Citation	Güiza F, Van Eyck J, (2013), Predictive data mining on monitoring data from the intensive care unit J Clin Monit Comput. 2013 Aug;27(4):449-53
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://dx.doi.org/10.1007/s10877-012-9416-3
Journal homepage	http://www.springer.com/medicine/anesthesiology/journal/10877
Author contact	fabian.guiza@kuleuven.be
IR	https://lirias.kuleuven.be/handle/123456789/

(article begins on next page)

Predictive data mining on monitoring data from the intensive care unit.

Fabian Güiza, PhD¹; Jelle Van Eyck²; Geert Meyfroidt, MD, PhD³

1. Department of intensive care medicine, University of Leuven (KU Leuven),
Belgium Herestraat 49, 3000 Leuven, Belgium

2. Department of Computer Science, University of Leuven (KU Leuven), Belgium

Corresponding author: Fabian Güiza

fabian.guiza@kuleuven.be

tel: +32 16 344232

fax: +32 16 344015

Keywords:

Intensive care, Artificial intelligence, Machine learning, Patient monitoring, Critical illness, Computerized patient medical records

Abstract

The widespread implementation of computerized medical files in intensive care units (ICUs) over recent years has made available large databases of clinical data for the purpose of developing clinical prediction models. The typical intensive care unit has several information sources from which data is electronically collected as time series of varying time resolutions. We present an overview of research questions studied in the ICU setting that have been addressed through the automatic analysis of these large databases. We focus on automatic learning methods, specifically data mining approaches for predictive modeling based on these time series of clinical data. On the one hand we examine short and medium term predictions, which have as ultimate goal the development of early warning or decision support systems. On the other hand we examine long term outcome prediction models and evaluate their performance with respect to established scoring systems based on static admission and demographic data.

Introduction

Over the last few decades many intensive care units (ICUs) worldwide have implemented electronic systems to collect high frequency measurements and closely monitor their critically ill patients [1]. This has led to the development of large databases to store the vast amounts of data routinely generated from the different information sources in the ICU. These electronic patient data management systems (PDMS) integrate in relational databases readings from all information sources including among others: demographics and other data elicited upon admission, information on medication and therapies administered, results from laboratory analyses, annotations from the medical staff, and measurements of physiological signals. Having the ICU data in this electronic integrated format offers several advantages: a decrease in the amount of administrative work leading to more time spent on patient care [2]; increased data quality over manual recording [3]; increased patient safety [4]; availability of decision support [5]; and the opportunity of studying them with automatic data analysis techniques.

The analysis of these databases is well suited for data mining, which encompasses machine learning, artificial intelligence and in general automatic statistical analysis techniques [6-8]. Early work on artificial intelligence applied to the ICU domain focused on knowledge-driven techniques [9,10], however over the last years research has shifted towards data-driven methodologies, that is, data mining. Various techniques for mining clinical data have been applied in the ICU setting to address a wide gamut of clinical research questions [11-13].

We believe the time is right to ask whether by now there are some clear indications of an actual added benefit in the use of these rich databases and data mining techniques over routinely used scoring systems, based on for example manually collected admission and demographic data. These benefits could span from local effects, if they impact daily clinical practice, to global effects if they impact overall hospital administrative or financial decisions.

To address this question, we will mainly focus on the usage of these data mining techniques and databases for predictive modeling in the ICU. Several approaches have been followed in recent studies to evaluate the developed techniques, with

authors comparing to existing scoring systems and prognostic models, and in some cases to actual physicians.

We identify two main research directions: First, short and intermediate term predictive modeling, with the ultimate goal of generating intelligent alerts and decision support systems that can be used by the physician on the daily evaluation of the individual ICU patient. Second, long term predictive modeling, which can provide population level prognosis and serve as benchmarking tools to compare medical performance across institutions. In what follows we will look at recent exponents of these two trends of clinical predictive modeling as used in the ICU domain, and suggest directions for future work.

