



Published in final edited form as:

Informatics (MDPI). 2021 March ; 8(1): . doi:10.3390/informatics8010016.

Application of Machine Learning in Intensive Care Unit (ICU) Settings Using MIMIC Dataset: Systematic Review

Mahanazuddin Syed^{1,*†}, Shorabuddin Syed^{1,†}, Kevin Sexton^{1,2,3}, Hafsa Bareen Syeda¹, Maryam Garza¹, Meredith Zozus⁴, Farhanuddin Syed⁵, Salma Begum⁶, Abdullah Usama Syed⁷, Joseph Sanford^{1,8}, Fred Prior¹

¹Department of Biomedical Informatics, University of Arkansas for Medical Sciences (UAMS), Little Rock, Arkansas 72205, USA.

²Department of Surgery, University of Arkansas for Medical Sciences (UAMS), Little Rock, Arkansas 72205, USA.

³Department of Health Policy and Management, University of Arkansas for Medical Sciences (UAMS), Little Rock, Arkansas 72205, USA.

⁴Department of Population Health Sciences, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229, USA.

⁵Shadan Institute of Medical Sciences, College of Medicine, Hyderabad, Telangana 500086, India.

⁶Department of Information Technology, University of Arkansas for Medical Sciences (UAMS), Little Rock, Arkansas 72205, USA.

⁷Department of Information Science, University of Arkansas at Little Rock (UALR), Little Rock, Arkansas 72205, USA.

⁸Department of Anesthesiology, University of Arkansas for Medical Sciences (UAMS), Little Rock, Arkansas 72205, USA.

Abstract

Modern Intensive Care Units (ICUs) provide continuous monitoring of critically ill patients susceptible to many complications affecting morbidity and mortality. ICU settings require a high staff-to-patient ratio and generates a sheer volume of data. For clinicians, the real-time interpretation of data and decision-making is a challenging task. Machine Learning (ML) techniques in ICUs are making headway in the early detection of high-risk events due to increased processing power and freely available datasets such as the Medical Information Mart for Intensive

*Correspondence: msyed2@uams.edu; Tel.: +1 (501) 686-8447.

Author Contributions: Conceptualization, M.S., F.P., and S.S.; methodology, M.S. and S.S.; formal analysis, all authors; investigation, all authors; resources, K.S.; data curation, all authors; writing—original draft preparation, M.S. and S.S.; writing—review and editing, all authors; validation, all authors; visualization, all authors; supervision, J.S.; project administration, A.U.S. and S.B.; funding acquisition, None. All authors have read and agreed to the published version of the manuscript.

[†]These authors contributed equally to this work.

Supplementary Materials: Table S1: Summary of 61 studies qualified for quantitative descriptive analysis. The following are available online at <https://www.mdpi.com/2227-9709/8/1/16/s1>.

Conflicts of Interest: The authors declare no conflict of interest.

Care (MIMIC). We conducted a systematic literature review to evaluate the effectiveness of applying ML in the ICU settings using the MIMIC dataset. A total of 322 articles were reviewed and a quantitative descriptive analysis was performed on 61 qualified articles that applied ML techniques in ICU settings using MIMIC data. We assembled the qualified articles to provide insights into the areas of application, clinical variables used, and treatment outcomes that can pave the way for further adoption of this promising technology and possible use in routine clinical decision-making. The lessons learned from our review can provide guidance to researchers on application of ML techniques to increase their rate of adoption in healthcare.

Keywords

intensive care unit; critical care; MIMIC; machine learning; deep learning; systematic review; sepsis; acute kidney injury

1. Introduction

Artificial intelligence (AI) encompasses a broad-spectrum of technologies that aim to imitate cognitive functions and intelligent behavior of humans [1]. Machine Learning (ML) is a subfield of AI that focuses on algorithms that allow computers to define a model for complex relationships or patterns from empirical data without being explicitly programmed [2]. ML, powered by increasing availability of healthcare data, is being used in a variety of clinical applications ranging from diagnosis to outcome prediction [1,3]. The predictive power of ML improves as the number of samples available for learning increases [4,5].

ML algorithms can be supervised or unsupervised based on the type of learning rule employed. In supervised learning, an algorithm is trained using well-labeled data. Thereafter, the machine predicts on unseen data by applying knowledge gained from the training data [6]. Most adopted supervised ML models are Random Forest (RF), Support Vector Machines (SVM), and Decision Tree algorithms [6]. In unsupervised learning, there is no ground truth labeling required. Instead, the machine learns from the inherent structure of the unlabeled data [7]. Either type of ML is an iterative process in which the algorithm tries to find the optimal combination of both model variables and variable weights with the goal of minimizing error in the predicted outcome [5,6]. If the algorithm performs with a reasonably low error rate, it can be employed for making predictions where outputs are not known. However, while developing a ML model, an optimal bias-variance tradeoff should be selected to optimize prediction error rate [8]. Improper selection of bias and variance results in two problems: (1) underfitting and (2) overfitting [9]. Finding the “sweet spot” between the bias and variance is crucial to avoid both underfitting and overfitting [8,10].

Deep learning (DL), a subcategory of machine learning, achieves great power and flexibility compared to conventional ML models by drawing inspiration from biological neural networks to solve a wide variety of complex tasks, including the classification of medical imaging and Natural Language Processing (NLP) [10–14]. Most widely used DL models are variants of Artificial Neural Network (ANN) and Multi-Layer Perceptron (MLP). In general, ML models are data driven and they rely on a deep understanding of the system for prediction, thereby, empowering users to make informed decision.

To provide better patient care and facilitate translational research, healthcare institutions are increasingly leveraging clinical data captured from Electronic Health Records (EHR) systems [15]. Of these systems, the Intensive Care Unit (ICU) generates an immense volume of data, and requires a high staff-to-patient ratio [16,17]. To avoid adverse events and prolonged ICU stays, early detection and intervention on patients vulnerable to complications is crucial; for these reasons, the ML literature is increasingly using ICU patient data for clinical event prediction and secondary usage, such as sepsis and septic shock [18]. ML techniques in ICUs are making headway in the early detection of high-risk events due to increased processing power and freely available datasets such as the Medical Information Mart for Intensive Care (MIMIC) [19]. The data available in the MIMIC database includes highly structured data from time-stamped, nurse-verified physiological measurements made at the bedside, as well as unstructured data, including free-text interpretations of imaging studies provided by the radiology department [13].

