IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Bayesian Networks Modeling Using Partial Least Squares Approach to Predict Stroke Disease

To cite this article: Dedi Trisnawarman *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **508** 012121

View the article online for updates and enhancements.

Bayesian Networks Modeling Using Partial Least Squares Approach to Predict Stroke Disease

Dedi Trisnawarman[1], Tris Sutrisno[2], Novario Jaya Perdana[3], Sugesti[4]
[1]Information Systems Department, Faculty of Information Technology
Universitas Tarumanagara, Indonesia
[2,3]Informatics Engineering Department, Faculty of Information Technology
Universitas Tarumanagara, Indonesia
[4]STMIK Dharma Putra Tangerang, Indonesia

[1]dedit@fti.untar.ac.id, [2]tris@fti.untar.ac.id, [3]novariojp@fti.untar.ac.id, [4]gesti74@gmail.com

Abstract. Stroke is one of the leading causes of death in Indonesia. However, the incidence of stroke continues to increase. It could affect not only old people but also young people. In Indonesia, it is estimated that 500,000 people suffer from stroke every year, and about 25% or 125,000 people die and the rest suffer minor or severe defects. This study aims to build a decision-making model to help diagnose the possibility of stroke attacks. The model combines Bayesian Networks (BNs) and Partial Least Square (PLS) methods for predicting the attack on suspected patients. The model have been tested using PLS-PM approach. Hospital medical record was used as the testing data with the help of expert verification. The results showed that the model structure of BNs built on expert assumptions can be tested using the PLS approach, and stroke disease can be predicted using interrelated indicators in the model structure of BNs.

Keywords: Bayesian Networks, PLS, Prediction, Stroke

1. Introduction

Stroke is one of the leading causes of death in Indonesia, yet, the incidence of stroke continues to increase. In most cases, people often ignore to control the high blood pressure or hypertension, which is one of the risk factors that dominates the incidence of stroke. Stroke rate worldwide is estimated at 200 per 100,000 populations, in a year. As of now, stroke is on the third rank of the world's most deadly disease after heart attack and cancer. Meanwhile, in Indonesia, stroke ranks first as the cause of deaths in hospital [1].

Based on the results of National Standard Health Research (*Riskesdas*) in 2013, the prevalence of stroke in Indonesia increased by age. The highest stroke cases diagnosed by health-care workers is from people aged 75 years and over (43.1%) and the lowest were within age group of 15 and 24 years old which is only 0.2% [2]. The stroke attack cases were found more on men (7.1%) rather than women (6.8%). Furthermore, the prevalence of stroke in urban areas is higher (8.2%) compared to rural areas (5.7%).

The development of information technology and science, especially in the field of computing, allows rapid progress in creating tools for decision making. Likewise in the medical field, an easy to use independently tool is needed to help diagnose someone having a stroke. In this research, Bayesian networks (BNs) is used to find solutions for several categories of diagnosis. This method provides a compact probabilistic framework for building probability theory to support reasoning under conditions of uncertainty. This method is ideal for domains that have uncertainties such as making medical decisions. For example, to assess mammography in breast cancer, the patient's alarm system indicated the emergence of abdominal pain and using MRI to diagnose liver lesions. [3], to predict stroke patients based on risk factors for heart disease [4].

BNs is a simple Probabilistic Graphical Model (PGM) built from probabilistic theory and graph theory. The probabilistic theory is directly related to data while graph theory is directly related to the form of representation that wants to be obtained [5]. For example, a BNs can represent a probabilistic relationship between disease and symptoms. BNs can be used to calculate the probability of the presence of various symptoms of the disease.

Modelling and computation can be used to predict something that has not yet happened based on historical data that has existed before. Therefore stroke can be predicted using modelling and computation. It is necessary to build a prediction model using computing to predict stroke attack on patients.

2. Literature review

2.1. Bayesian Network for clinical decision making

BNs models has been widely used in the medical domain. The application of the BNs model in clinical decision support systems first developed in the late 1980s [6]. Since then, there have been hundreds of publications about the BN model for the medical field.

