Proceedings of the 32nd European Safety and Reliability Conference (ESREL 2022)
Edited by Maria Chiara Leva, Edoardo Patelli, Luca Podofillini, and Simon Wilson
 Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-5183-4_R12-12-372-cd

A data driven approach to elicit causal links between performance shaping factors and human failure events

Karl Johnson

Civil & Environmental Engineering, University of Strathclyde, United Kingdom. E-mail: karl.johnson@strath.ac.uk

Caroline Morais

Agency for Petroleum, Natural Gas and Biofuels (ANP), Brazil. E-mail: cmorais@anp.gov.br

Lesley Walls

Management Science, University of Strathclyde, United Kingdom. E-mail: lesley.walls@strath.ac.uk

Edoardo Patelli

Civil & Environmental Engineering, University of Strathclyde, United Kingdom. E-mail: edoardo.patelli @strath.ac.uk

Within the field of human reliability analysis (HRA), there is an acknowledged demand to move further towards data driven models. There have been several independent research projects focused on gathering the required empirical data, to support existing theoretical models used in HRA, as well as allow the use of probabilistic tools, such as Bayesian Networks, to model such data. However, with regards to Bayesian Networks, there is a reliance upon expert elicitation to design the structure of the network, that is, the causal links between the considered factors are determined by an expert, with the data used only to estimate the conditional probability tables.

This work aims to provide a methodology/framework to elicit causal links between performance shaping factors from data, producing a HRA model constructed entirely from data, with the ability to integrate the knowledge provided by experts. The Multi-Attribute Technological Accidents Dataset (MATA-D) has been used as the data source, therefore the model is produced under a framework based on the Cognitive Reliability and Error Analysis Method (CREAM). This model is produced through a combination of information theory and structure learning algorithms for Bayesian Networks.

The proposed model/methodology aims to support experts in their evaluation of human error probability, and reveal causal links between performance shaping factors, that may not have otherwise been considered.

Keywords: Bayesian networks, human factors, human reliability analysis, structure learning, CREAM.

1. Introduction

Various human reliability analysis (HRA) methods have been developed to aid in the incorporation of the human contribution to risk into overall system safety analysis. The methods take both qualitative and quantitative approaches to identify and analyze the causes, consequences and likelihood of human actions and decisions. Of these methods some propose guidance for the assessment of a human error probability (HEP), which is based on an operator's actions, the opportunities for error, the occurrence of errors, as well as incorporating factors that may impact performance. (Boring and Blackman 2007) Such factors are referred to as Performance shaping factors (PSFs), with positive or negative influence on human performance these include organizational, technological, and personal factors (Morais, Yung, et al. 2022). Some examples of PSFs that are often observed include, experience, stress, complexity of task, adequacy of procedure, workload, etc., (Park, Jung and Kim 2020) the identification and quantification of the effects of these factors is a key step in the process of various HRA approaches (Griffith and Mahadevan 2011). Such factors are interdependent, therefore there is also the need to consider the interrelationships among these factors. (Groth and Mosleh 2012)

This has led to the need for casual models as these explicitly capture the factors that influence human performance and their interdependencies. These models enable not only the calculation of HEP, but also explain

why human errors occur and how we can help prevent them. (Groth and Mosleh 2012) A popular choice when producing these models has been through the use of Bayesian networks, their graphical structure allows the causal links between PSFs and events, to be easily recognized by those not involved directly in the models building. The level of influence these factors have on each other can also be simply extracted from the conditional probability tables included in the model.

The use of Bayesian not only addresses this challenge within HRA, but also aids in the demand to move further towards data driven models. Various models have been proposed where analysts have been able to combine information from several sources, including empirical data and expert opinion. Despite these steps forward, the use of empirical data has primarily been reserved for the estimation of the conditional probabilities. There is a reliance on expert opinion to identify the causal links between the PSFs, and therefore determine the structure of such models. Due to this there maybe some potential causal links between factors not considered due to various types of bias.

This paper, through the use of structure learning algorithms for Bayesian networks, proposes a method to elicit the causal links between performance shaping factors from data. Which in turn will allow the construction of a HRA model entirely from data, with the ability to integrate the knowledge provided by experts.

2. Theoretical Background

This section will include a more detailed explanation of some of the terms mentioned before, as well as outline the structure learning algorithms used in this work.