The primary aim of this study was to conduct a systematic literature review on the effectiveness of applying ML technologies using MIMIC dataset. Specifically, we summarized the clinical area of application, disease type, clinical variables, data type, ML methodology, scientific findings, and challenges experienced across the existing ICU-ML literature.

2. Methods

This systematic literature review followed the guidelines of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework for preparation and reporting [20].

2.1. Eligibility Criteria

This study focused on peer-reviewed publications that applied ML techniques to analyze retrospective ICU data available from publicly available MIMIC dataset, which includes discrete structured clinical data, physiological waveforms data, free text documents, and radiology imaging reports.

2.2. Data Sources and Search Strategy

Three search engines were used: PubMed, Web of Science, and the CINAHL. We restricted our search to research articles published in English and in peer-reviewed journals or conferences available from the inception of each database through 30 October 2020. The search syntax was built with the guidance of a professional librarian and included search terms: “Machine Learning”, “Deep Learning”, “Artificial Intelligence”, “Neural Network”, “Supervised Learning”, “Support Vector Machine”, “SVM”, “Intensive Care Unit”, “ICU”, “Critical Care”, “Intensive Care”, “MIMIC”, “MIMIC-II”, “Medical Information Mart for Intensive Care”, “Beth Israel Deaconess Medical Center”. Figure 1 illustrates the process of identifying eligible publications.

2.3. Study Selection

Following the systematic search process, 322 publications were retrieved. Of that, 113 duplicate publications were removed, leaving 209 potentially relevant articles for the title and abstract screening. Two teams (HB, SS and MS, SB) screened these articles independently, leading to the removal of another 89 publications, and 120 publications were retained for a full-text assessment. These were assessed for eligibility, resulting in 61 total publications that were included in the final analysis. Disagreements were resolved by an independent review by third person (AS).

2.4. Data Collection and Analysis

A quantitative descriptive analysis was performed on the qualified studies that had applied ML technique in the ICU settings using MIMIC database. Data elements extracted for this analysis included: (a) type of disease (b) type of data (c) clinical area of application (d) ML techniques and (e) year of publication. After the extraction and analysis, we summarized and reported the findings in tables in accordance with the aim of the study.

3. Results

The search strategy yielded a total of 322 articles, which were published and made available as of 30 October 2020. Of which, 61 publications were selected for further analyses. These publications were categorized into seven themes based on the effectiveness of applying ML techniques in various ICU settings. These themes are identified based on ML algorithms to predict, monitor, and improve patient outcomes. Descriptions of each theme and related publications are listed in Table 1.

Majority of the studies in our review focused on predicting mortality (21 studies) and followed by risk stratification (10 studies). Multiple studies focused on predicting the onset date of specific diseases such as sepsis and septic shock (eight studies), cardiac diseases (seven studies), and acute kidney injury (AKI) (six studies).

Details of each study including clinical applications, ML models and clinical variables used, sample size, and model performance of the qualified studies is provided in the Supplementary Table S1.

In our review, both traditional ML and DL models were used in the classification tasks. Based on best-performing models reported by the studies, traditional ML was used in 36 studies and DL was used in 25 studies, respectively. In traditional ML algorithms, SVM and RF were most commonly used whereas in DL, Long Short-Term Memory (LSTM) was employed. Majority of the studies have used discrete clinical variables, and eight publications used both discrete and unstructured data such as discharge summaries, nursing notes, radiology reports, etc. as an input.

Thirty-six studies applied imputation methods, and the remaining 25 either completely removed the records with missing data or did not mention it. Fifty-two studies used cross-validation to evaluate model performance. Forty-eight studies used feature identification to improve accuracy.

4. Discussion

The aim of this systematic review was to provide an up-to-date and holistic view of the current ML applications in ICU settings using MIMIC data in the attempt to predict clinical outcomes. Our review revealed ML application was widely adopted in areas such as mortality, risk stratification, readmission, and infectious disease in critically ill patients using retrospective data. This review may be used to provide insights for choosing key variables and best performing models for further research.

The application of ML techniques within the ICU domain is rapidly expanding with improvement of modern computing, which has enabled the analysis of huge volumes of complex and diverse data [1]. ML expands on existing statistical techniques, utilizing methods that are not based on a priori assumptions about the distribution of the data, but deriving insights directly from the data [80,81].

With ICUs being complex settings that generate a variety of time-sensitive data, more and more ML-based studies have begun tapping the openly available, large tertiary care hospital data (MIMIC). Our screening resulted in 61 publications that utilized MIMIC data to train and test ML models enabling reproducibility. The majority of these publications focused on predicting mortality, sepsis, AKI, and readmissions.

4.1. Mortality Prediction

Mortality prediction for ICU patients is critical and crucial for assessing severity of illness and adjudicating the value of treatments, and timely interventions. ML algorithms developed for predicting mortality in ICUs focused mainly on in-hospital mortality and 30 days mortality at discharge. Studies by Marafino et al. [22], Pirracchio et al. [23], Hoogendoorn et al. [24], Awad et al. [26], Davoodi et al. [29], and Weissman et al. [31] predicted in-hospital mortality, whereas Du et al. [25] predicted 28 days mortality at discharge, and Zahid et al. [30] predicted both 30 days and in-hospital mortality. Most studies focusing on predicting in-hospital mortality looked at mortality after 24 h of ICU admission. However, one in particular, Awad et al. [26], predicted mortality within 6 h of admission. Marafino et al. [22] predicted mortality using only nursing notes from the first 24 h of ICU admission, whereas Weissman et al. [31] improved mortality prediction by combining structured and unstructured data generated within the first 48 h of the ICU stay. Davoodi et al. [29] and Hoogendoorn et al. [24] predicted after 24 h and within a median of 72 h, respectively. Studies by Tang et al. [33], Caicedo-Torres et al. [36], Sha et al. [38], and Zhang et al. [41] predicted in-hospital mortality irrespective of the admission or discharge time.