Researchers have been built models to help diagnose a disease using BNs [7, 8, 9, 10, 11, 12, 13, 14]. For example, the BNs model has been developed to diagnose pyloric stenosis [7], and pneumonia [14]. The application of BNs to diagnose and selecting the procedures for patients suspected of having gallbladder disease described in [15]. Kline et al., [16] also applied the BNs model to predict the pre-test probability venous thromboembolism. Lukas et al. [17] described a decision-theoretic model for the management of non-Hodgkin's gastric lymphoma (NHL). The BNs model has also been developed in predicting disease risks [16, 18,] and the risks of specific medical outcomes [19, 20]. Other clinical applications of BNs includes models for monitoring patients in intensive care [11], radiology [4], biomedical therapy and informatics [21, 22]. A clinical decision-making application called Mentor was used to predict mental disabilities in newborns baby [23]. It was built based on the BN model using data that were validated by respected experts in their fields. While another system called DiaVal was developed to help diagnose heart attack using BNs. In the system, the BNs structure was built based on expert knowledge using a causal representation of cardiac pathophysiology [24]. Another case of the application of BNs model was to used in a model for diagnose Alzheimer's disease [25].

2.2. Building a Bayesian Network

Building a BNs model requires collecting complex domain-specific informations from different sources into an intuitive, coherent and easy to understand form. This work requires a structured approach to capture relevant knowledge for qualitative and quantitative parts [26]. The qualitative part is known as Directed Acyclic Graph (DAG), Figure 1 ilustrates a simple DAG BNs.

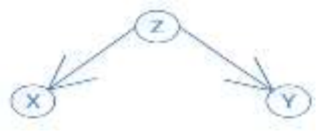

Figure 1. Simple DAG BNs

BNs have one important property that the graph can be considered as representing the joint probability distribution for all the variables. The property is part of the quantitative BNs and is expressed in Equation 1.

$$P(x_1 \ldots x_n) = \prod_{k=1}^{\infty} \boldsymbol{P(xi|Parents(Yi)})) \quad (1)$$

Where :

$P\ (x_1, \ldots, x_n)$	= Opportunities based on attributes $x_1, \ldots, x_n$
n	= Number of attributes
x_i	= i data value
$Parents\ (Y_i)$	= Immediate the predecessor or parent of the Y_i attribute

In practice, it takes expert knowledge to build a BNs model. It is also possible to use existing data from the problem domain to build a BNs. In both cases, experts in their fields must be involved in the construction of the BNs model.

The role of experts is to provide analysts with their specific knowledge. It helps analysts to synthesize domain knowledge and translate the information into BNs components such as nodes, causal relationships, and probabilities.

After identifying all variable from the problem domain, the next step is to describe the relationship between variables. One common approach is the causal relationships analysis calculated by domain experts [26, 27]. Experts must identify variables that cause other variables to take values or prevent taking certain values. Besides using experts' opinions, the dependency structure of BNs studied for is historical data.

2.3. Partial least square (PLS)

PLS is a multivariate statistical technique that can handle a lot of response and explanatory variables at once. PLS is a good alternative for multiple regression analysis and principal component regression because the PLS method is more robust, meaning that the model parameters do not change much when the new sample taken from the total population. PLS was first developed in the 1960s by Herman O. A. Wold in the field of econometrics [28].

PLS is a predictive technique that can handle many independent variables, even if multicollinearity happens among these variables. PLS is a powerful analytical method because it can be applied to all data scales, does not require many assumptions and the sample size does not have to be large. PLS, aside of being used as confirmatory method of a theory, can also be used to build relationships that have no underlying theory or for propositions test. PLS can also be used for structural modelling with reflective or formative indicator [28].

Partial least square (PLS) is a more appropriate approach for predictive purposes. This model developed as an alternative to situations where the theory is weak, or the available indicators do not meet the reflexive measurement model. Wold (1985) in Sawatsky [29] mentions PLS as "soft modelling". PLS-

SEM allows users to use a measurement scale other than intervals such as nominal data, ordinal and ratio data where this is not permitted in SEM based covariates that we have known so far. By using PLS, it is possible to do research using indicators that are reflective or formative.