2.1 Bayesian networks

Bayesian networks are a type of probabilistic graphical model composed of a directed acyclic graph (DAG), and a set of probability statements. The directed acyclic graph consists of nodes (the variables or factors) and directed arcs (causal influence between nodes). The arcs represent conditional dependencies; arcs are added between nodes to indicate that one node directly influences the other. When an arc does not exist between two nodes, it does not necessarily mean they are completely independent, as they may be connected via other nodes. The probability statements are typically given as conditional probability tables (CPTs) to encode the strength of relationships between the dependent variables (Friedman, Goldszmidt and Wyner 1999).

2.2 Structure learning

As previously discussed the majority of applications of Bayesian Networks within HRA methods still rely on expert opinion for determining the structure of the model itself. There are many different approaches to structure learning in use in other fields, however few examples of even partial use of these are found within HRA. There are three classes of algorithms typically used to learn the structure of Bayesian networks from data. These are constraint-based algorithms, which use conditional independence tests to learn the dependence structure of the data, score-based algorithms, which use goodness-of-fit scores as objective functions to maximise; and then hybrid algorithms that combine aspects of both approaches. (Scutari_, Graafland and Gutiérrez 2019)

2.2.1 Constraint-based algorithms

Constraint-based algorithms identify conditional independence constraints with statistical tests, and link nodes that are found to not be independent. An example of one of the better established versions of this approach is the PC-Stable (Scutari_, Graafland and Gutiérrez 2019). The first two steps of the approach identify which pairs of variables (X_i, X_j) are connected by an arc, without direction. These are variables that cannot be separated by any subset of the other variables, this is tested heuristically by performing a series of conditional independence tests, Test(X_i, X_j | S; D) with increasingly large separating sets S. The next step identifies the v-structures, for each triplet $X_i - X_k - X_j$ such that X_i is not adjacent to X_j and that $X_k \notin S_{XiXj}$, replace it with the v-structure $X_i \rightarrow X_k \leftarrow X_j$. The final step is setting the remaining arc directions by applying recursively the following two rules:

i. If X_i is adjacent to X_j and there is a strictly directed path from X_i to X_j then replace $X_i - X_j$ with $X_i \rightarrow X_j$ (to avoid introducing cycles);
ii. if X_i and Xj are not adjacent but $X_i \rightarrow X_k$ and $X_k - X_j$, then replace the latter with $X_k \rightarrow X_j$ (to avoid introducing new v-structures).

In this work the NPC algorithm as the constraint-based algorithm has been used. The basic process of the PC and the NPC algorithm is the same, the NPC algorithm includes a criterion (called the necessary path condition) which addresses issues faced by the PC algorithm on limited datasets (Steck 2001).

2.2.1 Score-based algorithms

Scored-based algorithms make use of general-purpose optimisation techniques to Bayesian Network structure learning. Each candidate graph is given a network score reflecting its goodness of fit, which the algorithm then attempts to maximise (Scutari_, Graafland and Gutiérrez 2019). In this paper the K2 algorithm is presented as a score based approach. This method requires an initial node ordering (and a maximum number of parent nodes) input which significantly reduces the computational complexity, however does have a major impact on the final structure. Firstly the candidate parents for the node X_i is set as the empty set, then considering each node X_j according to the sequence specified in the initial node ordering ($j < i$), and greedily adding parent nodes to the parents set if it maximizes the score, finally stopping the algorithm when the maximum number of parents is reached, there are no more parents to add, or there is no parent addition that may improve the score (Benmohamed E. 2020).

To determine the initial ordering input some ideas from information theory are borrowed as proposed by Benmohamed et al (Benmohamed E. 2020). The mutual information for each conditional relationship is calculated $I(X,Y)=H(X)-H(X|Y)$, where $H(X)$ is the entropy of X, and $H(X|Y)$ represents the conditional entropy of X given Y. Pairwise triangular-structures are tested, for each nodes X and Y, it is verified if they form with another node Z a cycle. By the directivity property, the amount of mutual information between the input messages and the output messages is likely to become smaller once the exchanged message has gone through multi-level treating, which is formulated as

$$I(X;Y) \leq I(X;Z)+I(Z;Y)$$

After some manipulation and application of rules it leads to the below, which if satisfied allows us to say there is a dependency relationship between X and Y.

$$(I(X;Y) > I(X;Z)) \text{ v } (I(X;Y) > I(Z;Y))$$

A dependency matrix, D, can now be constructed, which is a matrix where 1 if variables i and j are dependent and 0 otherwise. As MI is symmetric we cannot determine if X or Y is the parent in the relationship. The conditional relative average entropy is then compared for each dependency relationship to determine which is the parent node.

$$CR(Y,X) = H(X|Y)/ (H(X)*|X|),$$

This is then compiled into an input order that satisfies this most.