For mortality prediction, all of the studies used three main categories of clinical variables: (1) demographics, (2) vital signs, and (3) laboratory test variables. In addition to the most commonly used data elements, other clinical information such as medications, intake/output variables, risk scores, and comorbidities were also utilized. Weissman et al. [31] and Zhang et al. [41] used clinical variables from both structured and unstructured data types for mortality prediction.

Multiple studies predicted mortality on disease-specific patient cohorts. Celi et al. [21] and Lin et al. [35] predicted in-hospital mortality on AKI patients. Lin et al. [35] predicted mortality based on five important variables (urine output, systolic blood pressure, age, serum bicarbonate level, and heart rate). In addition, the study by Lin et al. [35] also revealed that the effect of kidney injury markers, such as cystatin C and neutrophil gelatinase-associated lipocalin on subclinical injury, had not yet been analyzed, which can provide AKI prognostic information. This is due to lack of data availability in MIMIC. Garcia-Gallo et al. [37] and Kong et al. [40] predicted mortality on sepsis patients, and specifically, Garcia-Gallo et al. [37] identified patients that are on 1-year mortality trajectory. Anand et al. [34] claimed that the risk of mortality in diabetic patients could be better predicted using a combination of limited variables: HbA1c, mean glucose during stay, diagnoses upon admission, age, and type of admission. To compute the “diagnosis upon admission” variable, the study utilized Charlson Comorbidity Index, Elixhauser Comorbidity Index, and Diabetic Severity Index. The authors further claimed that combining diabetic-specific metrics and using the fewest possible variables would result in better mortality risk prediction in diabetic patients.

In our review, studies have used both traditional ML (10 studies) and DL methods (11 studies) to predict mortality. In traditional ML techniques, Random Forest, Decision Tree, and Logistic Regression were the most commonly used algorithms. However, recent studies by Caicedo-Torres et al. [36], Du et al. [25], and Zahid et al. [30] have used DL methods for mortality prediction with a promising accuracy ranging from 0.86–0.87 as reported in the Supplementary Table S1. Traditional ML models can be easily interpretable when compared to DL models that have many levels of features and hidden layers to predict outcomes. Understanding the features that contribute towards the prediction plays an important role for clinical decision-making [82,83]. For example, one of the most cited studies by Pirracchio et al. [23] developed a mortality prediction algorithm (Super Learner) using a combination of traditional ML models; the results of which were easily interpretable by clinical researchers. In general, DL techniques are employed to improve prediction accuracy by training on large volumes of data [12]. Zahid et al. [30] developed a DL model (Self-Normalizing Neural Network (SNN)) that performed marginally better than the Pirracchio et al. [23] mortality prediction rate (Area Under the Receiver Operating Characteristic curve (AUROC) of SNN: 0.86 and Super Learner: 0.85). However, interpreting the results of DL models is challenging because of multiple hidden layers and they are often treated as black-box models. To address this limitation, Caicedo-Torres et al. [36] and Sha et al. [38] demonstrated the interpretability of the model in visualizations that will allow clinicians to make informed decisions.

4.2. Acute Kidney Injury (AKI) Prediction

AKI is one of the common complications among adult patients in the intensive care unit (ICU). AKI patients are at risk for adverse clinical outcomes such as prolonged ICU and hospitalization stays, high morbidity, and mortality. Application of ML in AKI care has been mainly focused on early prediction of an AKI event and risk stratification. In our review, studies employed traditional ML techniques to predict AKI events and XGBoost was the most commonly used algorithm.

Using the MIMIC dataset, Zimmerman et al. [68], Sun et al. [69], and Li et al. [84] predicted AKI after 24 h of ICU admission. Sun et al. [69] and Li et al. [84] used clinical unstructured notes generated during the first 24 h of ICU stay, whereas Zimmerman et al. [68] used structured clinical variables for prediction. The AUROC of predicting AKI within the first 24 h in Sun et al. [69], Zimmerman et al. [68], and Li et al. [84] was reported as 0.83, 0.783, and 0.779, respectively. Additional details on the type of clinical variables, sample size, and ML model are listed in Supplementary Table S1.

To define and classify AKI, three standard guidelines have been published and used in clinical settings: (1) Risk, Injury, Failure, Loss, End-Stage (RIFLE), (2) Acute Kidney Injury Network (AKIN), and (3) Kidney Disease: Improving Global Outcomes (KDIGO). In our results, most studies used KDIGO guidelines to create ground truth labels and is based on serum creatinine (SCr) and urine output. The SCr is one of the important predictor variables in AKI; however, it is a late marker of AKI, which delays diagnosis and care [85]. In clinical settings, it is highly desirable to early predict the AKI event for better intervention strategies. To address the aforementioned clinical need, Zimmerman et al. [68] predicted SCr values for 48 and 72 h based on 24 h SCr values and other clinical variables. Li et al. [84] extracted key features from clinical notes, such as diuretic and insulin medications using NLP instead of completely depending on SCr. Even though urine output is one of the defined metrics of AKI, Zimmerman et al. [68] reported it as not a significant predictor [86]. Further investigation should focus on the effect of urine output on predicting AKI and its impact.

4.3. Sepsis and Septic Shock

Sepsis is one of the leading causes of death among ICU patients and hospitalized patients overall [87]. As sepsis progresses, patients from pre-shock state are highly likely to develop septic shock. Early recognition of sepsis and initiation of treatment will reduce mortality and morbidity [88]. In our review, eight studies applied ML techniques to predict sepsis or septic shock events. Of these, four applied traditional ML algorithms and the other four used DL methods. XGBoost and LSTM were the most commonly used algorithms, of which the details of variables, sample size, and model performances are provided in the Supplementary Table S1.

The Scherpf et al. [55] model predicted sepsis 3 h prior to the onset with an AUROC of 0.81. The results of our review also reveal that most studies focused on early predicting of pre-shock state using hemodynamic measurements. The common variables used in ML models are arterial pressure, heart rate, labs, risk scores including Glasgow Coma Scores (GCS) and Sequential Organ Failure Assessment (SOFA) scores, and respiratory rate.