Short and intermediate term predictive modeling

These modeling techniques are concerned with predicting the evolution of the individual patient and have as ultimate goal the early identification of changes in the health state of the patient at the level of minutes, hours or days.

Predictions stem from the analysis of raw signal data generated from the different ICU information sources, where each input to be analyzed is a time-series with its own frequency and range of values. Monitors capture and report these signals at different time resolutions. Some time-varying phenomena such as medication dosage, records of administered therapies or results from laboratory analyses are encoded at a low frequency with one data instance every several hours. Other signals, such as temperature, even though recorded with high frequency (once every minute), are not expected to change quite that rapidly and their slow trends are of most interest.

Monitored signals are known to be susceptible to various types of noise resulting from patient transport, transducer disconnections or movement, and device failure among others. There exist many alarms provided by the monitoring devices, which are typically threshold alarms operating on the raw captured signals, and as such are prone to generation of many false alerts [14]. To address this issue one line of research has focused on intelligent noise and artifact removal from the raw signal measurements. The use of these advanced techniques allows for the filtered signals to be fed back to the standard monitors and thereby dramatically reduce the number of false alarms, while still retaining the sensitivity required to correctly detect clinically relevant events [15]. Even though such dedicated advanced filtering is not commonly used in prediction modeling, most methods do make use of some noise and artifact removal step. Common approaches are applying low pass filters, moving averages or transforming the raw signal to a lower time resolution by aggregating several measurements. Once the signals have been de-noised they can be used as input for predictive modeling.

The vast majority of short and medium term predictive modeling is based on data mining techniques that were developed for classification and regression tasks in general, where the inputs are assumed to be time-independent, and have then been applied to use in the time-series domain. As such they were not designed to deal directly with the temporal aspect of the data and consider instead as mostly independent the consecutive measurements of the signal. Despite of this shortcoming, they have been used successfully in a wide range of prediction tasks in the ICU setting. Some examples include:

- The use of Bayesian networks (BN) for prediction of fluid requirement on day two of ICU stay, as a study of inflammatory response in 3000 patients, which resulted in a predictive accuracy of 78% [16].
- A rule-learning algorithm for prediction of impending physiologic instability in over 12000 ICU patients, resulting in 90% sensitivity and 60% specificity [17].
- Prediction of tacrolimus dosage based on blood concentrations of 50 liver transplantation patients via Support Vector Machines (SVM) and multiple linear regression (MLR) resulting in high accuracy with negligible differences between predicted and observed values [18].
- Prediction 15 and 30 minutes in advance of events of hypotension in a multi-center database of over 260 traumatic brain injured patients via a Bayesian artificial neural network (ANN) resulting in 41% sensitivity and 86% specificity [19].
- Prediction of these hypotension episodes, one or two hours in advance via ANN resulting in an area under the receiver operating characteristic curve (AUC) of 0.92, a 83% sensitivity and 86% specificity [20].
- Prediction of second day ICU discharge after non-emergency cardiac surgery via Gaussian processes (GP) on a validation cohort of 500 patients, resulting in an AUC of 0.76, and demonstrating a significantly better discriminative power than the EuroSCORE and the ICU nurses, and equal performance compared to ICU physicians [21].

Other data mining techniques have been developed with the purpose of dealing with time-series data directly. A recent exponent is the case of prediction in 830 patients of the need for dialysis between days 5 and 10 of ICU stay via echo-state networks, which are a form of recurrent artificial neural networks (ANN) [22]. Performance comparison with regards to sensitivity and specificity do not indicate a clear benefit of this technique over the more routinely used SVMs or BNs, as they all resulted in AUCs of 0.82. They do however show improvements regarding the time needed to build a model and regarding the amount of pre-processing required, with echo-state networks being several orders of magnitude faster.

