Available online at **www.ijmrhs.com**

ISSN No: 2319-5886

International Journal of Medical Research & Health Sciences, 2017, 6(6): 1-7

Use of Machine-Learning Approaches to Predict Clinical Deterioration in Critically Ill Patients: A Systematic Review

Tadashi Kamio[1,2*], Tomoaki Van[2] and Ken Masamune[2]

[1] *Department of Anesthesiology and Critical Care Medicine, Jichi Medical University Saitama Medical Center, Saitama, Japan*

[2] *Institute of Advanced BioMedical Engineering and Science, Tokyo Women's Medical University, Tokyo, Japan*

**Corresponding e-mail: tadashi-kamio@mail.goo.ne.jp*

ABSTRACT

Introduction: *Early identification of patients with unexpected clinical deterioration is a matter of serious concern. Previous studies have shown that early intervention on a patient whose health is deteriorating improves the patient outcome, and machine-learning-based approaches to predict clinical deterioration may contribute to precision improvement. To date, however, no systematic review in this area is available.* ***Methods:*** *We completed a search on PubMed on January 22, 2017 as well as a review of the articles identified by study authors involved in this area of research following the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guidelines for systematic reviews.* ***Results:*** *Twelve articles were selected for the current study from 273 articles initially obtained from the PubMed searches. Eleven of the 12 studies were retrospective studies, and no randomized controlled trials were performed. Although the artificial neural network techniques were the most frequently used and provided high precision and accuracy, we failed to identify articles that showed improvement in the patient outcome. Limitations were reported related to generalizability, complexity of models, and technical knowledge.* ***Conclusions:*** *This review shows that machine-learning approaches can improve prediction of clinical deterioration compared with traditional methods. However, these techniques will require further external validation before widespread clinical acceptance can be achieved.*

Keywords: Machine learning, critical care, intensive care unit, clinical deterioration, prediction

Abbreviations: HR: Heart Rate; RR: Respiration Rate; SpO_2: Oxygen Saturation; SBP: Systolic Blood Pressure; DBP: Diastolic Blood Pressure; MAP: Mean Arterial Pressure; T: Temperature; GCS: Glasgow Coma Score; HRV: Heart Rate Variability; RRV: Respiratory Rate Variability; ECG: Electrocardiogram; PPG: Photoplethysmogram; BUN: Blood Urea Nitrogen; Ht: Hematocrit; PTT: Partial Thromboplastin Time; WBC: White Blood Cell Count; k-NN: k-Nearest Neighbors; ANN: Artificial Neural Network; SVM: Support Vector Machine; RIPPER: Repeated Incremental Pruning to Produce Error Reduction; AUROC: Area Under The Receiver Operating Characteristic Curves; PPV: Positive Predictive Values; NPV: Negative Predictive Values

INTRODUCTION

Early identification, recognition, and acknowledgement of patients with unexpected clinical deterioration are a matter of serious concern. Early intervention on a patient whose health is deteriorating will likely improve the patient outcome, and delayed intervention has been associated with increased morbidity and mortality [1-4]. The modified early warning score (MEWS) can be used on all hospitalized patients to allow early detection of clinical deterioration and of potential needs for higher level of care [5]. However, the basic approach of data collection and management has remained largely unchanged over the past 40 years [6]. Moreover, a study on computerized physiological monitoring systems found that medical and nursing staff had difficulty identifying the onset of adverse trends as they develop but could identify when a trend had commenced when they retrospectively looked at them [7]. This result suggests that relying on the staff to identify gradual deterioration without some forms of assistance such as a track and trigger system may mean that patients whose health is deteriorating will be missed.

On the other hand, other newer monitoring technologies based on statistical machine-learning theory are available, which provide early recognition of patient health deterioration [8,9]. Machine learning is a scientific discipline that focuses on how computers learn from data [10]. The adoption of data-intensive machine-learning methods can be found throughout science, technology, and commerce, leading to more evidence-based decision making across many walks of life [11]. Machine-learning technology is currently well suited for analyzing medical data, and many works have been performed in medical diagnosis. The algorithms of machine learning will improve prognosis, displace much of the work of radiologists and anatomical pathologists, and improve diagnostic accuracy [12].