For predicting pre-shock state, Liu et al. [54] and Kam et al. [53] used a combination of these variables along with lab findings with the Area under the Curve (AUC) performance reported as 0.93 and 0.929, respectively. One of the interesting findings of the Liu et al. [54] study was that serum lactate was the primary predictor variable indicating a patient's risk level of entering septic shock, and is used as a biomarker for sepsis patient risk stratification. The study also reported, "A patient with serum lactate concentration one standard deviation above the population mean is approximately five times as likely to transition into shock than a patient with average serum lactate concentration" [54]. The hemodynamic measurements

can be derived from waveform data or can be extracted as discrete data elements from EHR. Ghosh et al. [52] used three waveforms: mean arterial pressure, heart rate, and respiratory rate to derive hemodynamic predictor variables, whereas Liu et al. [54] and Kam et al. [53] used discrete measurements.

4.4. ICU Readmission

Intensive Care Units (ICU) provide care to critically ill patients, which is often costly and labor-intensive. Prolong ICU stays increase cost burden to both patients and hospitals. Early predicting unplanned readmissions may help in ICU resources allocation and improve patient health outcomes. Details of the studies qualified in this theme are listed in the Supplementary Table S1. Desautels et al. [76] identified patients who are likely to suffer unplanned ICU readmission: his model reported an AUROC of 0.71. Rojas et al. [78] and Lin et al. [79] focused on identifying patients that were re-admitted within 30 days of discharge. The best AUROC reported by Lin et al. [79] and Rojas et al. [78] is 0.791 and 0.78, respectively. The common predictor variables used in all three studies include: vital signs, demographics, comorbidities, and labs. Our findings revealed that there has been limited research done on predicting readmissions and the reported model AUROCs in literature are not promising (less than 80%) using MIMIC data.

4.5. ML Model Optimization

The performance of a given model heavily depended on data pre-processing, feature identification, and model validation. The missing data problem is arguably the most common issue encountered by machine learning practitioners when analyzing real-world healthcare data [89]. Researchers in general choose to address the missing data by either imputing or removing the observations [89]. The imputation can be done using simple-to-complex techniques: for example, in the study done by Lin et al. [35], missing observations were imputed using the mean value of the variable, whereas Davoodi et al. [29] and Zhang et al. [67] used sophisticated imputation techniques, Gaussian and Multivariate Imputation by Chained Equation (MICE), respectively. Substituting observed values with estimated observations introduces bias that may distort the data distribution or introduce spurious associations influencing model accuracy. To minimize this, imputation methods should be carefully selected, especially for prospective data. Imputation methodology depends on aim of the study, importance of data elements, percentage of missing data, and ML model used.

Feature importance technique is often employed to identify the highest ranked features. ML models with only important features improve the accuracy and computing time [90]. Cross-validation (CV) of a ML algorithm is vital to estimate a model's predictive power and generalized performance on the unseen data [91,92]. K-fold CV is often used to reduce the pessimistic bias by using more training data to teach the model. Our analysis found 52 studies used various validation techniques. Five-fold and 10-fold CV were the most common validation method used.

This review has some inherent limitations. First, there is the possibility of studies missed due to the search methodology. Second, we removed sixteen publications where full text was not available, and this may have introduced bias. Finally, a comparison of ML model

performance was not possible in the quantitative analysis even though the studies used the MIMIC dataset for training and validating ML models. This is due to the fact that ML performance is dependent on the data elements selected for prediction, model parameters used, and size of the dataset.

4.6. Key Points and Recommendations

The aim of the study was to perform a comprehensive literature review on ML application in ICU settings using MIMIC dataset. The key points of our review and recommendations for future research provided therein are enlisted below.

- Recent proliferation of publicly available MIMIC datasets allowed researchers to provide effective ML-based solutions in an attempt to solve complex healthcare problems. However, reproducibility of ML models is lacking due to inconsistent reporting of clinical variables selected, data pre-processing, and model specifications during the development. Future studies should follow standard reporting guidelines to accurately disclose model specifications.
- Significant work has been done in predicting mortality within 6 to 72 h of hospital admission on retrospective data. However, prospective implementation is lacking. To adapt to dynamics of clinical events, we recommend exposing these models to prospective trials before moving it to routine clinical practice.
- ML model performance heavily depends on clinical variables utilized. We identified and summarized the variables used by different model across the themes. Future studies should focus on performing a detailed analysis of these variables for improved performance.
- Unstructured clinical notes have valuable and time-sensitive information critical for decision-making. Eight studies in our review tapped into clinical notes to mine important information. However, recent advancements in NLP techniques like Bidirectional Encoder Representations from Transformers (BERT) and Embeddings from Language Models (ELMo) have not been explored.
- Interpretable ML models allow clinicians to understand and improve model performance. However, only two studies have resorted to visualization-based interpretations in the review.

5. Conclusions

ML is gaining traction in the ICU setting. This systematical review aimed to assemble the current ICU literature that utilized ML methods to provide insights into the areas of application and treatment outcomes using MIMIC dataset. Our work can pave the way for further adoption and overcome hurdles in employing ML technology in clinical care. This study identified the most important clinical variables used in the design and development of ML models for predicting mortality and infectious disease in critically ill patients, which can provide insights for choosing key variables for further research. We also discovered that predicting disease classification and treatment outcomes using supervised and unsupervised ML is possible with high predictive value on retrospective data. Prospective validation is still

lacking, possibly due to the limitations with implementation and real-time disparate data processing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