The reported studies are not meant as an exhaustive exploration of all used techniques but rather as a sample of the diversity and maturity of the field of predictive modeling in ICU. In all cases good predictive results were obtained, and the developed models exhibit discriminative performance above that which is considered to be of use in clinical practice. This would indicate that most of these methods are ready to be translated into intelligent systems to be used in the ICU. However many of these studies have been postulated only as proofs of concept and the step towards implementation in the actual ICU setting remains future work. Limitations that are frequently mentioned for the lack of cases that have taken this step are foremost, the limited number of patients used to develop the models, which in most studies translates to insufficient model validation. In the ideal scenario large databases with respect to the number of patients would be used to develop and internally validate the model while equally large datasets would be used for external validation [23]. This is rarely done in practice and the majority of studies rely only on internal validation techniques on small population samples. Likewise, large emphasis has been placed on models that deliver high accuracies and have high discriminative performance with high AUCs, while other important performance measures like model calibration are

typically overlooked [24]. This aspect of a model's performance is crucial if it is to be used for prediction on individual patients.

The ultimate goal of these models is that they transition into alert or support systems to be used in the ICU. This would allow for follow-up studies to show whether the use of these intelligent systems has a positive impact on ICU patient outcome, by leading to better choice of therapies and treatment. However for the most part this step remains future work, as the number of systems to have been actually deployed for real clinical use is limited. We therefore advocate that whenever possible, successful modeling strategies should carry on further than the studies performed in retrospective analyses of medical databases and into randomized controlled studies to evaluate their effect on clinical practice.

Long term predictive modeling

Models based on demographic and administrative static data are considered golden standards for long term or outcome prediction. The main reason being that they have been developed and validated in very large databases that can go back several decades. Likewise, these data types predate the electronic era, which eases collection costs and feasibility when compared to monitored clinical data.

Nevertheless studies carried out to compare the prognostic value of purely static and demographic data versus that of physiological time-dependant data, show improved performance of the latter. In a large study of over 47000 patients, prediction of survival at 180 days after hospital discharge resulted in an AUC of 0.73 for an administrative-only model while the use of clinical variables improved performance to 0.83 [25]. Similarly a very large mortality prediction study on fifty-five Dutch ICUs and over 66000 patients showed improved performance and robustness of a model based on clinical data versus a model based on administrative data, with AUCs of 0.85 and 0.77 respectively [26].

Several data mining techniques have been used for prediction of ICU mortality. In a study of over 38000 patients data from several information sources during the first 24 hours of ICU stay were used to develop ANN, SVM, decision trees (DT) and conventionally used logistic regression (LR) models, all of which resulted in similar discriminatory performance with AUCs above 0.87. These models had similar performance as the routinely used scoring system APACHE III, albeit requiring less predictive variables [27].

Locally developed models better adjust to the characteristics of the studied population and typically result in more accurate predictions than severity scores developed in large patient populations, even when externally validated. A study on 352 patients with haematological malignancies, resulted in better predictive performance for hospital mortality when using locally built SVM and MLR models (0.80 AUCs) as compared to standard ICU severity of illness scores such as APACHE II and SAPS II (0.62 AUCs) [28]. The advantage of these highly performant locally learned models, is that they are more useful in the clinical setting even though they cannot be transported to other ICUs and are therefore not universally applicable.

The majority of long term ICU prognosis refers to mortality prediction for different risk subpopulations. Other well established models for long term prediction refer to outcome of the traumatic brain injured patient. In this scenario, as discussed, there exist well established golden-standards with which to compare model performance.

Validated prognostic models have been based mainly on admission characteristics, although here again an improvement in performance is observed once clinical data is included in the analysis [29].

Long term prediction outcomes, such as mortality are commonly used for benchmarking purposes, for evaluating the financial and patient care performance of an ICU or hospital as a whole. However, unless models are sufficiently discriminative and well-calibrated to provide accurate predictions for the individual patient they are of little use in daily clinical practice. For such a highly performant model, a difference in prognosis between the model's prediction and the physician's opinion could lead to more in-depth tests to further evaluate the health-state of the patient. These models would also be of value for counseling of relatives and patients, or when deployed in hospitals where there is a general lack of expert clinicians.