Artificial intelligence (AI) is a field that develops intelligent algorithms and machines. Most of the researchers today agree that no intelligence exists without learning [13]. Therefore, machine learning is one of the major branches of AI, and it is one of the most rapidly developing subfields in AI research. AI in medicine has also become a very important field in computer-aided medical research, which covers various fields (e.g., automated diagnosis and therapy recommendation, image recognition and interpretation, patient management, and telemedicine/telehealth, among others) [14]. Combination of big data and AI is believed to drastically change from conventional practice [15].

In general, more data are needed for these approaches to obtain meaningful results [16]. However, most of the data generated in the medical-care process have historically been underused due to the difficulty in accessing, organizing, and using the data entered on paper charts [17]. In contrast, several commercial and noncommercial intensive care unit (ICU) databases have been developed [17]. The high level of monitoring in an ICU provides a unique opportunity for machine learning to provide new insights. Machine-learning-based approaches to predict clinical deterioration may help more precisely determine anomalies in physiological parameters; therefore, dozens of algorithms have been proposed in this area [8,9].

Although some reviews have reported machine-learning methods in critical care, these reviews have not systematically described the performance outcomes [18,19]. Therefore, considering the rapid growth of AI or machine-learning applications in critical care, the current review specializes on machine-learning-based approaches to predict clinical deterioration in critical care. The objectives of this systematic review are to identify all studies that used machine-learning algorithms to predict clinical deterioration in critically ill patients and to compare the reported performance and utility of these clinical applications.

MATERIALS AND METHODS

We searched for studies regarding AI or machine-learning approaches being used to analyze ICU patient data in January 22, 2017 (date of search). A protocol for the review was written a priori and was followed in detail. We identified, categorized, and analyzed the factors into various themes, and the results were presented according to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) reporting guidelines [20].

Data sources and search strategy

Two reviewers independently screened the titles and abstracts of articles obtained by the following search procedure. We created a new PubMed query to provide a better context of the area under study so that our clinical query should focus on AI or machine-learning approaches. The new query that appeared under the search details when using the PubMed search engine was [“machine learning” OR “AI” OR “artificial neural network (ANN)”] AND (“ICU” OR “intensive care” OR “critical care” OR “critically ill”) (anywhere in the full text of the paper). All citations were imported to an electronic database (Endnote X3, Thomson Reuters, New York, NY, USA).

Inclusion/exclusion criteria

The following criteria were used to screen the papers for inclusion in the systematic review:

1) The paper was published in a peer-reviewed journal

2) The literature was restricted to studies published in English

3) The study populations were adult ICU patients (>18 years)

4) At least one of the study objectives was to use AI or machine-learning approaches to predict clinical deterioration using some vital-sign data

5) The data presented (or could be used to calculate) summary statistics (e.g., sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), or receiver operating characteristic (ROC) curves, etc.).

We included studies published between 2000 and January 22, 2017 (date of search) targeted at critically ill adult patients and excluded studies when they focused on mortality prediction models or alarm classification. The review was restricted to studies from 2000 because the developments of computerized critical-care database, monitoring systems, and computer technology for machine learning started in this decade [12]. This restriction in the time period reduced heterogeneity due to technical development. Studies that did not satisfy all five inclusion criteria were excluded from the review. Furthermore, we excluded articles in which no full text was available through a license at our institutes.

Data extraction

The titles and abstracts of the identified citations were read to screen the articles based on the selection criteria described in the previous section. The remaining articles were read in full text to extract information from each article. We extracted data that described the year, country, design, study population, machine-learning techniques, variables, and predictive performance. The extracted information is listed in Tables 1 and 2. For each study, summary statistics were either extracted from the article. No meta-analysis was performed because of the heterogeneity among studies.

RESULTS

Quantity of research available

Figure 1 shows the results of the article screening and selection process. The electronic search yielded 273 citations in which 244 citations were excluded based on the title and abstract. From these citations, we reviewed 29 full-text articles and selected 12 articles for final inclusion in the study [21-32].