This study was supported in part by the Translational Research Institute (TRI), grant UL1 TR003107 received from the National Center for Advancing Translational Sciences of the National Institutes of Health (NIH). The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol* 2017;2(4):230–43. [PubMed: 29507784]
2. Martins J, Magalhães C, Rocha M, Osório NS. Machine Learning-Enhanced T Cell Neopeptide Discovery for Immunotherapy Design. *Cancer informatics* 2019;18:1176935119852081-. [PubMed: 31205413]
3. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. *Database : the journal of biological databases and curation* 2020;2020:baaa010. [PubMed: 32185396]
4. Figueroa RL, Zeng-Treitler Q, Kandula S, Ngo LH. Predicting sample size required for classification performance. *BMC medical informatics and decision making* 2012;12:8-. [PubMed: 22336388]
5. Senders JT, Arnaut O, Karhade AV, Dasenbrock HH, Gormley WB, Broekman ML, et al. Natural and Artificial Intelligence in Neurosurgery: A Systematic Review. *Neurosurgery* 2018;83(2):181–92. [PubMed: 28945910]
6. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making* 2019;19(1):281-. [PubMed: 31864346]
7. Cleret de Langavant L, Bayen E, Yaffe K. Unsupervised Machine Learning to Identify High Likelihood of Dementia in Population-Based Surveys: Development and Validation Study. *Journal of medical Internet research* 2018;20(7):e10493–e. [PubMed: 29986849]
8. Belkin M, Hsu D, Ma S, Mandal S. Reconciling modern machine-learning practice and the classical bias-variance trade-off. *Proceedings of the National Academy of Sciences of the United States of America* 2019;116(32):15849–54. [PubMed: 31341078]
9. Borstelmann SM. Machine Learning Principles for Radiology Investigators. *Acad Radiol* 2020;27(1):13–25. [PubMed: 31818379]
10. Chartrand G, Cheng PM, Vorontsov E, Drozdal M, Turcotte S, Pal CJ, et al. Deep Learning: A Primer for Radiologists. *RadioGraphics* 2017;37(7):2113–31. [PubMed: 29131760]
11. Torres JJ, Varona P. Modeling Biological Neural Networks. In: Rozenberg G, Bäck T, Kok JN, editors. *Handbook of Natural Computing*. Berlin, Heidelberg:Springer Berlin Heidelberg; 2012. p.533–64.
12. Chen D, Liu S, Kingsbury P, Sohn S, Storlie CB, Habermann EB, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. *NPJ digital medicine* 2019;2:43-. [PubMed: 31304389]
13. Jang J, Ankit A, Kim J, Jang YJ, Kim HY, Kim JH, et al. A Unified Deep-Learning Model for Classifying the Cross-Country Skiing Techniques Using Wearable Gyroscope Sensors. *Sensors (Basel, Switzerland)* 2018;18(11):3819.
14. Giger ML. Machine Learning in Medical Imaging. *Journal of the American College of Radiology* 2018;15(3):512–20. [PubMed: 29398494]

15. Downing NL, Rolnick J, Poole SF, Hall E, Wessels AJ, Heidenreich P, et al. Electronic health record-based clinical decision support alert for severe sepsis: a randomised evaluation. *BMJ Quality & Safety* 2019;28(9):762.
16. Kizzier-Carnahan V, Artis KA, Mohan V, Gold JA. Frequency of Passive EHR Alerts in the ICU: Another Form of Alert Fatigue? *Journal of patient safety* 2019;15(3):246–50. [PubMed: 27331600]
17. Huddar V, Desiraju BK, Rajan V, Bhattacharya S, Roy S, Reddy CK. Predicting Complications in Critical Care Using Heterogeneous Clinical Data. *IEEE Access* 2016;4:7988–8001.
18. Giannini HM, Ginestra JC, Chivers C, Draugelis M, Hanish A, Schweickert WD, et al. A Machine Learning Algorithm to Predict Severe Sepsis and Septic Shock: Development, Implementation, and Impact on Clinical Practice. *Crit Care Med* 2019;47(11):1485–92. [PubMed: 31389839]
19. Johnson AEW, Pollard TJ, Shen L, Lehman L-WH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. *Scientific data* 2016;3:160035-. [PubMed: 27219127]
20. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 2009;339:b2700. [PubMed: 19622552]
21. Celi LA, Galvin S, Davidzon G, Lee J, Scott D, Mark R. A Database-driven Decision Support System: Customized Mortality Prediction. *J Pers Med* 2012;2(4):138–48. [PubMed: 23766893]
22. Marafino BJ, Boscardin WJ, Dudley RA. Efficient and sparse feature selection for biomedical text classification via the elastic net: Application to ICU risk stratification from nursing notes. *J Biomed Inform* 2015;54:114–20. [PubMed: 25700665]
23. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van der Laan MJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. *Lancet Respir Med* 2015;3(1):42–52. [PubMed: 25466337]
24. Hoogendoorn M, El Hassouni A, Mok K, Ghassemi M, Szolovits P. Prediction using patient comparison vs. modeling: a case study for mortality prediction. *Conf Proc IEEE Eng Med Biol Soc* 2016;2016:2464–7.
25. Hao D, Ghassemi MM, Mengling F. The effects of deep network topology on mortality prediction. *Conf Proc IEEE Eng Med Biol Soc* 2016;2016:2602–5.
26. Awad A, Bader-El-Den M, McNicholas J, Briggs J. Early hospital mortality prediction of intensive care unit patients using an ensemble learning approach. *Int J Med Inform* 2017;108:185–95. [PubMed: 29132626]
27. Meyer A, Zverinski D, Pfahringer B, Kempfert J, Kuehne T, Sundermann SH, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. *Lancet Respir Med* 2018;6(12):905–14. [PubMed: 30274956]
28. Purushotham S, Meng C, Che Z, Liu Y. Benchmarking deep learning models on large healthcare datasets. *J Biomed Inform* 2018;83:112–34. [PubMed: 29879470]
29. Davoodi R, Moradi MH. Mortality prediction in intensive care units (ICUs) using a deep rule-based fuzzy classifier. *J Biomed Inform* 2018;79:48–59. [PubMed: 29471111]
30. Zahid MAH, Lee J, editors. Mortality prediction with self normalizing neural networks in intensive care unit patients. *Proceedings of 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*; 2018 4–7 March 2018. p.226–9.
31. Weissman GE, Hubbard RA, Ungar LH, Harhay MO, Greene CS, Himes BE, et al. Inclusion of Unstructured Clinical Text Improves Early Prediction of Death or Prolonged ICU Stay. *Crit Care Med* 2018;46(7):1125–32. [PubMed: 29629986]
32. Jain SS, Sarkar IN, Stey PC, Anand RS, Biron DR, Chen ES. Using Demographic Factors and Comorbidities to Develop a Predictive Model for ICU Mortality in Patients with Acute Exacerbation COPD. *AMIA Annu Symp Proc* 2018;2018:1319–28. [PubMed: 30815176]
33. Tang F, Xiao C, Wang F, Zhou J. Predictive modeling in urgent care: a comparative study of machine learning approaches. *JAMIA Open* 2018;1(1):87–98. [PubMed: 31984321]
34. Anand RS, Stey P, Jain S, Biron DR, Bhatt H, Monteiro K, et al. Predicting Mortality in Diabetic ICU Patients Using Machine Learning and Severity Indices. *AMIA Jt Summits Transl Sci Proc* 2018;2017:310–9. [PubMed: 29888089]