3. Methodology

3.1 Testing of Algorithms

Having determined which structure learning algorithms to implement, there was a need to test the success of these on various sized datasets. To do this, first a network was created. From this network three sample dataset were created with 50, 150 and 1500 entries respectively. This dataset was generated using the process presented by Oehm (Oehm D. 2015).The NPC algorithm and K2 algorithm were then tested on each dataset, if an arc exists between a parent and child node is a binary problem, where we have 1 if a node exists from parent node A to child node B, and 0 if ones does not. Therefore, using the following,

- true positives (TP) occur when the true value is 1 and the model correctly predicts 1
- false negatives (FN) occur if the true value is 1 but the model wrongly predicts 0
- true negatives (TN) occur when true value is 0 and the model correctly predicts 0
- false positives (FP) occur when true value should be 0 but the model predicts 1. (Morais, Yung, et al. 2022)

From true and false predictions there are four metrics used to assess the algorithms performance: *accuracy*, *precision*, *recall*, and

F-measures score (Goh 2017). *Accuracy* is the fraction of correctly predicted data points out of all predictions and defined as follows:

$$Accuracy = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN})} \quad (1)$$

Precision is a good metric to indicate the proportion of positive identifications that are actually correct, or to monitor when the cost of a false positive is high (Ping Shun 2018) and is defined as follows:

$$Precision = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FP})} \quad (2)$$

Recall is a good measure to indicate the proportion of actual positives are identified correctly (Ping Shun 2018), defined as

$$Recall = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FN})} \quad (3)$$

F_1 score is useful if a balance between *precision* and *recall* is needed (Ping Shun 2018), defined as,

$$F_1 = 2 \cdot \frac{(\mathrm{Precision} \cdot \mathrm{Recall})}{(\mathrm{Precision} + \mathrm{Recall})} \quad (4)$$

	Metrics	50 samples	200 samples	2000 samples
NPC Algorithm	Accuracy	73%	86%	91%
	Precision	7%	52%	55%
	Recall	14%	72%	77%
	F_1 Score	9%	60%	64%
K2 Algorithm	Accuracy	77%	83%	87%
	Precision	14%	52%	55%
	Recall	15%	72%	74%
	F_1 Score	15%	60%	63%

Table 1 Performance metric on test sets

For this test network, the algorithms performed bed for all metrics with the 2000 samples. However, the algorithms also performed well with 200 samples, which is approximately how many samples the human reliability dataset used in this work. The algorithms performed better in the recall metric than precision, this means that the algorithm is more likely to produce a false positive than a false negative. Within an application of this approach within human reliability, expert judgement can then be used to prune the network, checking and removing arcs that are considered 'false positives'.

3.2 *Dataset*

The results presented in section 4 have been produced using the Multi-Attribute Technological Accidents Dataset (MATA-D) (Moura, et al. 2016). This dataset is a collection of 238 major accident reports from a range of different industries, which were then classified using the CREAM (Cognitive Reliability and Error Analysis Method) framework. The data is originally classified into 53 factors, which are further breakdowns of the following 15 categories; Action, Observation, Interpretation, Planning, Temporary Person Related Functions, Permanent Person Related Functions, Equipment, Procedures, Temporary Interface, Permanent Interface, Communication, Organisation, Training, Ambient Conditions, Working Conditions. Within this work, it has been decided that the data will be used in the grouped format, there are two significant reasons for this. Firstly, in the ungrouped dataset is dominated by negative results (that is cases where the factor was not identified to have influence the accident), which would have a significant impact on the dataset. Secondly, as this work is meant to demonstrate the usability and potential of this approach, it was decided that the a network with 16 nodes (15 factors, and a failure event) would be easier to review and evaluate.

3.4 Implementation

All computational work was carried out using Matalab software, the data from the MATA-Dataset was manipulated and then extracted from the original source in Excel. The NPC algorithm used in this work is based upon work initially carried out by Guangdi Li (Li 2009).