3. Research Methods

This research aims to build a decision-making model to help diagnose stroke attack. The method proposed in this research is the combination of BNs and PLS methods. The research data is from medical record data about stroke and expert opinions. The overview of the proposed model is shown in Figure 2.

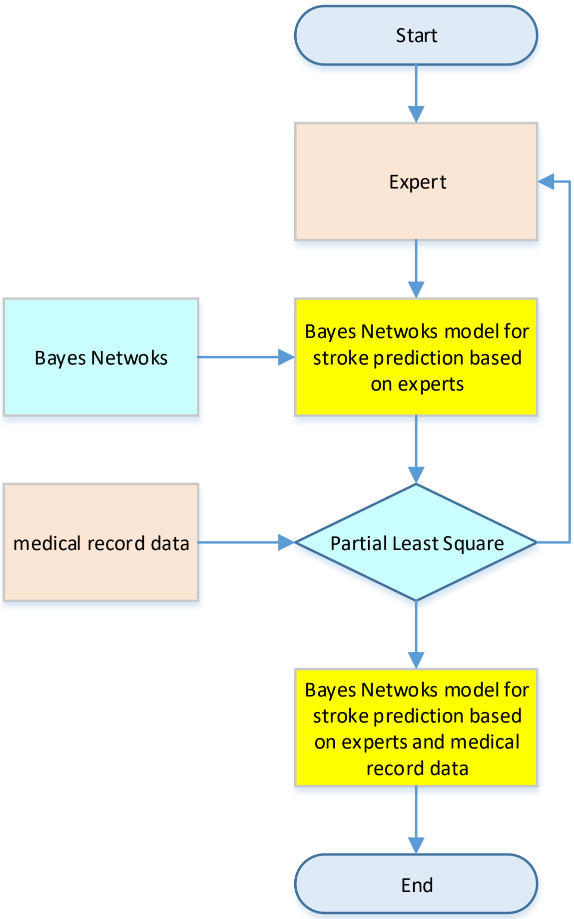

Figure 2. Stroke Partial Least Square (PLS) Bayesian Networks (BN) Model

The modelling process is carried out in two stages. The first stage is to build the structure of BNs and conduct testing using PLS. The second stage is to run the BNs learning parameter process.

The process begins with expert assumptions about the attributes that interrelate with stroke, such as the size of diastolic and systolic blood pressure, body paralysis, face droops, slurred speech, loss of body feeling, decreased consciousness, impaired function, and pain head of unknown cause. The next step is to build a directed acyclic graph (DAG) model based on the BNs method. Built DAG model validated using the PLS method based on medical records of stroke patients. Validation results are then analysed by involving experts to rebuild then the new model based on the results of the analysis. Validated model can then be used to help diagnose stroke patients.

4. Results and Discussion

4.1. BNs structure model

There are two ways to build a BNs structure model. The first one is to use assumptions from the opinions of experts, while the second one is based on data and computational algorithms such as Hill Climbing, K2 or TAN.

In this study, the BNs Structure model is built based on assumptions from experts, as shown in Figure 3.

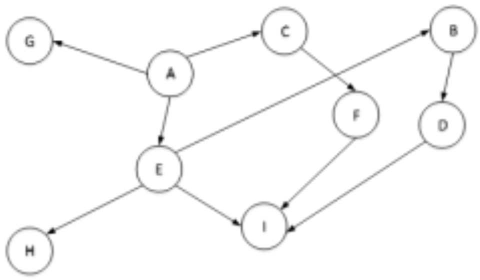

Figure 3. Bayesian Networks Prediction Model for Stroke Disease

DAG Stroke information:
A: Hypertension
B: Face/Mouth Droop
C: Paralysis of the body
D: Slurred Speech
E: Decreased awareness
F: Loss of body feeling
G: Headache
H: Noble function disorder
I: Stroke (result)

Information:

- Headaches are affected by hypertension
  (Systolic and diastolic blood pressure)
- Ruptured blood vessels affect body paralysis and decreased consciousness
- Noble function disorders are affected by hypertension which causes blood vessels to rupture and experience decreased consciousness
- The face/mouth droops cause the slurred speech and affected by a decrease in consciousness affected by ruptured blood vessels
- Loss of feeling on one side of the body is affected by paralysis of the side of the body that is affected by broken blood vessels and affects stroke

The data used in this study is from a private hospital located in Jakarta. The result of pre-processing data contained 70 rows data with nine attributes, namely: diastolic, systolic, body paralysis, face/mouth droops, slurred speech, loss of feeling, decreased consciousness, noble function disorder, and headache.

