Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Developing a Bayesian network as a decision support system for evaluating patient with diabetes mellitus admitted to the intensive care unit – a proof of concepts

Jakobsen, Rune Sejer; Hejlesen, Ole; Stausholm, Mads Nibe; Cichosz, Simon Lebech

Published in:
Proceedings from The 16th Scandinavian Conference on Health Informatics 2018 Aalborg, Denmark August 28–29, 2018

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Jakobsen, R. S., Hejlesen, O., Stausholm, M. N., & Cichosz, S. L. (2018). Developing a Bayesian network as a decision support system for evaluating patient with diabetes mellitus admitted to the intensive care unit – a proof of concepts. In A. Bygholm, L. Pape-Haugaard, K. Niss, O. Hejlesen, & C. Zhou (Eds.), Proceedings from The 16th Scandinavian Conference on Health Informatics 2018 Aalborg, Denmark August 28–29, 2018 (pp. 70-74). Linköping University Electronic Press. http://www.ep.liu.se/ecp/151/ecp18151.pdf

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Developing a Bayesian Network as a Decision Support System for Evaluating Patients with Diabetes Mellitus Admitted to The Intensive Care Unit – a Proof of Concept

Rune Sejer Jakobsen, Ole Hejlesen, Mads Nibe Stausholm, Simon Lebech Cichosz

Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Abstract

Evidence is increasing about an unsatisfying performance from the existing non-disease-specific scoring systems in the intensive care unit (ICU). Evidence is furthermore increasing about differences in the mortality rate between diabetics and non-diabetics dependent on the level of blood glucose (BG), but few scoring systems include these variables in the assessment of the patients. 142,404 ICU admissions were included from the eICU database in the development of an unsupervised trained Bayesian Network (BN). The BN suggested that abnormalities in the level of BG should be associated with differences in the mortality rate between diabetics and non-diabetics. The BN showed promising predictive ability with an AUC on 0.86 for predicting death (sensitivity: 75.06, specificity: 78.40 %). 48.43 % of the length of stays (LOS) were correctly predicted. The results were slightly below the results from the APACHE IV scoring system but showed great ability of risk stratification. The BN showed a potential for predicting the patient outcome and might enable an improved method for risk stratifying the patients admitted to the ICU.

Keywords:

Intensive Care Unit, APACHE IV, Mortality, Diabetes Mellitus, Blood Glucose.

Introduction

The ICU is a preventive and therapeutic care setting that take care of patients who suffers from a very vulnerable and potential reversible failure of the organ systems that is more severe and beyond a treatable limit from a regular bed department [1] [2]. The immune system of the patients at the ICU is often at a worsened state [1]. The ICU includes observation, diagnosis, treatment and caretaking of the patient [2].

In the last century has the ICU been under a rapidly development, including a development of multiple scoring systems [3]. Most of the different scoring systems includes a score – an integer – and a probability model, which predict the mortality rate for the patient [3].

Most of the existing scoring systems are developed to evaluate general disease severity for a mixed patient group in the ICU [4]. Some studies suggest, that the accuracy of the models is sufficient when evaluating a mixed group of critical ill patient admitted to the ICU [5] [6]. On the other hand, is the evidence also increasing about the models performs insufficient related to disease specific complications, such as Cardiac Arrest, H1N1 (influenza A), Acute Respiratory Distress Syndrome and patients requiring respiratory support [7] [8] [9] [10] [11]. This might support the discussion of whether the old models developed for evaluating general disease severity is suited for evaluating specific disease severity in the ICU [4].

A disease-specific patient group, meeting an increased interest in the ICU, is the patients suffering from Diabetes Mellitus (DM), where abnormalities in the level of BG are reported to have different impact on the mortality rate dependent on whether the patient suffers from DM [12] [13]. DM is the seventh most common cause of death in USA and is furthermore related to other comorbidities, such as retinopathy, neuropathy, renal failure, coagulative- and cardiovascular complications [14]. Although, the actual impact from DM, and the underlying mechanisms, on both the severity of the condition and the mortality rate not fully understood [15]. It might therefore be essential develop a model that enables an analysis of correlations between changes in multiple variables for the patients with DM. The disease-specific scoring system for diabetics should additionally be able to analyze the joint probability of DM and the level of BG in the ICU to improve the understanding of how patients suffering from DM proceeds in the ICU compared to non-diabetics.