4. Results

4.1 Expert's Network

In Hollnagel's book regarding the Cognitive Reliability and Error Analysis method (CREAM) (Hollnagel 1998), relationships between the factors/classification groups were proposed. It was suggested that this could be achieved by noting that to each consequent described by a classification group must correspond to one or more antecedents, from (an)other classification group(s). An (incomplete) example table was also proposed that summarised the relationships between the

antecedents and consequents (Hollnagel 1998). Entries within the table show (forward) links, the columns describe the antecedents with the factors listed in the top row while the rows describe the consequents with the factors listed in the left column. This table was then completed by Morais (Morais, Estrada-Lugo, et al. 2022) using the classification scheme provided by Hollnagel (Hollnagel 1998). This table provides an expert's opinion on the links between the 53 performance shaping factors proposed with CREAM, from this these factors can be grouped as stated before.

	Grouped Antecedents				
Grouped Consequents	Human Action	Observation	Interpretation	Planning	...
Human Action	1	1	1	1	⋯
Observation	0	1	1	1	⋯
Interpretation	0	1	1	0	⋯
Planning	0	0	1	1	⋯
Temporary Person Related Functions	0	0	0	0	⋯
...	...	...	...	...	...

Table 2 Example of antecedents-consequents table

From this table, it is the possible to produce a (Bayesian) network to graphically display the relationships. Due to the grouping of the factors and the original classification scheme, the produced network would include some cycles. These were links in both directions from *Observation* to *Interpretation*, and between *Temporary Person Related Functions* and *Commnication*, there was also a cycle created between *Permanent Interface*, *Communication* and *Working Conditions*. To provide an example of a Bayesian Network produced by expert opinion the forward link from *Observation* to *Interpretation*, from *Temporary Person Related Functions* to *Commnication*, and from *Working Conditions* to *Permanent Inteface*, were chosen to be removed. This gives rise to the network presented in *Figure 1*. This network will be used as a comparison with the learnt networks, and will later be aggregated with the learnt networks to fill in any gaps caused by lack of data with expert opinion.

For the learnt networks the same 15 (grouped) performance shaping factors, as well as 'failure event' have been used as the nodes for the Bayesian Networks. From the expert opinion's network it can be seen that each of the factors links to Human Action, either directly or via another node, however within MATA-Dataset not ever accident/failure event had a human action attributed as a factor, therefore an additional final 'failure event' node was needed.

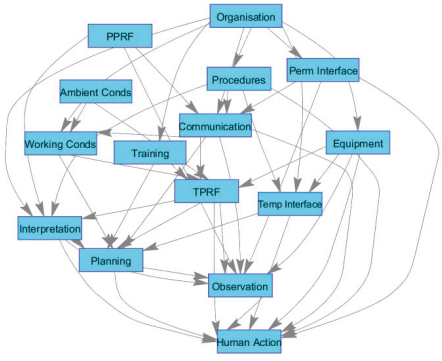

Figure 1 Bayesian Network based on Hollnage'sl classification scheme

4.2 Learnt Networks

The K2 algorithm produced the network shown in *Figure 2*, the inputs for this were the node (Equipment, Procedures, Organisation, Training, Communication, Observation, Interpretation, Human Action, Temporary Interface, Planning, Temporary Person Related Functions, Working Conditions, Permanent Person Related Functions, Ambient Conditions, Permanent Interface, Failure Event) which was determined via the use of information theory as stated before, and a maximum of three parents for each node.

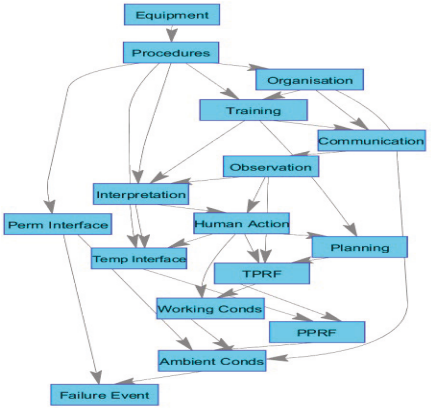

Figure 2 Network learnt via K2 algorithm

The network shown in Figure 3 was produced using the NPC algorithm, with a 5% significance test.