35. Lin K, Hu Y, Kong G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *Int J Med Inform* 2019;125:55–61. [PubMed: 30914181]
36. Caicedo-Torres W, Gutierrez J. ISeeU: Visually interpretable deep learning for mortality prediction inside the ICU. *J Biomed Inform* 2019;98:103269. [PubMed: 31430550]
37. Garcia-Gallo JE, Fonseca-Ruiz NJ, Celi LA, Duitama-Munoz JF. A machine learning-based model for 1-year mortality prediction in patients admitted to an Intensive Care Unit with a diagnosis of sepsis. *Med Intensiva* 2018.
38. Sha Y, Wang MD. Interpretable Predictions of Clinical Outcomes with An Attention-based Recurrent Neural Network. *Acm bcb* 2017;2017:233–40. [PubMed: 32577628]
39. Ahmed FS, Ali L, Joseph BA, Ikram A, Ul Mustafa R, Bukhari SAC. A statistically rigorous deep neural network approach to predict mortality in trauma patients admitted to the intensive care unit. *J Trauma Acute Care Surg* 2020;89(4):736–42.
40. Kong G, Lin K, Hu Y. Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. *BMC Med Inform Decis Mak* 2020;20(1):251.
41. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. *BMC Med Inform Decis Mak* 2020;20(1):280.
42. Marafino BJ, Davies JM, Bardach NS, Dean ML, Dudley RA. N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. *J Am Med Inform Assoc* 2014;21(5):871–5. [PubMed: 24786209]
43. Dervishi A. Fuzzy risk stratification and risk assessment model for clinical monitoring in the ICU. *Comput Biol Med* 2017;87:169–78. [PubMed: 28599216]
44. Chen Z, Bird VY, Ruchi R, Segal MS, Bian J, Khan SR, et al. Development of a personalized diagnostic model for kidney stone disease tailored to acute care by integrating large clinical, demographics and laboratory data: the diagnostic acute care algorithm - kidney stones (DACA-KS). *BMC Med Inform Decis Mak* 2018;18(1):72. [PubMed: 30119627]
45. Cramer EM, Seneviratne MG, Sharifi H, Ozturk A, Hernandez-Boussard T. Predicting the Incidence of Pressure Ulcers in the Intensive Care Unit Using Machine Learning. *EGEMS (Wash DC)* 2019;7(1):49. [PubMed: 31534981]
46. Xia J, Pan S, Zhu M, Cai G, Yan M, Su Q, et al. A Long Short-Term Memory Ensemble Approach for Improving the Outcome Prediction in Intensive Care Unit. *Comput Math Methods Med* 2019;2019:8152713. [PubMed: 31827589]
47. Rongali S, Rose AJ, McManus DD, Bajracharya AS, Kapoor A, Granillo E, et al. Learning Latent Space Representations to Predict Patient Outcomes: Model Development and Validation. *J Med Internet Res* 2020;22(3):e16374. [PubMed: 32202503]
48. Lee DH, Yetisgen M, Vanderwende L, Horvitz E. Predicting severe clinical events by learning about life-saving actions and outcomes using distant supervision. *J Biomed Inform* 2020;107:103425.
49. Su L, Liu C, Li D, He J, Zheng F, Jiang H, et al. Toward Optimal Heparin Dosing by Comparing Multiple Machine Learning Methods: Retrospective Study. *JMIR Med Inform* 2020;8(6):e17648. [PubMed: 32568089]
50. Eickelberg G, Sanchez-Pinto LN, Luo Y. Predictive modeling of bacterial infections and antibiotic therapy needs in critically ill adults. *J Biomed Inform* 2020;109:103540.
51. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of Sepsis in the Intensive Care Unit With Minimal Electronic Health Record Data: A Machine Learning Approach. *JMIR Med Inform* 2016;4(3):e28. [PubMed: 27694098]
52. Ghosh S, Li J, Cao L, Ramamohanarao K. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *J Biomed Inform* 2017;66:19–31. [PubMed: 28011233]
53. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. *Comput Biol Med* 2017;89:248–55. [PubMed: 28843829]
54. Liu R, Greenstein JL, Granite SJ, Fackler JC, Bembea MM, Sarma SV, et al. Data-driven discovery of a novel sepsis pre-shock state predicts impending septic shock in the ICU. *Sci Rep* 2019;9(1):6145. [PubMed: 30992534]