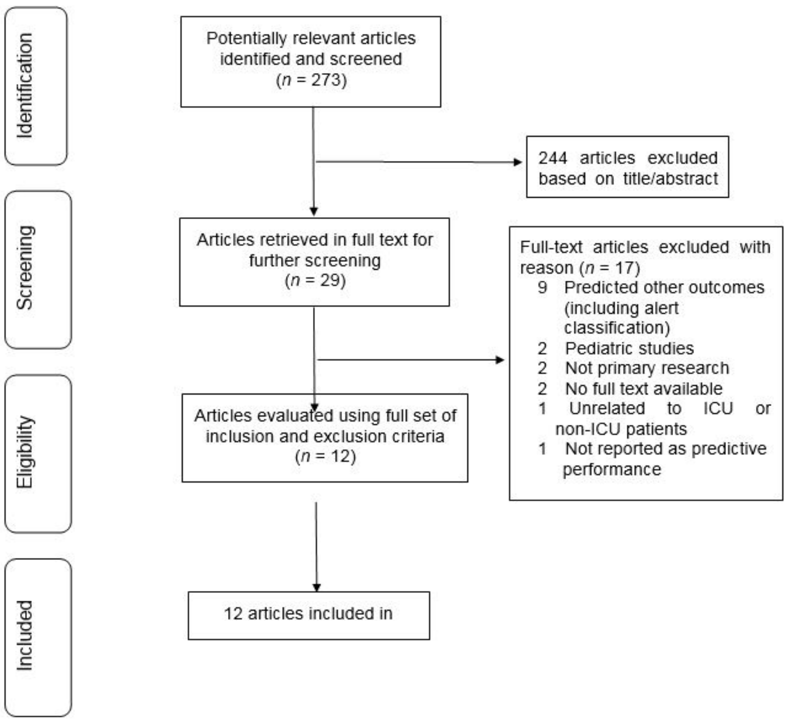

Figure 1 PRISMA flow diagram

This figure presents the trial flow diagram to identify the eligible articles for this study. A total of 273 articles were identified from the literature searches. Finally, based on the predefined criteria, 12 articles were included in the review. The selection criteria included empirical qualitative studies published from 2000 to 2017. The earliest eligible articles were published in 2008.

The most common reason for exclusion of articles after the full-text review was that the articles did not examine prediction of clinical deterioration or focused on the prediction of other outcomes (e.g., mortality, length of stay, and false-alarm classification, among others).

Study characteristics

The publication years ranged from 2000 to 2017. The included studies came from a variety of settings and countries. As listed in Table 1, six studies were published in the USA [21,24,28,29,31,32], two in Australia [22,30], and one each in Puerto Rico [23], Korea [25], Singapore [26], and UK [27]. Although only one study was a prospective observational cohort study [26], the others were observational studies that reported the prediction of clinical deterioration in critically ill patients. Of the 12 studies, four studies focused on prediction of hypotensive events [23,27-29], and three studies focused on sepsis prediction [22,24,30]. Three studies reported baseline characteristics of study patients; however, the remaining studies did not describe patient demographics.

Table 1 Study characteristics

Year	First Author	Region	Data acquisition	Patient characteristics reported	Population
2016	Chen	USA	Retrospective	Yes	1880 patients in surgical-trauma step-down unit
2016	Ghosh	Australia	Retrospective	No	1,310 ICU patients
2016	Ordoñez	Puerto Rico	Retrospective	No	58 ICU patients
2016	Desautels	USA	Retrospective	No	19,828 ICU patients
2016	Lee	Korea	Retrospective	No	15 cardiovascular ICU patients
2012	Ong	Singapore	Prospective (non-randomized observational cohort)	Yes	925 critically ill patients in an emergency department
2012	Donald	UK	Retrospective	Yes	119 ICU patients
2011	Lee	USA	Retrospective	No	1,357 records in ICU
2010	Lee	USA	Retrospective	No	1,311 records in ICU
2010	Tang	Australia	Retrospective	No	26 sepsis patients in an emergency department
2009	Crump	USA	Retrospective	No	36 ICU patients
2008	Eshelman	USA	Retrospective	No	12,695 ICU patients

Variable choice

All studies utilized several variables (Table 2). Nine studies used heart rate [21,22,24,26-29,31,32], and three studies used heart-rate variability for modeling applications [25,26]. Age and gender were used in the predictive models in Ong, et al. [26], Donald, et al. [27], and Crump, et al. [31]. Eshelman, et al. [32] proposed an algorithm consisting of a set of 15 rules.