4.2. PLS-PM

The results of executing the R program using the PLS-PM algorithm shown in Figure 4.

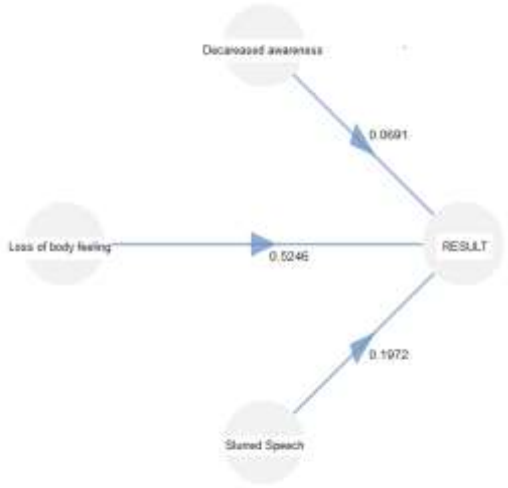

Figure 4. Results of program execution R

The calculations result using Bayesian networks shows the results of predictions containing an error of 21.42857%.

5. Conclusions

A model for diagnosing stroke attack, called the Stroke Partial Least Square (PLS) Bayesian Networks (BN) Model have been proposed. The model has been tested using PLS-PM approach, and correlations between interrelated indicators have been calculated using nine variables that are interconnected for disease prediction stroke. The results of the calculation of learning parameters using the Bayesian networks structure with variable stroke indicators showing error of 21.42857%.

6. References

[1]. http://www.yastroki.or.id/
[2]. Kesehatan, K. and RI, K.K., 2013. Riset kesehatan dasar. *Jakarta: Badan Penelitian dan Pengembangan Kesehatan Departemen Kesehatan Republik Indonesia*.
[3]. Gutierrez, J., & Yoo, C. 2011. Stroke prediction in a sample of HIV/AIDS patients: logistic regression, Bayesian networks or a combination of both. In 19th International Congress On Modelling And Simulation (pp. 1030-1036).
[4]. Heckerman, D., Geiger, D. and Chickering, D.M., 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning*, *20*(3), pp.197-243.
[5]. Burnside, E., Rubin, D., & Shachter, R. 2000. A Bayesian network for mammography. In Proceedings of the AMIA Symposium (p. 106). American Medical Informatics Association.
[6]. Ogunsanya, O.V., 2012. *Decision Support using Bayesian Networks for Clinical Decision Making* (Doctoral dissertation, Queen Mary University of London).
[7]. Alvarez, S.M., Poelstra, B.A. and Burd, R.S., 2006. Evaluation of a Bayesian decision network for diagnosing pyloric stenosis. *Journal of pediatric surgery*, *41*(1), pp.155-161.
[8]. Guo, Y., Bai, G. and Hu, Y., 2012, December. Using bayes network for prediction of type-2 diabetes. In *Internet Technology And Secured Transactions, 2012 International Conference for* (pp. 471-472). IEEE.
[9]. Cruz-Ramírez, N., Acosta-Mesa, H.G., Carrillo-Calvet, H., Nava-Fernández, L.A. and Barrientos-Martínez, R.E., 2007. Diagnosis of breast cancer using Bayesian networks: A case study. *Computers in Biology and Medicine*, *37*(11), pp.1553-1564.
[10]. Athanasiou, M. and Clark, J.Y., 2009. A Bayesian network model for the diagnosis of the caring procedure for wheelchair users with spinal injury. *Computer methods and programs in biomedicine*, *95*(2), pp.S44-S54.

[11]. Kahn, C.E., Laur, J.J. and Carrera, G.F., 2001. A Bayesian network for diagnosis of primary bone tumors. *Journal of digital imaging*, *14*(1), pp.56-57.