55. Scherpf M, Gräßer F, Malberg H, Zaunseder S. Predicting sepsis with a recurrent neural network using the MIMIC III database. *Comput Biol Med* 2019;113:103395. [PubMed: 31480008]
56. Fagerström J, Bång M, Wilhelms D, Chew MS. LiSep LSTM: A Machine Learning Algorithm for Early Detection of Septic Shock. *Sci Rep* 2019;9(1):15132. [PubMed: 31641162]
57. Song W, Jung SY, Baek H, Choi CW, Jung YH, Yoo S. A Predictive Model Based on Machine Learning for the Early Detection of Late-Onset Neonatal Sepsis: Development and Observational Study. *JMIR Med Inform* 2020;8(7):e15965. [PubMed: 32735230]
58. Yao RQ, Jin X, Wang GW, Yu Y, Wu GS, Zhu YB, et al. A Machine Learning-Based Prediction of Hospital Mortality in Patients With Postoperative Sepsis. *Front Med (Lausanne)* 2020;7:445. [PubMed: 32903618]
59. Lee J, Mark RG. An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care. *Biomed Eng Online* 2010;9:62. [PubMed: 20973998]
60. Lee J, Mark R. A Hypotensive Episode Predictor for Intensive Care based on Heart Rate and Blood Pressure Time Series. *Comput Cardiol* (2010) 2011;2010(26–29 Sept. 2010):81–4. [PubMed: 21874161]
61. Paradkar N, Chowdhury SR. Coronary artery disease detection using photoplethysmography. *Conf Proc IEEE Eng Med Biol Soc* 2017;2017:100–3.
62. Kaji DA, Zech JR, Kim JS, Cho SK, Dangayach NS, Costa AB, et al. An attention based deep learning model of clinical events in the intensive care unit. *PLoS One* 2019;14(2):e0211057. [PubMed: 30759094]
63. Barrett LA, Payrovnaziri SN, Bian J, He Z. Building Computational Models to Predict One-Year Mortality in ICU Patients with Acute Myocardial Infarction and Post Myocardial Infarction Syndrome. *AMIA Jt Summits Transl Sci Proc* 2019;2019:407–16. [PubMed: 31258994]
64. Payrovnaziri SN, Barrett LA, Bis D, Bian J, He Z. Enhancing Prediction Models for One-Year Mortality in Patients with Acute Myocardial Infarction and Post Myocardial Infarction Syndrome. *Stud Health Technol Inform* 2019;264:273–7. [PubMed: 31437928]
65. Cherifa M, Blet A, Chambaz A, Gayat E, Resche-Rigon M, Pirracchio R. Prediction of an Acute Hypotensive Episode During an ICU Hospitalization With a Super Learner Machine-Learning Algorithm. *Anesth Analg* 2020;130(5):1157–66. [PubMed: 32287123]
66. Morid MA, Sheng ORL, Del Fiore G, Facelli JC, Bray BE, Abdelrahman S. Temporal Pattern Detection to Predict Adverse Events in Critical Care: Case Study With Acute Kidney Injury. *JMIR Med Inform* 2020;8(3):e14272. [PubMed: 32181753]
67. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. *Crit Care* 2019;23(1):112. [PubMed: 30961662]
68. Zimmerman LP, Reyfman PA, Smith ADR, Zeng Z, Kho A, Sanchez-Pinto LN, et al. Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements. *BMC Med Inform Decis Mak* 2019;19(Suppl 1):16. [PubMed: 30700291]
69. Sun M, Baron J, Dighe A, Szolovits P, Wunderink RG, Isakova T, et al. Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes and Structured Multivariate Physiological Measurements. *Stud Health Technol Inform* 2019;264:368–72. [PubMed: 31437947]
70. Wang Y, Wei Y, Yang H, Li J, Zhou Y, Wu Q. Utilizing imbalanced electronic health records to predict acute kidney injury by ensemble learning and time series model. *BMC Med Inform Decis Mak* 2020;20(1):238.
71. Mikhno A, Ennett CM. Prediction of extubation failure for neonates with respiratory distress syndrome using the MIMIC-II clinical database. *Conf Proc IEEE Eng Med Biol Soc* 2012;2012:5094–7.
72. Behar J, Oster J, Li Q, Clifford GD. ECG signal quality during arrhythmia and its application to false alarm reduction. *IEEE Trans Biomed Eng* 2013;60(6):1660–6. [PubMed: 23335659]
73. Wang X, Gao Y, Lin J, Rangwala H, Mittu R, editors. A Machine Learning Approach to False Alarm Detection for Critical Arrhythmia Alarms. *Proceedings of 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*; 2015 9–11 Dec. 2015. p.202–7.

74. Ding Y, Ma X, Wang Y. Health status monitoring for ICU patients based on locally weighted principal component analysis. *Comput Methods Programs Biomed* 2018;156:61–71. [PubMed: 29428077]
75. Ren O, Johnson AEW, Lehman EP, Komorowski M, Aboab J, Tang F, et al., editors. Predicting and Understanding Unexpected Respiratory Decompensation in Critical Care Using Sparse and Heterogeneous Clinical Data. *Proceedings of 2018 IEEE International Conference on Healthcare Informatics (ICHI)*; 2018 4–7 June 2018. p.144–51.
76. Desautels T, Das R, Calvert J, Trivedi M, Summers C, Wales DJ, et al. Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. *BMJ Open* 2017;7(9):e017199.
77. McWilliams CJ, Lawson DJ, Santos-Rodriguez R, Gilchrist ID, Champneys A, Gould TH, et al. Towards a decision support tool for intensive care discharge: machine learning algorithm development using electronic healthcare data from MIMIC-III and Bristol, UK. *BMJ Open* 2019;9(3):e025925.
78. Rojas JC, Carey KA, Edelson DP, Venable LR, Howell MD, Churpek MM. Predicting Intensive Care Unit Readmission with Machine Learning Using Electronic Health Record Data. *Ann Am Thorac Soc* 2018;15(7):846–53. [PubMed: 29787309]
79. Lin YW, Zhou Y, Faghri F, Shaw MJ, Campbell RH. Analysis and prediction of unplanned intensive care unit readmission using recurrent neural networks with long short-term memory. *PLoS One* 2019;14(7):e0218942. [PubMed: 31283759]
80. Sanchez-Pinto LN, Luo Y, Churpek MM. Big Data and Data Science in Critical Care. *Chest* 2018;154(5):1239–48. [PubMed: 29752973]
81. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *Journal of global health* 2018;8(2):020303-. [PubMed: 30405904]
82. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Future healthcare journal* 2019;6(2):94–8.
83. Elshawi R, Al-Mallah MH, Sakr S. On the interpretability of machine learning-based model for predicting hypertension. *BMC medical informatics and decision making* 2019;19(1):146-. [PubMed: 31357998]
84. Li Y, Yao L, Mao C, Srivastava A, Jiang X, Luo Y, editors. Early Prediction of Acute Kidney Injury in Critical Care Setting Using Clinical Notes. *Proceedings of 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*; 2018 3–6 Dec. 2018. p.683–6.
85. de Geus HR, Betjes MG, Bakker J. Biomarkers for the prediction of acute kidney injury: a narrative review on current status and future challenges. *Clin Kidney J* 2012;5(2):102–8. [PubMed: 22833807]
86. Tsai T-Y, Chien H, Tsai F-C, Pan H-C, Yang H-Y, Lee S-Y, et al. Comparison of RIFLE, AKIN, and KDIGO classifications for assessing prognosis of patients on extracorporeal membrane oxygenation. *Journal of the Formosan Medical Association* 2017;116(11):844–51. [PubMed: 28874330]
87. Ogundipe F, Kodadhala V, Ogundipe T, Mehari A, Gillum R. Disparities in Sepsis Mortality by Region, Urbanization, and Race in the USA: a Multiple Cause of Death Analysis. *Journal of Racial and Ethnic Health Disparities* 2019;6(3):546–51. [PubMed: 30607577]
88. Liu V, Escobar GJ, Greene JD, Soule J, Whippy A, Angus DC, et al. Hospital Deaths in Patients With Sepsis From 2 Independent Cohorts. *JAMA* 2014;312(1):90–2. [PubMed: 24838355]
89. Tang F, Ishwaran H. Random Forest Missing Data Algorithms. *Statistical analysis and data mining* 2017;10(6):363–77. [PubMed: 29403567]
90. Khan NM, Madhav CN, Negi A, Thaseen IS, editors. Analysis on Improving the Performance of Machine Learning Models Using Feature Selection Technique. *Proceedings of Intelligent Systems Design and Applications*; 2020 2020//. p.69–77.
91. Tabe-Bordbar S, Emad A, Zhao SD, Sinha S. A closer look at cross-validation for assessing the accuracy of gene regulatory networks and models. *Scientific reports* 2018;8(1):6620-. [PubMed: 29700343]
92. Raschka S Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *ArXiv* 2018;abs/1811.12808.