Machine-learning approaches and predictive performance

Several machine-learning techniques have been used for predicting clinical deterioration in critically ill patients. As listed in Table 2, the most common methods used are ANNs [25,27-29], and two studies used support vector machine (SVM) [26,30]. The remaining studies used the random forest classification model [21], coupled hidden Markov models [22], k-nearest neighbors (k-NN) [23], continuous nonlinear function approximations [24], Bayesian network models [31], and repeated incremental pruning to produce error reduction (RIPPER) algorithm [32]. Discrimination, as measured by the area under ROC curve (AUROC), was reported in seven studies of the prediction models, and these values are listed in Table 2. One research, which used trained ANNs to predict ventricular tachycardia events 1 h before its onset, achieved 85.3% accuracy for the test set [25]. Ong, et al. reported that machine-learning scores that adopted the SVM were more accurate than the MEWS in predicting cardiac arrest within 72 h (0.781 versus 0.680; differences in AUROC were as follows: 0.101, 95%; CI: from 0.006 to 0.197; P=0.037) [26].

Table 2 Prediction models

Author	Prediction outcome	Used variables	Machine-learning techniques	Predictive performance (best performance)
Chen [21]	Cardiorespiratory insufficiency	HR, RR, SpO_2, SBP, DBP	Random forest classification model	AUROC=0.94
Ghosh [22]	Onset of septic shock	MAP, HR, RR	Coupled hidden Markov models	Likelihood 0.71

Ordoñez [23]	Hypotension scenario within an hour	SpO_2, SBP, DBP	k-NN	Accuracy=0.85, Precision=0.82, Recall=0.87 and F-Measure=0.86
Desautels [24]	Sepsis onset	SBP, pulse pressure, HR, RR, T, SpO_2, age, GCS	Continuous nonlinear function approximations	AUROC=0.880 at onset time
Lee [25]	Ventricular tachycardia 1 h before occurrence	HRV, RRV	ANN	Accuracy=0.853, Sensitivity=0.882, Specificity=0.824, PPV=0.833, NPV=0.875, AUCROC=0.93
Ong [26]	Cardiac arrest within 72 h	HRV, age, gender, medical history, HR, BP, SpO_2, RR, GCS	SVM	AUROC=0.781 (compared with 0.680 for MEWS)
Donald [27]	Hypotensive events	Age, sex, SBP, BP (mean), HR	Bayesian ANN	Sensitivity=0.40, Specificity=0.86
Lee (2011)	Hypotensive events	HR, SBP, DBP, MAP	ANN	AUROC=0.934, Accuracy=0.861, Sensitivity=0.851, Specificity=0.862, PPV=0.151, NPV=0.995
Lee [28,29]	Hypotensive events	MAP, HR, pulse pressure, relative cardiac output	ANN	AUROC=0.918, Sensitivity=0.826, Specificity=0.859
Tang [30]	Discriminate severe sepsis patients from SIRS patients	ECG signal, PPG waveform	SVM	Sensitivity=0.94, Specificity=0.62, PPV=0.85, NPV=0.83, Accuracy=0.84
Crum [31]	Setting alerts from personal baselines	Age, gender, T, HR, SpO_2, admission diagnosis	Bayesian network models	AUROC=0.91
Eshelman [32]	Identifying hemodynamically unstable patients	BUN, WBC, PTT, Ht, HR, SBP, oxygenation index	RIPPER algorithm	Sensitivity=0.60, Specificity=0.9285, PPV=0.7970

Although all studies realized fine results, several limitations were present in the machine-learning techniques. All 12 studies were carried out in a single-center study, which may have limited generalization of the results to other hospitals and hospital systems. Because machine-learning techniques are highly dependent on the underlying data, lack of standardization may lead to significant discrepancies. Furthermore, three studies highlighted the problems of assessing the influence of different probability cutoff thresholds [22,27,28].