[12]. Luciani, D., Cavuto, S., Antiga, L., Miniati, M., Monti, S., Pistolesi, M. and Bertolini, G., 2007. Bayes pulmonary embolism assisted diagnosis: a new expert system for clinical use. *Emergency Medicine Journal*, *24*(3), pp.157-164.

[13]. Negrín, M.A. and Vázquez-Polo, F.J., 2006. Bayesian cost-effectiveness analysis with two measures of effectiveness: the cost-effectiveness acceptability plane. *Health Economics*, *15*(4), pp.363-372.

[14]. Aronsky, D., Fiszman, M., Chapman, W.W. and Haug, P.J., 2001. Combining decision support methodologies to diagnose pneumonia. In *Proceedings of the AMIA Symposium* (p. 12). American Medical Informatics Association.

[15]. Haddawy, P., Kahn Jr, C.E. and Butarbutar, M., 1994. A Bayesian network model for radiological diagnosis and procedure selection: Work-up of suspected gallbladder disease. *Medical physics*, *21*(7), pp.1185-1192.

[16]. Kline, J.A., Novobilski, A.J., Kabrhel, C., Richman, P.B. and Courtney, D.M., 2005. Derivation and validation of a Bayesian network to predict pretest probability of venous thromboembolism. *Annals of emergency medicine*, *45*(3), pp.282-290.

[17]. Lucas, P.J., Boot, H. and Taal, B., 1998. A decision-theoretic network approach to treatment management and prognosis. *Knowledge-based Systems*, *11*(5-6), pp.321-330.

[18]. Sanders, D.L. and Aronsky, D., 2006. Prospective evaluation of a Bayesian Network for detecting asthma exacerbations in a Pediatric Emergency Department. In *AMIA Annual Symposium Proceedings* (Vol. 2006, p. 1085). American Medical Informatics Association.

[19]. O Gevaert. 2006. Predicting the outcome of pregnancies of unknown location: Bayesian networks with expert prior information compared to logistic regression. Hum Reprod, 7(21):182431

[20]. Hoot, N. and Aronsky, D., 2005. Using Bayesian networks to predict survival of liver transplant patients. In *AMIA annual symposium proceedings* (Vol. 2005, p. 345). American Medical Informatics Association.

[21]. Beinlich, I.A., Suermondt, H.J., Chavez, R.M. and Cooper, G.F., 1989. The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks. In *AIME 89* (pp. 247-256). Springer, Berlin, Heidelberg.

[22]. Verduijn, M., Peek, N., Rosseel, P.M., de Jonge, E. and de Mol, B.A., 2007. Prognostic bayesian networks: I: Rationale, learning procedure, and clinical use. *Journal of Biomedical Informatics*, *40*(6), pp.609-618.

[23]. Mani, S., McDermott, S. and Valtorta, M., 1997. MENTOR: a bayesian model for prediction of mental retardation in newborns. *Research in Developmental Disabilities*, *18*(5), pp.303-318.

[24]. Diez, F.J., Mira, J., Iturralde, E. and Zubillaga, S., 1997. DIAVAL, a Bayesian expert system for echocardiography. *Artificial Intelligence in Medicine*, *10*(1), pp.59-73.

[25]. Guerrero, J.M., Martínez-Tomás, R., Rincón, M. and Peraita, H., 2016. Diagnosis of Cognitive Impairment Compatible with Early Diagnosis of Alzheimer’s Disease. *Methods of information in medicine*, *55*(01), pp.42-49.

[26]. Laskey, K.B. and Mahoney, S.M., 2000. Network engineering for agile belief network models. *IEEE Transactions on knowledge and data engineering*, *12*(4), pp.487-498.

[27]. Oteniya, L., 2008. Bayesian belief networks for dementia diagnosis and other applications: a comparison of hand-crafting and construction using a novel data driven technique.

[28]. Pirouz, D.M., 2006. An overview of partial least squares.

[29]. Sawatsky, M.L., Clyde, M. and Meek, F., 2015. Partial least squares regression in the social sciences. *The Quantitative Methods for Psychology*, *11*(2), pp.52-62