The objective of this paper was therefore to develop a new model for decision support and risk stratification of patients suffering from DM admitted to the ICU. The model should enable an assessment of the impact from a joint probability of DM and the level of BG and furthermore seek to improve the accuracy of the predictions regarding mortality rate and LOS, compared to the APACHE IV scoring system.

Materials and Methods

The data acquired was collected from 2014-2015 and contains data about critical ill patients admitted to the ICU in

USA. Data was collected in the eICU Collaborative Database. The eICU Collaborative Database included data from > 200,000 patient admissions. The mean age of all the patients was 61.97 years, 45,90 % was female and 54.10 % was male.

Data screening

142,404 ICU admissions were registered with information of whether the patients suffered from DM and an actual mortality rate and an actual LOS. Admissions with missing information regarding DM, mortality rate and LOS were excluded. 108,899 admissions were registered without DM and 33,505 admissions registered with diabetes. There was no discrimination between Type 1 Diabetes Mellitus (T1DM) and Type 2 Diabetes Mellitus (T2DM) in the eICU database. 4.93 % of the patients suffering from DM died in the hospital and 6.17 % of the patients without DM died in the hospital.

Bayesian Network

A BN was used as it enabled an assessment of correlations in data, including analyzing the joint probability of two or more variables in the ICU. BN's used the Bayesian rule to update the probability of each node when evidence e was presented. The Bayesian rule stated:

$$P\,(x|e) = \frac{P(e|x)P(x)}{P(e)} \qquad (1)$$

Which was that the probability of having the disease given symptoms, $P(x|e)$, could be calculated if the probability of having the symptoms, $P(e|x)$, the probability of having the disease, $P(x)$, and the probability of having the symptoms, $P(e)$, were known.

When evidence was presented in the network, would the probabilities be updated dependent on how the connections in the network was modelled. This concept was called probabilistic inference. The probabilistic inference was used for reasoning in the network. The Bayesian rule can be rewritten as [16]:

$$P(x|e) \propto P(e^{c}|x) \times P(x|e^{P}) = \prod_{j=1}^{|C|} P(e_{C_j}|P_{mn})\left[\sum_{all\,P_{mn}} P(x|P_{mn}) \sum_{j=1}^{|C|} P(P_i|e_{p_i})\right] \qquad (2)$$

This indicated, that the probability of having the disease given symptoms, $P(x|e)$, had a proportional dependency on the children nodes, $P(e^{C}|x)$, and the parent nodes, $P(x|e^{P})$. [16] Information could not flow between children nodes in a BN, which instead was written as the product of all the likelihoods, $\prod_{j=1}^{|C|} P(e_{C_j}|\,P_{mn})$. C_J represented the $\boldsymbol{j}$ 'th child note, e_{C_J} represented the value of the probabilities of cardinality it states. P_{mn} represented the value of the state $\boldsymbol{n}$ on the parent node P_m. All priors were summarized and multiplied by their condition probability, $\sum_{all\,P_{mn}} P(x|P_{mn}) \sum_{j=1}^{|C|} P(P_i|e_{p_i})$. The children nodes and the parent nodes were normalized to get exactly $P(x|e)$. [16]

The Bayesian rule and the probabilistic inference were used to let information pass though the network and predict how evidence affect the probabilistic outcome throughout the entire network [16]. This means that a quantification of the condition, a risk stratification and a predicted score for mortality rate and LOS could be calculated dependent on entered evidence e in the BN.

Structure learning

The development of the graphical structure of the BN was performed in Hugin Expert Developer (https://www.hugin.com/). Structure learning was performed unsupervised using the Greedy Search-and-score algorithm provided by Hugin Expert. A maximum of 13 parents were chosen in the Greedy Search-and-score algorithm for the BN and Bayesian Information Criterion (BIC) was chosen as the penalized likelihood criteria. BIC penalizes the complexity of the model and rewards the model if it fits the data well. [17]

Parameter learning

The distributions of data-points between the states in each node – the parameters – were learned by using the Expectation-Maximization (EM) algorithm. The EM algorithm handled missing data very well, making it suited for handling real-life data. The EM-algorithm was furthermore considered as the natural choice of algorithm for parameter learning if the structure learning was performed by the Greedy Search-and-Score algorithm [18]. The EM-algorithm counts the parameter for the BN.