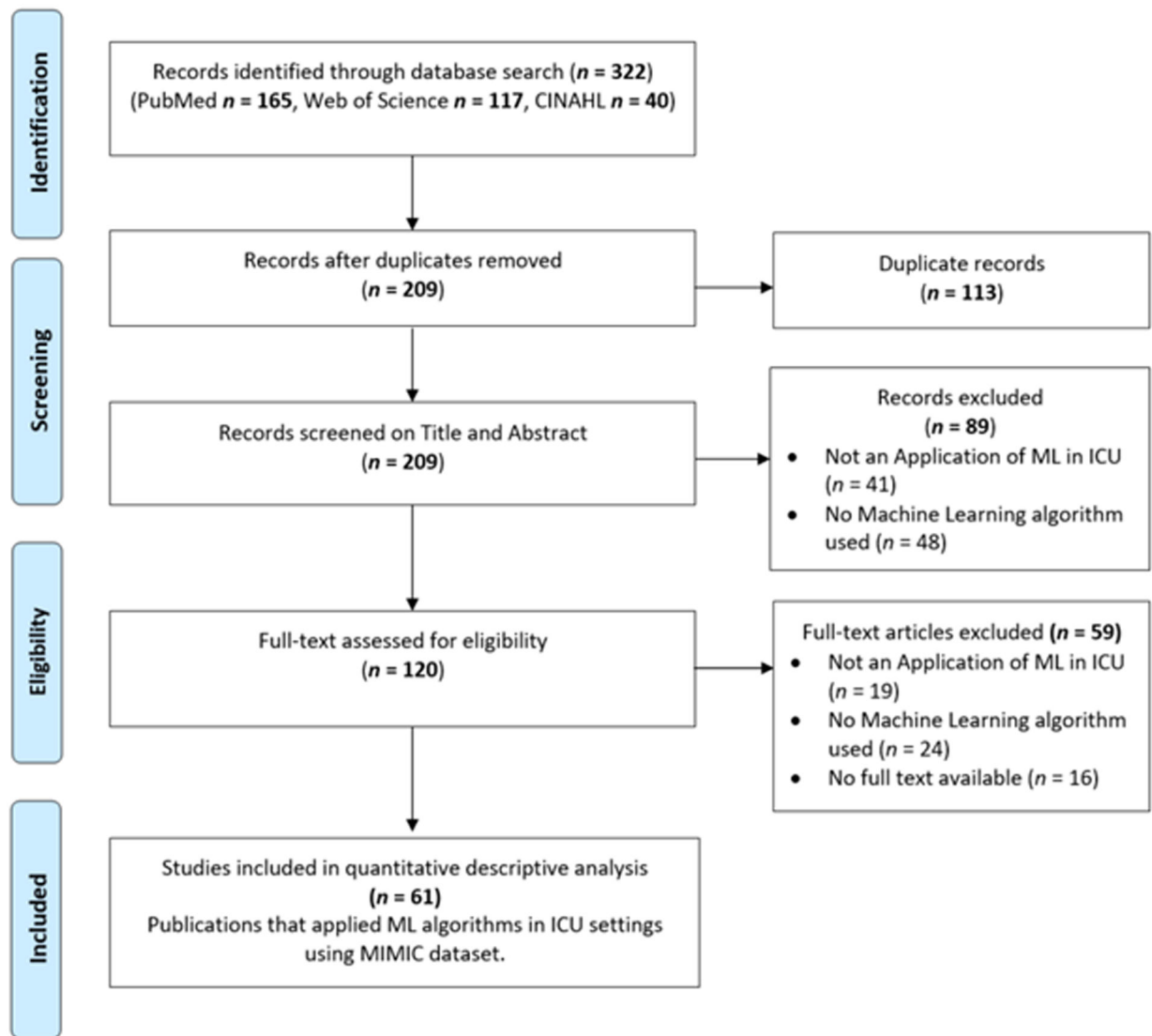


Figure 1.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of systematic identification, screening, eligibility, and inclusion of publications that applied Machine Learning (ML) techniques in Intensive Care Unit (ICU) settings using Medical Information Mart for Intensive Care (MIMIC) dataset.

Table 1.

An overview of the 61 publications in the literature, classified into seven themes and their descriptions. The themes are listed according to the frequency of publication (percentage and absolute count).

S.NO	Theme	Description	Publication References	Publications Count (%)
1	Mortality Prediction	Publications focused on prediction hospital and ICU mortality.	[21–41]	21 (34.4)
2	Risk Stratification	Publications focused on assessing the risk of treatments.	[17,42–50]	10 (16.4)
3	Sepsis and Septic Shock Prediction	Publications focused on prediction of sepsis or septic shock.	[51–58]	8 (13.1)
4	Cardiac episodes Prediction	Publications focused on predicting cardiac events such as cardiac disease, myocardial infarction.	[59–65]	7 (11.5)
5	Acute Kidney Injury Prediction	Publications focused on early prediction and onset prediction of acute kidney injury.	[66–70]	6 (9.8)
6	Resource Management	Publications focused on ICU staff and resource management by suppressing false alarms in ICU.	[71–75]	5 (8.2)
7	ICU Readmission	Publications focused on prediction of discharge and readmission.	[76–79]	4 (6.6)