It can be noted that there are significantly less arcs present in the learnt networks than the one based upon the expert's classification scheme. For the K2 the number of arcs has been restricted by the maximum of three parent nodes, and for the NPC algorithm increasing the significance threshold increases the complexity of the network. These restrictions not only meant the arcs given in the network are those with more evidence in the dataset, but it also presents an example of the networks that can be simply reviewed/evaluated. The structure is expected in many regards, factors such as equipment, procedures, organization and training appear towards the top of the hierarchal structure of the network. These are factors that can be generally considered to have an impact throughout a process. Both networks also display a chain of links between training, communication, observation and human action, which is a generally accepted chain of influence (Sasaki N. 2017). From the learnt network there are also several links not present in the network based on the expert's classification scheme, these are links where some further discussion and/or research are warranted as their presence in the structure here suggests there is evidence within the dataset. It is suggested that these causal relationships require investigation and further data gathering efforts to determine their value.

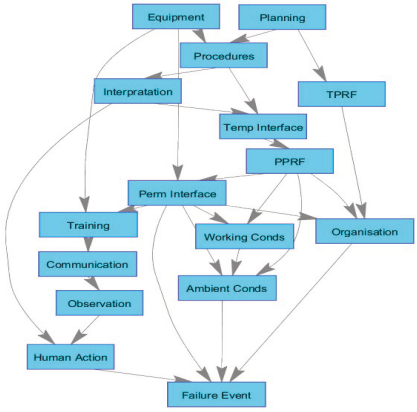

Figure 3 Network learnt via NPC algorithm

4.4 Aggregated Network

To combine the information learnt from both structure learning approaches, the produced networks can aggregated together, the arcs appearing in both of the learnt networks can be considered to be part of the core structure (Acid S. 2004). The network can then be expanded through either the inclusion of the other arcs present in the two learnt structures, or from the network produced from the expert classification. Presented below in *Figure 4* is a complete network incorporating both learnt networks and the experts network. The 'core' arcs are retained with the addition of any arcs that are present in one of the learnt networks and the expert opinion network, as a demonstration of how expert knowledge can be simply integrated into this process. The network here is complete, with all chains leading to 'Failure event'.

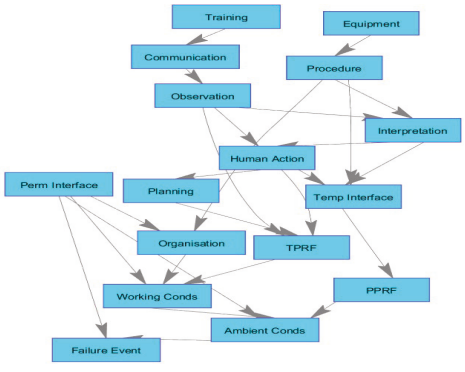

Figure 4 Aggregated Network

5. Conclusion

This work demonstrates the possibility to learn the structure of a Bayesian network from human reliability data, and in turn elicit the causal links between performance shaping factors. This can be considered a step away from reliance upon expert opinion, whilst proceeding with a modelling format that easily allows the integration of their knowledge. Further efforts to expand the MATA-Dataset, and integrate other sources of empirical data are necessary to further improve confidence in the presented models, and increase opportunities for learning. The MATA-D currently contains 238 accident reports, and their classification via the CREAM scheme, as presented in the testing of the algorithms there is improved accuracy as the number of inputs increases. The expansion of this will also allow the expansion of the presented networks from the 15 grouped nodes to the 53 individual factors. A significant amount of information is lost by grouping the factors in this way, and some of the unexpected arcs in the learnt networks, may follow a more expected logic if broken down into their individual factors.

Combining the structure learning methods presented here, and the processes followed in many other models for determining the conditional probability tables, would allow the development of a human reliability model entirely from data. This would limit confirmation bias and financial bias that may have been present in previous approaches, leading to a data supported estimation of human error probability, and the identification of the key factors that need to be targeted to reduce this.

Acknowledgement

This work was partially supported by the EPSRC grant EP/T517938/1.

References

Acid S., de Campos L.M., Fernández-Luna J.M., Rodríguez S., Rodriguez J.M., Salcedo J.l. 2004. "A comparison of learning algorithms for Bayesian networks: A case study based on data from an emergency medical service." *Artificial Intelligence in Medecine* 30 (3): 215-232.

Benmohamed E., Ltifi H., Ben Ayed M. 2020. *Journal of King Saud University - Computer and Information Sciences* 34 (3).

Boring, Ronald Laurids, and Harold S Blackman. 2007. "The origins of the SPAR-H method's performance shaping factor multipliers." *2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting* (Institute of Electronics and Electrical Engineers) 177-184.

Constantinou, Anthony Costa , Norman Fenton, and Martin Neil . 2016. "Integrating Expert Knowledge with Data in Bayesian Networks: Preserving Data-Driven Expectations when the Expert Variables Remain Unobserved." *Expert Systems with Applications* 56: 197-208.

