 0.12]
Lehman et al., 2012	1379	4153	775	8432	0.64 [0.62, 0.66]	0.67 [0.66, 0.68]

B. Nursing data

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
De Silva et al., 2021	18080	4708	3056	13895	0.86 [0.85, 0.86]	0.75 [0.74, 0.75]
Fu et al., 2021	19	204	23	1392	0.45 [0.30, 0.61]	0.87 [0.85, 0.89]
Huang et al., 2021	148	94	86	324	0.63 [0.57, 0.69]	0.78 [0.73, 0.81]
Kumar et al., 2021	93	76	18	817	0.84 [0.76, 0.90]	0.91 [0.89, 0.93]
Lehman et al., 2012	1551	3776	603	8809	0.72 [0.70, 0.74]	0.70 [0.69, 0.71]

C. Combined data†

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Khine et al., 2019	18288	3746	4861	19403	0.79 [0.78, 0.80]	0.84 [0.83, 0.84]
Lehman et al., 2012	1637	3524	517	9061	0.76 [0.74, 0.78]	0.72 [0.71, 0.73]
Rojas et al., 2018	317	441	816	8379	0.28 [0.25, 0.31]	0.95 [0.95, 0.95]

Combined data B

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
De Silva et al., 2021	23490	5049	3378	19123	0.87 [0.87, 0.88]	0.79 [0.79, 0.80]
Ghassemi et al., 2014	537	489	90	4041	0.86 [0.83, 0.88]	0.89 [0.88, 0.90]
Huang et al., 2021	154	88	80	330	0.66 [0.59, 0.72]	0.79 [0.75, 0.83]
Weissman et al., 2018	1299	3061	77	2050	0.94 [0.93, 0.96]	0.40 [0.39, 0.41]

Combined data C

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ghassemi et al., 2014	574	566	53	3964	0.92 [0.89, 0.94]	0.88 [0.87, 0.88]
Weissman et al., 2018	1307	2259	69	2852	0.95 [0.94, 0.96]	0.56 [0.54, 0.57]

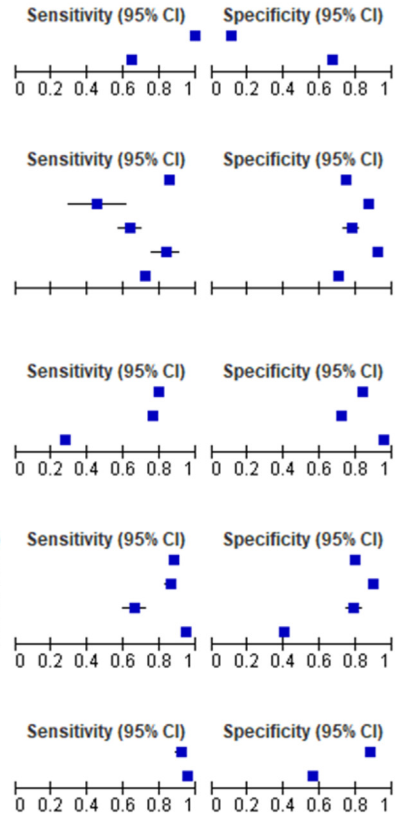

Fig. 2. Paired forest plot of the sensitivity and specificity.
†Combined data A includes nursing data and physiological data, B includes nursing data and clinical notes, and C includes nursing data, physiological data, and clinical notes.

In systematic review of diagnostic test accuracy, heterogeneity between studies is common and presumed to exist (Campbell et al., 2015; Macaskill et al., 2022). In this study, heterogeneity was observed by graphical assessment through paired forest plots and receiver operating characteristic curves. The cause of heterogeneity might be explored by data source, outcomes, and type of nursing data. Subgroup analysis based on the data source, outcomes, and nursing data type was not conducted to investigate heterogeneity because of the limited number of studies included in each data type. Further studies are required to identify covariates affecting this heterogeneity.