Validation

The holdout method was used for validation, which separated the dataset into a training set and a test set [19]. The data was distributed as 90 % for training and 10 % for test. This method ensures that the model was less likely to be overfitted.

The accuracy of the predictions from the BN was compared with the predictions from the APACHE IV scoring system. Two APACHE IV models were developed for the test: one for mortality and one for LOS. The two APACHE IV models were illustrated in Figure 1.

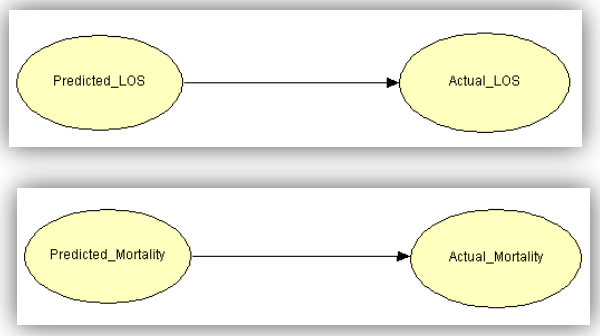

Figure 1 – Presented the two models for testing the accuracy of the predictions from APACHE IV

The same unique patientunitstayID was used for both testing the BN and the APACHE IV scoring system. 128,164 admissions were used for training and 14,240 admissions were used for testing.

Results

Three tests were performed for the BN: (1) One for testing the impact of a joint probability of DM and the level of BG on the mortality rate, (2) one for testing the performance of predicting the mortality rate and (3) One for testing the performance of predicting the LOS.

The impact from DM and level of BG on the mortality rate

Figure 2 illustrates the impact on the mortality rate from a joint probability of the level of BG and respectively diabetics and non-diabetics.

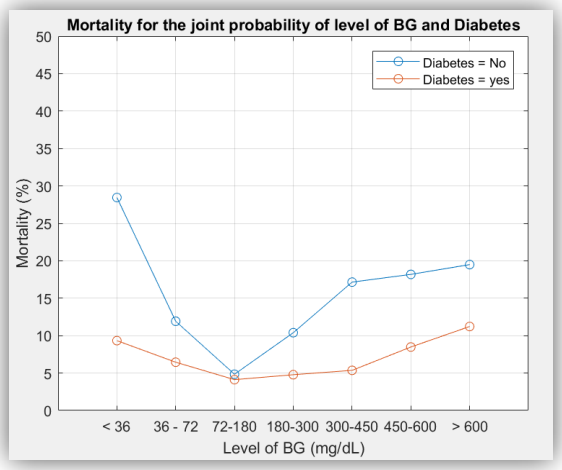

Figure 2 – presented the correlation between mortality rate and the joint probability of the level of BG and diabetes. 72 – 180 mg/dL was associated with a normal level of BG

Performance test for predicting mortality rate

A performance test was also carried out to evaluate the accuracy of the predictions from the BN regarding the mortality rate for both diabetics and non-diabetics, which were also compared with a performance test for the APACHE IV scoring system. The performance test consisted of a ROC-curve with an associated AUC, and a result table which illustrated the relationship from different cut-offs between sensitivity and specificity. The ROC-curve for predicting mortality for both diabetics and non-diabetics were a test of the accuracy of predicting the [mortality: dead]. The ROC-curve was presented in Figure 3:

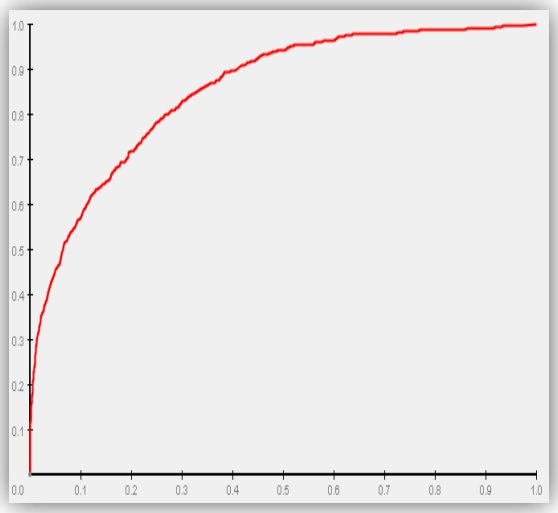

Figure 3 – Presented the ROC-curve of [mortality: dead] with an AUC on 0.85561. The ROC-curve illustrates sensitivity (true positive rate) vs. 1 – specificity (false positive rate)

The ROC-curve had an AUC on 0.86 for the BN developed in this project. In comparison did the test of APACHE IV show an AUC on 0.87 for [Mortality: Dead], which was tested following Figure 1.