Conclusion

Data mining techniques for clinical prediction result in accurate models that are useful in clinical practice, as they often are equivalent to experienced physicians, and typically outperform generalized risk scores. This performance is a direct result of their use of monitored clinical data.

A common finding in studies of short, medium and long term clinical prediction, is the inability of one particular modeling data mining strategy to consistently outperform the rest. The same methodology can rank either amongst the best or the worst performing, depending on the particular population studied. No clear conclusions have emerged as to which methods are better suited for specific data types or prediction tasks. At this stage it is probably best practice to internally validate several variants of different algorithms on a large development cohort and chose the model that best exploits the local characteristics of the population under study, as local models have been found to perform better than generalized ones. However, these models require updating in order to remain relevant [30], and currently there is a general lacking in studies of historic transportability of models to determine their stability with regards to changes in time of the population they describe. An open question is the frequency with which local models need to be re-trained on more recent data samples that better represent the current ICU population.

Even though current data mining techniques that make use of time-series of clinical data are able to outperform traditional techniques based purely on static demographic data, and to achieve similar performance to expert physicians, they remain mostly proof of concept studies. By now it has been abundantly shown that these models excel in predictive performance, and we believe the time is right to translate these modeling techniques to intelligent systems that can be used in clinical practice. The focus should shift towards evaluating the potential benefit of these tools when acting as decision support systems in the ICU, as ultimately, the clinical benefit of these systems can only be properly evaluated through prospective randomized clinical trials.

References:

1. Manjoney R. Clinical information systems market – an insider's view. *J Crit Care* 2004; 19: 215–220.
2. Bosman RJ, Rood E, Oudemans-van Straaten HM. Intensive care information system reduces documentation time of the nurses after cardiothoracic surgery. *Intens Care Med* 2003; 29: 83–90.
3. Roncati Zanier E, Ortolano F, Ghisoni L et al. Intracranial pressure monitoring in intensive care: clinical advantages of a computerized system over manual recording. *Crit Care* 2007; 11: R7 (doi:10.1186/cc5155).
4. Colpaert K, Claus B, Somers A et al. Impact of computerized physician order entry on medication prescription errors in the intensive care unit: a controlled cross-sectional trial. *Crit Care* 2006; 10: R21.
5. Berger MM, Revelly J-P, Wasserfallen J-B et al. Impact of a computerized information system on quality of nutritional support in the ICU. *Nutrition* 2006; 22: 221–229.
6. Mitchell T (ed). *Machine Learning*. McGraw Hill Publishing Company, 1997.
7. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery: An overview. In Fayyad U, Piatetsky-Shapiro G, Smyth P & Uthurusamy (eds) *Advances in Knowledge Discovery and Data Mining*, The MIT press, pp 495–515, 1996.
8. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer-Verlag, 2001.
9. Uckun S. Intelligent system in patient monitoring and therapy management: a survey of research projects. *International Journal of Clinical Monitoring and Computing*. Volume 11, Number 4 (1994), 241-253
10. Hayes-Roth B, Uckun S, Larsson JE, Drakopoulos J, Gaba D, Barr J, Chien J. Guardian: an experimental system for intelligent ICU monitoring. *Proc Annu Symp Comput Appl Med Care*. 1994:1004.
11. Lavrac N. Selected techniques for data mining in medicine. *Artif Intell Med* 1999; 16: 3–23.
12. Lucas P. Bayesian analysis, pattern analysis, and data mining in health care. *Curr Opin Crit Care* 2004; 10: 399–403.
13. Meyfroidt, G., Guiza Grandas, F., Ramon, J., Bruynooghe, M. Machine learning techniques to examine large patient databases. *Best Practice & Research Clinical Anaesthesiology*. 2009. 23 (1), 127-143.
14. Ansermino J, Daniels J, Hewgill R, et al. An evaluation of a novel software tool for detecting changes in physiological monitoring. *Anesth Analg* 2009; 108: 873–880.
15. Borowski, M., Görges, M., Fried, R., Such, O., Wrede, C., Imhoff, M. (2011), *Medical Device Alarms*, *Biomed Tech*, 56, 73-83.
16. Celi LA, Hinske LC, Alterovitz, G, Szolovits P: An artificial intelligence tool to predict fluid requirement in the intensive care unit: a proof-of-concept study. *Crit Care* 2008, 12:R151.
17. L. J. Eshelman, K. P. Lee, J. J. Frassica, W. Zong, L. Nielsen, and M. Saeed, “Development and evaluation of predictive alerts for hemodynamic instability in ICU patients,” *Proceedings of the Annual Symposium proceedings/AMIA Symposium*, pp. 379–383, 2008.
18. Van Looy S, Verplancke T, Benoit D, Hoste E, Van Maele G, De Turck F, Decruyenaere J. A novel approach for prediction of tacrolimus blood