Nursing data reflect the subjective and objective decision-making among nurses as well as care plans for patients' conditions (Capan et al., 2017). In contrast, clinical notes written by physicians and therapists are more likely to include objective information based on physiological data, laboratory test results, and radiological images. A previous study investigated the development of prediction models for mortality in the surgical intensive care unit using physician documentation and severity of illness scores. The study found that the area under the curve scores of developed models using physician notes, the severity of illness scores, and severity of illness scores with physician notes were 0.84, 0.86, and 0.88, respectively (Parreco et al., 2018). These results suggest that the combined physician notes and severity scores improved performance in predicting mortality as opposed to the sole use of physician notes or severity of illness scores. This study showed that the combination of nursing data and clinical notes or physiological data demonstrated better predictive performance than physiological data alone. In prediction models, using nursing data and clinical notes including richer information than structured data (Waudby-Smith et al., 2018) as predictors increases the predictive power using machine learning techniques.

4.2. Limitations

There were several limitations in this systematic review. First, most of the included studies developed models using a widespread Medical Information Mart for Intensive Care database that is accessible to the public. As a result, it was limited in analyzing various types of nursing data in other clinical settings. Second, despite comprehensively searching the literature using search terms, it is possible that we may miss on our search when authors did not mention terms, such as nursing records, nursing data, nursing assessments, or nurses. Thus, researchers ought to clearly articulate the types of data as nursing, nursing documents, and nursing assessments when they use nursing data in their studies. Finally, the original studies included in this review poorly reported participants, predictors, outcomes, and analysis. This results in the risk of bias being assessed as high or unclear. Therefore, to ensure clear and accurate reporting in future research on prediction models, reporting guidelines should be followed (Collins et al., 2015).

5. Conclusions

The results of the systematic review demonstrate that prediction models using nursing data and combined data demonstrate an increased likelihood of clinical outcome prediction in intensive care patients compared to the sole use of physiological data. Furthermore, nursing data such as nursing assessment, nursing notes, and documentation frequency that reflect patient health concerns among nurses can be applied to various types of prediction models. It is necessary to establish a strategy to utilize nursing data, other clinical notes, and physiological data as predictors considering clinical