DISCUSSION

In this study, we systematically reviewed the literature to identify all studies that used machine-learning algorithms to predict clinical deterioration in critically ill patients and compared the reported performance. We identified 12 studies and showed that various models for predicting patient deterioration employed many algorithms to achieve good results. However, 11 of the 12 studies were retrospective studies, and we could not identify any randomized controlled trials. Although providing early predictors of potentially life-threatening conditions in critically ill patients could improve outcomes, whether these models can actually improve patient outcomes remains unknown. Only limited clinical evidence is available, and thus, more definitive clinical trials are needed to educate clinicians on the efficacy of predicting patient deterioration in ICUs.

Among these studies, ANN techniques were the most frequently used. ANNs are computational models inspired by the human brain [33]. In the current decade, the application of ANNs has dramatically progressed, and they are applied to medical data, which turns out to be highly effective in analyzing and modeling [34]. The reason behind this is the huge and rapid growth in the ability of networked and mobile computing systems to gather and transport vast amounts of data. Jordan, et al. reported that machine learning is likely to be one of the most transformative technologies in the 21st century [12]. According to the technological development, this trend will increasingly become active in the future.

On the other hand, the comparison among the different prediction models using machine-learning techniques is often difficult due to a number of factors: the employed datasets are not always the same, and the selected set of variables and algorithms do not always match. Standardization of data will probably be a future problem. Moreover, the difference in evaluation metrics may lead to important discrepancies, which means that these techniques will require further external validation before widespread clinical acceptance can be realized.

We identified potentially important studies to predict clinical deterioration in critically ill patients by systematically reviewing the literature and provided a comprehensive summary of machine-learning approaches. However, our study is not without limitations. Limiting data to studies from 2000 onwards may have excluded very early studies that examined the effect of predicting patient deterioration using machine-learning algorithms. In addition, we excluded articles in which no full text was available (n=2), which may have led to some underestimation of the number of models and external validations in the search period.

CONCLUSION

Our systematic review included 12 studies that used machine-learning algorithms to predict clinical deterioration of critically ill patients. This review shows that machine-learning approaches can improve prediction of clinical deterioration compared with traditional methods. However, no randomized controlled trials have evaluated the independent effects of these approaches, and thus, further studies are needed to demonstrate evidence that these models can improve patient outcomes.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

[1] Brennan, Troyen A., et al. "Incidence of adverse events and negligence in hospitalized patients: results of the Harvard Medical Practice Study I." *New England journal of medicine* 324.6 (1991): 370-376.

[2] Baker, G. Ross, et al. "The Canadian adverse events study: The incidence of adverse events among hospital patients in Canada." *Canadian medical association journal* 170.11 (2004): 1678-1686.

[3] Moon, A., et al. "An eight-year audit before and after the introduction of modified early warning score (MEWS) charts, of patients admitted to a tertiary referral intensive care unit after CPR." *Resuscitation* 82.2 (2011): 150-154.

[4] Paterson, R., et al. "Prediction of in-hospital mortality and length of stay using an early warning scoring system: Clinical audit." *Clinical Medicine* 6.3 (2006): 281-284.

[5] Duckitt, R. W., et al. "Worthing physiological scoring system: Derivation and validation of a physiological early-warning system for medical admissions. An observational, population-based single-centre study." *British journal of anaesthesia* 98.6 (2007): 769-774.

[6] de Georgia, Michael A., et al. "Information technology in critical care: Review of monitoring and data acquisition systems for patient care and research." *The Scientific World Journal* 2015 (2015).

[7] Alberdi, Eugenio, et al. "Computerization and decision making in neonatal intensive care: A cognitive engineering investigation." *Journal of Clinical Monitoring and Computing* 16.2 (2000): 85-94.

[8] Goldstein, Markus, and Seiichi Uchida. "A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data." *PloS one* 11.4 (2016): e0152173.

[9] Hu, Scott B., et al. "Prediction of clinical deterioration in hospitalized adult patients with hematologic malignancies using a neural network model." *PloS one* 11.8 (2016): e0161401.

[10] Deo, Rahul C. "Machine learning in medicine." *Circulation* 132.20 (2015): 1920-1930.