Different cut-offs could be chosen for sensitivity and specificity for the BN developed in this project. A moderate sensitivity and a moderate specificity also calculated because a high sensitivity caused a low specificity and a high specificity caused a low sensitivity. The relationship between sensitivity and specificity were presented in a result table, which was presented in Table 1:

Table 1 – Presented the relationship between sensitivity and specificity in terms of number of true positive (No. TP), number of false positive (No. FP), number of true negative (No. TN) and number of false negative (No. FN).

	Sensitivity	Specificity	No. TP	No. FP	No. TN	No. FN
High Sensitivity	100.00 %	0.12 %	12416	823	1	0
High Specificity	2.71 %	100.00 %	364	0	824	13052
Moderate Specificity	75.06 %	78.40 %	10070	178	646	3346

The APACHE IV scoring system had a moderate sensitivity on 75.57 % and a moderate specificity on 81.07 %.

Performance test for predicting LOS

The performance test of predicting LOS consisted of a confusion matrix, where predictions was from a max belief and a

calculation of the percentage of correct predictions. The confusion matrix for LOS was illustrated in Table 4:

Table 2 – Presented the confusion matrix of LOS

< 1. (actual)	**1. – 2. (actual)**	**> 2. (actual)**	
808	608	425	**< 1. (predicted)**
1640	1754	1442	**1. – 2. (predicted)**
1215	2029	4318	**> 2. (predicted)**

48.43 % were correctly predicted from a test of max belief. In comparison did the test of APACHE IV show that 52.54 % was correctly predicted from a test of max belief, which was tested following Figure 1.

Discussion

An AUC > 0.80 was associated with good discrimination capability [5], meaning the accuracy of predicting the mortality rate was considered satisfying for the BN even though the accuracy of the predictions from the BN was slightly below the predictions from the APACHE IV scoring system. The accuracy of predicting the LOS was considered satisfying, because the test was performed as a test of max belief, whereas the actual prediction from the BN was a distribution of likelihoods. The accuracy from the test of max belief of the prediction of LOS was slightly below the predictions from the APACHE IV scoring system.

The literature suggested that the non-diabetics was more vulnerable towards hyperglycemia and hypoglycemia [12, 13] which was also suggested by the BN. The increased mortality for non-diabetics with abnormalities in the level of BG might be explained by the diabetics are more accustomed towards both hyperglycemia and hypoglycemia [15].

It was possible that the mortality would be different for T1DM and T2DM, where e.g. the body mass index of especially T2DM could have an impact on the mortality rate [15]. Hba1c could also be a part of the assessment of the patient, which could explain occurrence of abnormalities in the level of BG prior to ICU admission or even cases of unknown DM [14]. This might affect the degree of how accustomed the patient was to experiencing abnormalities in the level of BG. Intensive insulin treatment could also be within the scope of assessing diabetics in the ICU, as the mortality rate and LOS might be dependent on the intensive insulin treatment [20] and because the intensive insulin treatment was strongly associated with normal DM caretaking [21, 22].

The Greedy Search-and-score algorithm was used for structure learning, but other algorithms exist for unsupervised structure learning. The PC-approach makes edges between all nodes with no conditional independence in the BN [18], meaning a very complex BN will be generated with very high number of CPT's, which would be practically impossible to compile for a BN with many nodes and states. This could be solved by inspecting the model and deleting some of the edges, which might hold the potential of ensuring all edges associated with domain knowledge were included in the model. It would be interesting for a future study to evaluate whether the PC-approach could improve the accuracy of the predictions, as an alternative structure of the BN could be produced [18]. The Rebane-Pearl might hold the potential for improving the transparency of the model, where edges were directed away from a root where one node could only have one parent (except for the root-node) [18]. The Rebane-Pearl algorithm was not used for unsupervised structure learning, as one patient could have more diagnoses and one variable for APS was associated with other APS variables for the same organ system, meaning one node was expected to have multipleparents.

The Greedy Search-and-Score algorithm was considered suited for weighting the accuracy and transparency, where an inspection of the generated structure for a disease-specific model for DM showed promising edges. The Greedy-search-and-Score did also show a promising accuracy of the predictions, meaning the algorithm was considered as a promising approach for performing unsupervised structure learning.