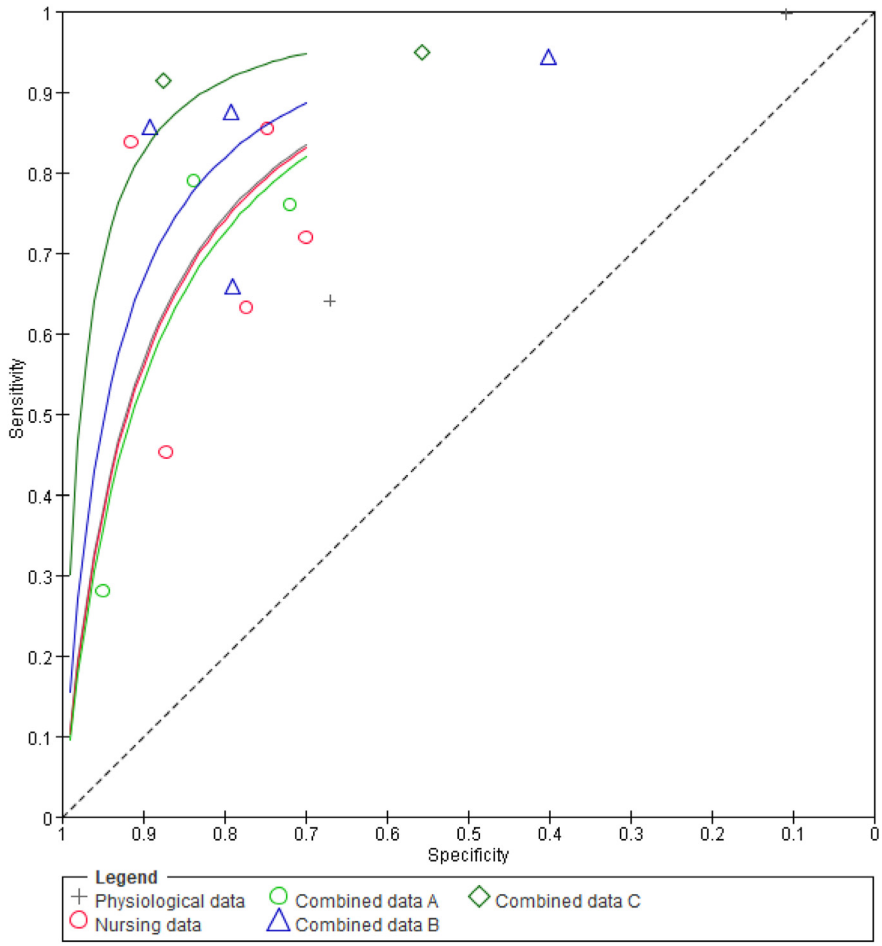

Fig. 3. Summary of receiver operating characteristic curves for diagnostic test accuracy.

context rather than physiological data alone. Future studies are needed to apply nursing data used in various clinical settings as predictors in the development of prediction models and evaluating their predictive performance.

Funding

This work was supported by the 2021 Faculty-Student Research Fund, granted from the Mo-Im Kim Nursing Research Institute, College of Nursing at Yonsei University. Mihui Kim received a scholarship from Brain Korea 21 FOUR Project funded by the National Research Foundation (NRF) of Korea, Yonsei University College of Nursing.

CRediT authorship contribution statement

Mihui Kim: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Sangwoo Park:** Conceptualization, Data curation, Investigation, Writing – original draft. **Changhwan Kim:** Formal analysis, Investigation, Validation, Writing – original draft. **Mona Choi:** Conceptualization, Formal analysis, Methodology, Supervision, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijnurstu.2022.104411.

References

Abhyankar, S., Demner-Fushman, D., Callaghan, F.M., McDonald, C.J., 2014. Combining structured and unstructured data to identify a cohort of ICU patients who received dialysis. J. Am. Med. Inform. Assoc. 21, 801–807. https://doi.org/10.1136/amiajnl-2013-001915.

Calvert, J., Mao, Q., Hoffman, J.L., Jay, M., Desautels, T., Mohamadlou, H., Chettipally, U., Das, R., 2016. Using electronic health record collected clinical variables to predict medical intensive care unit mortality. Ann. Med. Surg. 11, 52–57. https://doi.org/10.1016/j.amsu.2016.09.002.

Campbell, J.M., Klugar, M., Ding, S., Carmody, D.P., Hakonsen, S.J., Jadotte, Y.T., White, S., Munn, Z., 2015. Diagnostic test accuracy: methods for systematic review and meta-analysis. Int. J. Evid. Based Healthc. 13, 154–162. https://doi.org/10.1097/XEB.0000000000000061.

Capan, M., Wu, P., Campbell, M., Mascioli, S., Jackson, E.V., 2017. Using electronic health records and nursing assessment to redesign clinical early recognition systems. Health Syst. 6, 112–121. https://doi.org/10.1057/hs.2015.19.

Chan, P.S., Jain, R., Nallmothu, B.K., Berg, R.A., Sasson, C., 2010. Rapid response teams: a systematic review and meta-analysis. Arch. Intern. Med. 170, 18–26. https://doi.org/10.1001/archinternmed.2009.424.

Churpek, M.M., Yuen, T.C., Winslow, C., Meltzer, D.O., Kattan, M.W., Edelson, D.P., 2016. Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards. Crit. Care Med. 44, 368–374. https://doi.org/10.1097/CCM.0000000000001571.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. J. Br. Surg. 102, 148–158. https://doi.org/10.1002/bjs.9736.