Cussens, James. 2011. "Bayesian network learning with cutting planes." *Twenty-Seventh Conference on Uncertainty in Artificial Intelligence* 152-160.

Dellaert, Frank. 2002. "The Expectation Maximization Algorithm."

Friedman, Nir, Moises Goldszmidt, and Abraham Wyner. 1999. "Data Analysis with Bayesian Networks: A Bootstrap Approach." Proc Fifteenth Conf on Uncertainty in Artificial Intelligence (UAI).

Goh, Y.M., Ubeynarayana, C.U. 2017. "Goh, Y.M., Ubeynarayana, C.U.Construction accident narrative classification: An evaluation of text mining techniques. ." *Accident Analysis & Prevention* 108: 122-130.

Griffith, Candice, and Sankaran Mahadevan. 2011. "Inclusion of fatigue effects in human reliability analysis." *Reliability Engineering and Sytem Safety* 96 (1437-1447).

Groth, Katrina M., and Ali Mosleh. 2012. "Deriving causal Bayesian networks from human reliability analysis data: A methodology and example model." *Journal of Risk and Reliability* 226 (4): 361-379.

Groth, Katrina, and Laura Swiler. 2013. "Bridging the gap between HRA research and HRA practice: A Bayesian network version of SPAR-H." *Reiliability Engineering and System Safety* 115: 33-42.

Hollnagel, E. 1998. *Cognitive Reliability and Error Analysis Method (CREAM).* Oxford: Elsevier Science Ltd.

Li, G. 2009. "NPC algorithm for learning DAG in Bayesian network (https://www.mathworks.com/matlabcentral/fileexchange/24456-npc-algorithm-for-learning-dag-in-bayesian-network)."

Morais, Caroline, Hector Diego Estrada-Lugo, Silvia Tolo, Tiago Jacques, Raphael Moura, Michael Beer, and Edoard Patelli. 2022. "Robust data-driven human reliability analysis using credal networks." *Reliability Engineering and System Safety* 218.

Morais, Caroline, Kai Lai Yung, Karl Johnson, Raphael Moura, Michael Beer, and Edoardo Patelli. 2022. "Identification of human errors and influencing factors: A machine learning approach." *Safety Science* 146.

Moura, Raphael, Beer Michael, Edoardo Patelli, John Lewis, and Franz Knoll. 2016. "Learning from major accidents to improve system design." *Safety Science* 84: 37-45.

Oehm D. 2015. "Simulating data with Bayesian networks." 15 October.

Park, Jooyoung, Wondea Jung, and Jonghyun Kim. 2020. "Inter-relationships between performance shaping factors for human reliability analysis of nuclear power plants." *Nuclear Engineering and Technology* 52 (1): 87-100.

Peng-Cheng, Li, Chen Guo-hua, Dai Li-cao, and Zhang Li. 2012. "A fuzzy Bayesian network approach to improve the quantification of organizational influences in HRA frameworks." *Safety Science* 50 (7): 1669-1583.

Ping Shun, K. 2018. *Accuracy, Precision, Recall or F1? Towards Data Science.* https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9.

Sasaki N., Somemura H., Nakamura S., Yamamoto M.P., Shinmei I., Horikoshi M., Tanaka K. 2017. "Effects of Brief Communication Skills Training for Workers Based on the Principles of Cognitive Behavioral Therapy." *Journal of Occupational and Environmental Medicine* 59 (1): 61-66.

Scutari, Marco, Vitolo Claudia, and Tucker Allan. 2019. "Learning Bayesian networks from big data with greedy search: computational complexity and efficient implementation." *Statistics and Computing* 29: 1095–1108.

Scutari_, M., C.E. Graafland, and J.M. Gutiérrez. 2019. "Who Learns Better Bayesian Network Structures: Accuracy and Speed of Structure Learning Algorithms." *International Journal of Approximate Reasoning* 115: 235-253.

Steck, H. 2001. "Constrained-Based Structural Learning in Bayesian Networks Using Finite Data Sets." Munich.

Suzuki, Joe. 2017. "An Efficient Bayesian Network Structure Learning Strategy." *New Generation Computing* 35: 105–124.

Uusitalo, Laura. 2007. "Advantages and challenges of Bayesian networks in environmental modelling." *Ecological Modelling* 203 (3-4): 312-318.