Conclusion

The output of this paper was a BN developed through unsupervised structure learning (by the Greedy Search-and-score algorithm) and parameter learning (by the EM algorithm). The BN was able to quantify the severity of the condition, predict the outcome for the patient and potentially improve the risk stratification of the patient admitted to the ICU. The BN could possibly be a means for decision support for the clinician when evaluating patients suffering from DM admitted to the ICU.
The results showed high difference in the likelihood for mortality rate for the joint probability of DM and the level of BG.
Both the BN and the APACHE IV scoring system showed high accuracy in predicting the mortality rate and relatively poor accuracy of predicting the LOS, though predicting LOS for both the BN and the APACHE IV scoring system were better than just guessing. The accuracy of the predictions of mortality rate and LOS from the BN was slightly below the accuracy of the predictions from the APACHE IV scoring system. The BN was associated with an improved and alternative method for risk stratification where the BN was able to present changes in the likelihood from correlations between evidence in multiple variables for both diabetics and non-diabetics in the ICU.

Acknowledgement

The authors would like to thank the Hugin Expert team (https://www.hugin.com) for allowing us to use their software for developing the Bayesian Network.

References

[1] Last JM. A Dictionary of Public Health. Oxford University Press. 2007.

[2] Dansk Intensiv Database. Årsrapport. 2016.

[3] Rapsang AG, and Shyam DC. Scoring systems in the intensive care unit: A Compendium. 2014.

[4] Vincent JL, and Moreno R. Clinical review: Scoring systems in the critically ill. 2010.

[5] Zimmerman JE, Kramer AA, McNair DS, and Malila FM
Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically illpatients. 2006.

[6] Naqvi IH, Mahmood K, Ziaullaha S, Kashif SM, and Sharif A. Better prognostic marker in ICU - APACHE II, SOFA or SAPS II!. 2016.

[7] Enfield K, Miller R, Rice T, Thompson BT, and Truwit J.
Limited Utility of SOFA and APACHE II Prediction Models for ICU Triage in Pandemic Influenza. 2011.

[8] Saleh A, Ahmed M, and Mansoura AA.Comparison of the Mortality Prediction of Different ICU Scoring Systems (APACHE II and III, SAPS II, and SOFA) in Acute Respiratory Distress Syndrome Patients. 2016.

[9] Shahpori R, Stelfox HT, Doig CJ, and Boiteau PJE. Sequential Organ Failure Assessment in H1N1 pandemic planning. 2011.

[10] Bisbal M, Jouve E, Papazian L, Bourmont SD, Perrin G, Eon B, and Gainnier M. Effectiveness of SAPS III to predict hospital mortality for post-cardiac arrest patients. 2014.

[11] Rojek-Jarmuła A, Hombach R, and Krzych ŁJ. APACHE II score cannot predict successful weaning from prolonged mechanical ventilation. 2017.

[12] Tayek CJ, and Tayek JA. Diabetes patients and non-diabetic patients intensive care unit and hospital mortality risks associated with sepsis. 2012.

[13] Cichosz SL, and Schaarup C. Hyperglycemia as a Predictor for Adverse Outcome in ICU Patients With and Without Diabetes. 2017.

[14] Carpenter DL, Gregg SR, Xu K, Buchman TG and C. M. Coopersmith. Prevalence and Impact of Unknown Diabetes in the ICU. 2015.

[15] Siegelaar SE, Devries JH, and Hoekstra JB. Patients with diabetes in the intensive care unit; not served by treatment, yet protected?. 2010.

[16] Duda RO, Hart PE, and Stork DG. Pattern Classification. 2001.

[17] Cho SJ, and Boeck PD. A Note on N in Bayesian Information Criterion for Item Response Models. 2018.

[18] Hugin Expert. Hugin Expert – Manual. 2017.

[19] Schneider J. Cross Validation. 1997.

[20] Barski L, Kezerle L, Zeller L, Zektser M, and Jotkowitz A. New approaches to the use of insulin in patients with diabetic ketoacidosis. 2013.

[21] Almdal T, Kjeldsen HC, Vestergaard H, and Sachs C. Sundhed.dk - Type 1 diabetes. 2016.

[22] Almdal T, Kjeldsen HC, Sanbæk A, Vestergaard H, and Sachs C. Sundhed.dk - Type 2 diabetes. 2015.