[11] Jordan, Michael I., and Tom M. Mitchell. "Machine learning: Trends, perspectives, and prospects." *Science* 349.6245 (2015): 255-260.

[12] Obermeyer, Ziad, and Ezekiel J. Emanuel. "Predicting the future-big data, machine learning, and clinical medicine." *The New England journal of medicine* 375.13 (2016): 1216.

[13] Wang, Hua, Cuiqin Ma, and Lijuan Zhou. "A brief review of machine learning and its application." *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*. IEEE, 2009.

[14] Gorunescu, Florin. "Intelligent decision systems in medicine: A short survey on medical diagnosis and patient management." *E-Health and Bioengineering Conference (EHB), 2015*. IEEE, 2015.

[15] Jha, Saurabh, and Eric J. Topol. "Adapting to artificial intelligence: Radiologists and pathologists as information specialists." *JAMA* 316.22 (2016): 2353-2354.

[16] Webb, Geoffrey I., Michael J. Pazzani, and Daniel Billsus. "Machine learning for user modeling." *User modeling and user-adapted interaction* 11.1 (2001): 19-29.

[17] Anthony Celi, Leo, et al. ""Big data" in the intensive care unit. Closing the data loop." *American Journal of Respiratory and Critical Care Medicine* (2013): 1157-1160.

[18] Johnson, Alistair E. W., et al. "Machine learning and decision support in critical care." *Proceedings of the IEEE* 104.2 (2016): 444-466.

[19] Nizami, Shermeen, James R. Green, and Carolyn McGregor. "Implementation of artifact detection in critical care: A methodological review." *IEEE reviews in biomedical engineering* 6 (2013): 127-142.

[20] Moher, David, et al. "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement." *PLoS med* 6.7 (2009): e1000097.

[21] Chen, Lujie, et al. "Dynamic and personalized risk forecast in step-down units. Implications for monitoring paradigms." *Annals of the American Thoracic Society* 14.3 (2017): 384-391.

[22] Ghosh, Shameek, et al. "Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns." *Journal of biomedical informatics* 66 (2017): 19-31.

[23] Ordoñez, Patricia, et al. "Learning stochastic finite-state transducer to predict individual patient outcomes." *Health and Technology* 6.3 (2016): 239-245.

[24] Desautels, Thomas, et al. "Prediction of sepsis in the intensive care unit with minimal electronic health record data: A machine learning approach." *JMIR medical informatics* 4.3 (2016).

[25] Lee, Hyojeong, et al. "Prediction of ventricular tachycardia one hour before occurrence using artificial neural networks." *Scientific Reports* 6 (2016).

[26] Ong, Marcus Eng Hock, et al. "Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score." *Critical Care* 16.3 (2012): R108.

[27] Donald, Rob, et al. Early warning of EUSIG-defined hypotensive events using a Bayesian Artificial Neural Network. *Springer* Vienna, 2012.

[28] Lee, Joon, and Mark, R. G. "A hypotensive episode predictor for intensive care based on heart rate and blood pressure time series." *Computing in Cardiology, 2010*. IEEE, 2010.

[29] Lee, Joon, and Roger G. Mark. "An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care." *Biomedical engineering online* 9.1 (2010): 62.

[30] Tang, Collin H. H., et al. "Non-invasive classification of severe sepsis and systemic inflammatory response syndrome using a nonlinear support vector machine: A preliminary study." *Physiological measurement* 31.6 (2010): 775.

[31] Crump, Cindy, et al. "Using Bayesian networks and rule-based trending to predict patient status in the intensive care unit." *AMIA Annual Symposium Proceedings*. Vol. 2009. American Medical Informatics Association, 2009.

[32] Eshelman, Larry J., et al. "Development and evaluation of predictive alerts for hemodynamic instability in ICU patients." *AMIA*. 2008.

[33] Zhou, Zhi-Hua, et al. "Lung cancer cell identification based on artificial neural network ensembles." *Artificial Intelligence in Medicine* 24.1 (2002): 25-36.

[34] Pastur-Romay, Lucas Antón, et al. "Deep artificial neural networks and neuromorphic chips for big data analysis: Pharmaceutical and bioinformatics applications." *International Journal of Molecular Sciences* 17.8 (2016): 1313.