Collins, S.A., Cato, K., Albers, D., Scott, K., Stetson, P.D., Bakken, S., Vawdrey, D.K., 2013. Relationship between nursing documentation and Patients' mortality. Am. J. Crit. Care 22, 306–313. https://doi.org/10.4037/ajcc2013426.

Collins, S.A., Vawdrey, D.K., 2012. “Reading between the lines” of flowsheet data: nurses' optional documentation associated with cardiac arrest outcomes. Appl. Nurs. Res. 25, 251. https://doi.org/10.1016/j.apnr.2011.06.002.

De Georgia, M.A., Kaffashi, F., Jacono, F.J., Loparo, K.A., 2015. Information technology in critical care: review of monitoring and data acquisition systems for patient care and research. Sci. World J. 2015, 727694. https://doi.org/10.1155/2015/727694.

De Silva, K., Mathews, N., Teede, H., Forbes, A., Jonsson, 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. Comput. Biol. Med. 132, 104305. https://doi.org/10.1016/j.compbiomed.2021.104305.

Douw, G., Schoonhoven, L., Holwerda, T., Huisman-de Waal, G., van Zanten, A.R., van Achterberg, T., van der Hoeven, J.G., 2015. Nurses' worry or concern and early recognition of deteriorating patients on general wards in acute care hospitals: a systematic review. Crit. Care 19, 230. https://doi.org/10.1186/s13054-015-0950-5.

Dykes, P.C., Kim, H.-E., Goldsmith, D.M., Choi, J., Esumi, K., Goldberg, H.S., 2009. The adequacy of ICNP version 1.0 as a representational model for electronic nursing assessment documentation. J. Am. Med. Inform. Assoc. 16, 238–246. https://doi.org/10.1197/jamia.M2956.

El-Rashidy, N., El-Sappagh, S., Abuhmed, T., Abdelrazek, S., El-Bakry, H.M., 2020. Intensive care unit mortality prediction: an improved patient-specific stacking ensemble model. IEEE Access 8, 133541–133564. https://doi.org/10.1109/Access.2020.3010556.

Fu, L.H., Knaplund, C., Cato, K., Perotte, A., Kang, M.J., Dykes, P.C., Albers, D., Collins Rossetti, S., 2021. Utilizing timestamps of longitudinal electronic health record data to classify clinical deterioration events. J. Am. Med. Inform. Assoc. 28, 1955–1963. https://doi.org/10.1093/jamia/ocab111.

Ghassemi, M., Naumann, T., Doshi-Velez, F., Brimmer, N., Joshi, R., Rumshisky, A., Szolovits, P., 2014. Unfolding physiological state: mortality modelling in intensive care units. Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 75–84.

Hashir, M., Sawhney, R., 2020. Towards unstructured mortality prediction with free-text clinical notes. J. Biomed. Inform. 108, 103489. https://doi.org/10.1016/j.jbi.2020.103489.

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. J. Am. Med. Inform. Assoc. 28, 1660–1666. https://doi.org/10.1093/jamia/ocab051.

Jeddah, D., Chen, O., Lipsky, A.M., Forgacs, A., Celniker, G., Lilly, C.M., Pessach, I.M., 2021. Validation of an automatic tagging system for identifying respiratory and hemodynamic deterioration events in the intensive care unit. Healthc. Inform. Res. 27, 241–248. https://doi.org/10.4258/hir.2021.27.3.241.

Jeong, S., 2018. Scoring Systems for the Patients of intensive care unit. Acute Crit. Care 33, 102–104. https://doi.org/10.4266/acc.2018.00185.

Jo, Y., Loghmanpour, N., Rosé, C.P., 2015. Time series analysis of nursing notes for mortality prediction via a state transition topic model. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Association for Computing Machinery, Melbourne, Australia, pp. 1171–1180.