- concentration in liver transplantation patients in the intensive care unit through support vector regression. *Crit Care*. 2007;11(4):R83.
19. Donald R, Howells T, Piper I, Chambers I, Citerio G, Enblad P, Gregson B, Kiening K, Mattern J, Nilsson P, Ragauskas A, Sahuquillo J, Sinnott R, Stell A. Early warning of EUSIG-defined hypotensive events using a Bayesian Artificial Neural Network. *Acta Neurochir Suppl*. 2012;114:39-44.
 20. J. Lee and R. G. Mark, "An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care," *BioMedical Engineering Online*, vol. 9:62, 2010.
 21. Meyfroidt G, Güiza F, Cottem D et al: Computerized prediction of intensive care unit discharge after cardiac surgery: development and validation of a Gaussian processes model. *BMC Med Inform Decis Mak*. 2; 11: 64
 22. Verplancke T, Van Looy S, Steurbaut K, Benoit D, De Turck F, De Moor G, Decruyenaere J. A novel time series analysis approach for prediction of dialysis in critically ill patients using echo-state networks. *BMC Med Inform Decis Mak*. 2010 Jan 21;10:4.
 23. Steyerberg EW: Validation of prediction models. *Clinical prediction models: a practical approach to development, validation and updating* New York, Springer; 2009, 299-311.
 24. Steyerberg EW: Overfitting and optimism in predicting models. *Clinical prediction models: a practical approach to development, validation and updating* New York, Springer; 2009, 83-100.
 25. Bohensky MA, Jolley D, Pilcher DV, Sundararajan V, Evans S, Brand CA. Prognostic models based on administrative data alone inadequately predict the survival outcomes for critically ill patients at 180 days post-hospital discharge. *J Crit Care*. 2012 May 15
 26. Brinkman S, Abu-Hanna A, van der Veen A, de Jonge E, de Keizer NF. A comparison of the performance of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ratios of intensive care units. *Crit Care Med*. 2012 Feb;40(2):373-8.
 27. Kim S, Kim W, Park RW. A Comparison of Intensive Care Unit Mortality Prediction Models through the Use of Data Mining Techniques. *Healthc Inform Res*. 2011 Dec;17(4):232-43.
 28. T Verplancke, S Van Looy, D Benoit, S Vansteelandt, P Depuydt, F. De Turck, J Decruyenaere. Support vector machine versus logistic regression modeling for prediction of hospital mortality in critically ill patients with haematological malignancies. *BMC medical informatics and decision making*. 01/2009; 8(1):56.
 29. Lingsma HF, Roozenbeek B, Steyerberg EW, Murray GD, Maas AI. Early prognosis in Traumatic Brain Injury: from prophecies to predictions. *Lancet Neurology* 2010;9:543-554.
 30. Harrison DA, Brady AR, Parry GJ, et al: Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006; 34:1378-1388.

Conflicts of interest:

All authors declare they have no conflicts of interest.

Acknowledgments:

GM is funded by the Research Foundation - Flanders (FWO) (Senior Clinical Investigator, 1843113N).