Jones, D.A., DeVita, M.A., Bellomo, R., 2011. Rapid-response teams. N. Engl. J. Med. 365, 139–146. https://doi.org/10.1056/NEJMra0910926.

Kang, M.J., Dykes, P.C., Korach, T.Z., Zhou, L., Schnock, K.O., Thate, J., Whalen, K., Jia, H.M., Schwartz, J., Garcia, J.P., Knaplund, C., Cato, K.D., Rossetti, S.C., 2020. Identifying nurses' concern concepts about patient deterioration using a standard nursing terminology. Int. J. Med. Inform. 133, 104016. https://doi.org/10.1016/j.ijmedinf.2019.104016.

Khadanga, S., Aggarwal, K., Joty, S., Srivastava, J., 2019. Using Clinical Notes With Time Series Data for ICU Management. arXiv preprint arXiv:1909.09702.

Khine, A.H., Wettayaprasit, W., Duangsuwan, J., 2019. Ensemble CNN and MLP with nurse notes for intensive care unit mortality. JCSSE 2019 - 16th International Joint Conference on Computer Science and Software Engineering: Knowledge Evolution Towards Singularity of Man-Machine Intelligence, pp. 236–241.

Korach, Z.T., Cato, K.D., Collins, S.A., Kang, M.J., Knaplund, C., Dykes, P.C., Wang, L.Q., Schnock, K.O., Garcia, J.P., Jia, H.M., Chang, F., Schwartz, J.M., Zhou, L., 2019. Unsupervised machine learning of topics documented by nurses about hospitalized patients prior to a rapid-response event. Appl. Clin. Inform. 10, 952–963. https://doi.org/10.1055/s-0039-3401814.

Korach, Z.T., Yang, J., Rossetti, S.C., Cato, K.D., Kang, M.J., Knaplund, C., Schnock, K.O., Garcia, J.P., Jia, H., Schwartz, J.M., Zhou, L., 2020. Mining clinical phrases from nursing notes to discover risk factors of patient deterioration. Int. J. Med. Inform. 135, 104053. https://doi.org/10.1016/j.ijmedinf.2019.104053.

Kumar, V., Bajpai, R., Roy, R.B., 2021. Clinical notes mining for post discharge mortality prediction. IETE Tech. Rev. https://doi.org/10.1080/02564602.2021.1936224.

Lehman, L.-W., Saeed, M., Long, W., Lee, J., Mark, R., 2012. Risk stratification of ICU patients using topic models inferred from unstructured progress notes. AMIA Annual Symposium Proceedings. American Medical Informatics Association, p. 505.

Macaskill, P., Takwoingi, Y., Deeks, J.J., Gatsonis, C., 2022. Chapter 10: understaning meta-analysis Draft version (13 May 2022). In: JJ, D., PM, B., MM, L., Takwoingi (Eds.), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 2. Cochrane.

Marafino, B.J., Boscardin, W.J., Dudley, R.A., 2015. Efficient and sparse feature selection for biomedical text classification via the elastic net: application to ICU risk stratification from nursing notes. J. Biomed. Inform. 54, 114–120. https://doi.org/10.1016/j.jbi.2015.02.003.

Marafino, B.J., Park, M., Davies, J.M., Thombley, R., Luft, H.S., Sing, D.C., Kazi, D.S., DeJong, C., Boscardin, W.J., Dean, M.L., Dudley, R.A., 2018. Validation of prediction models for critical care outcomes using natural language processing of electronic health record data. JAMA Netw. Open 1, e185097. https://doi.org/10.1001/jamanetworkopen.2018.5097.

Moons, K.G., de Groot, J.A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., Reitsma, J.B., Collins, G.S., 2014. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med. 11, e1001744. https://doi.org/10.1371/journal.pmed.1001744.

Odell, M., Victor, C., Oliver, D., 2009. Nurses' role in detecting deterioration in ward patients: systematic literature review. J. Adv. Nurs. 65, 1992–2006. https://doi.org/10.1111/j.1365-2648.2009.05109.x.

Parreco, J., Hidalgo, A., Kozol, R., Namias, N., Rattan, R., 2018. Predicting mortality in the surgical intensive care unit using artificial intelligence and natural language processing of physician documentation. Am. Surg. 84, 1190–1194. https://doi.org/10.1177/000313481808400736.

Reitsma, J.B., Glas, A.S., Rutjes, A.W., Scholten, R.J., Bossuyt, P.M., Zwinderman, A.H., 2005. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. J. Clin. Epidemiol. 58, 982–990. https://doi.org/10.1016/j.jclinepi.2005.02.022.

Rocker, G., Cook, D., Sjokvist, P., Weaver, B., Finfer, S., McDonald, E., Marshall, J., Kirby, A., Levy, M., Dodek, P., Heyland, D., Guyatt, G., Level of Care Study, I., Canadian Critical Care Trials, G., 2004. Clinician predictions of intensive care unit mortality. Crit. Care Med. 32, 1149–1154. https://doi.org/10.1097/01.ccm.0000126402.51524.52.

Rojas, J.C., Carey, K.A., Edelson, D.P., Venable, L.R., Howell, M.D., Churpek, M.M., 2018. Predicting intensive care unit readmission with machine learning using electronic health record data. Ann. Am. Thorac. Soc. 15, 846–853. https://doi.org/10.1513/AnnalsATS.201710-787OC.

Schnock, K.O., Kang, M.J., Rossetti, S.C., Garcia, J., Lowenthal, G., Knaplund, C., Chang, F., Albers, D., Korach, T.Z., Zhou, L., Klann, J.G., Cato, K., Bates, D.W., Dykes, P.C., 2021. Identifying nursing documentation patterns associated with patient deterioration and recovery from deterioration in critical and acute care settings. Int. J. Med. Inform. 153, 104525. https://doi.org/10.1016/j.ijmedinf.2021.104525.

Steyerberg, E.W., 2019. Clinical Prediction Models : A Practical Approach to Development, Validation, and Updating. Springer.

Tran, N., Lee, J., 2018. Using multiple sentiment dimensions of nursing notes to predict mortality in the intensive care unit. 2018 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2018, pp. 283–286.

Waudby-Smith, I.E.R., Tran, N., Dubin, J.A., Lee, J., 2018. Sentiment in nursing notes as an indicator of out-of-hospital mortality in intensive care patients. PLoS One 13, e0198687. https://doi.org/10.1371/journal.pone.0198687.

Weissman, G.E., Hubbard, R.A., Ungar, L.H., Harhay, M.O., Greene, C.S., Himes, B.E., Halpern, S.D., 2018. Inclusion of unstructured clinical text improves early prediction of death or prolonged ICU stay. Crit. Care Med. 46, 1125–1132. https://doi.org/10.1097/CCM.0000000000003148.

Wellner, B., Grand, J., Canzone, E., Coarr, M., Brady, P.W., Simmons, J., Kirkendall, E., Dean, N., Kleinman, M., Sylvester, P., 2017. Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements. JMIR Med. Inform. 5, e45. https://doi.org/10.2196/medinform.8680.

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., Groupdagger, P., 2019. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann. Intern. Med. 170, 51–58. https://doi.org/10.7326/M18-1376.

Yu, J.Y., Jeong, G.Y., Jeong, O.S., Chang, D.K., Cha, W.C., 2020. Machine learning and initial nursing assessment-based triage system for emergency department. Healthc. Inform. Res. 26, 13–19. https://doi.org/10.4258/hir.2020.26.1.13.

Zalewski, A., Long, W., Johnson, A.E.W., Mark, R.G., Lehman, L.W.H., 2017. Estimating patient's health state using latent structure inferred from clinical time series and text. 2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017, pp. 449–452. https://doi.org/10.1109/BHI.2017.7897302.