A Kohonen Artificial Neural Network as a DSS model for predicting CAD

Sri Hartati

Computer Science and Electronics Department,
Faculty of Mathematics and Natural Sciences,
Universitas Gadjah Mada,
Yogyakarta, Indonesia
shartati@ugm.ac.d

***Abstract*— A clinical decision support system (DSS) is a computer tool which uses two or more items of data to generate patient or encounter-specific advice. There are a wide variety of DSSs, ranging from simple reminder/alert systems to complex guidelines that model chronic disease management. A major component of DSS is a model base which, in this case, a Kohonen self-organizing network is discussed. This paper presents a Kohonen artificial neural network as a model of DSS for the prediction of risk factor based coronary artery disease (CAD). The system was trained using data generated in real life. The results showed that the proposed method consists of 6 steps computation for training process. The model was successfully implemented and tested success rate of 89.47%. The model can be adopted as a model in DSS, this model assists the patients and doctors in managing CAD with care.**

***Keywords*— kohonen, kson, conorary artery disease,artificial neural network.**

I. Introduction

Decision support systems (DSS) are gaining an increased popularity in various domains, including business, engineering, the military, and medicine. They are especially valuable in situations in which the amount of available information is prohibitive for the intuition of an unaided human decision maker and in which precision and optimality are of importance. Decision support systems can aid human cognitive deciencies by integrating various sources of information, providing intelligent access to relevant knowledge, and aiding the process of structuring decisions. They can also support choice among well defined alternatives and build on formal approaches, such as the methods of engineering economics, operations research, statistics, and decision theory. They can also employ articial intelligence methods to address heuristically problems that are intractable by formal techniques.

A clinical decision support system (DSS) is a computer tool which uses two or more items of data to generate patient- or encounter-specific advice [16,17]. There are a wide variety of DSSs, ranging from simple reminder/alert systems to to complex guidelines that model chronic disease management.

Employing soft computing techniques in medical applications could reduce the cost, time, human expertise and medical error. According to [15], by using these techniques include neural networks methods; it will help the clinicians especially doctor to diagnose the disease easily and quickly. Many applications that have been employed artificial neural network (ANN) to diagnose disease like diagnosing cancer [7], brain tumour [15] and diabetes [10], skin disease [19] established a diagnosis scheme for dermatoscopic images known as the ABCD (asymmetry, border, colour and diameter) rule of dermatoscopy. According to [16] the ABCD rule shows the weaknesses of the analysis done by the doctor because there is no defined limit between the different states and the feature extraction depends entirely on the doctor's experience. ANN is choosing in diagnosing skin disease because the main feature of neural networks is parallel processing in a large group of relatively simple but highly interconnected processors and the ability to self-organize or adapt through learning algorithms that change the connectivity between the units [11]. Qureshi et al [11] pointed out that ANN is significantly more accurate at predicting stage progression in cancer than the clinicians and the application of ANN utilizing large numbers of variables may help the clinicians in improving accuracy of diagnosis and predicting prognosis. At this point, ANN can be seen as a role of complementing traditional methods to produce more accurate detection of clinicians' outcomes. By using BPNN [4], the data set that is difficult to classify becomes easy to classify. The advantage of ANN is tolerance to the variability of the input data. The ability of ANN is to learn trends from sample data and then apply this knowledge for future classification. It is a good technique for many clinical pattern recognition problems and being increasingly used in medical diagnoses [1,3, 12].

Coronary artery disease or CAD (also known as ischemic heart disease, coronary heart disease or heart disease), refers to abnormalities of the arteries that carry oxygen and other nutrients to the heart muscle. This disease has risk factors that are conditions that increase the risk of developing heart disease. Some can be controlled while others are uncontrollable. Someone at risk for Coronary Artery Disease if someone has high blood pressure, has high blood cholesterol levels, has diabetes, smokes, is overweight, leads a sedentary lifestyle, has stress (controllable factors) and has a family history of heart disease, and being a male over 45 or a female over 55[5].

This paper presents a Kohonen anrtificial neural network as a model of DSS for the prediction of risk factor based coronary artery disease. The system which is trained using data generated in real life

The development involves three main processes: pre-processing, design and construction, and testing and evaluation. Pre-processing is to perform data normalization and transformation. Input data or data set is obtained from the patient data. Input data is converted into categorical type and is used for training and testing. Meanwhile, design and construction incorporates interface design, coding as well as applying the Kohonen artificial neural network. Lastly, testing and evaluation to perform the system flow, neural network training and testing.

II. KSON As a Model of Decision Support System

Decision Support System (DSS) is composed of the following subsystems: Data Management, Model Management, Communication and Knowledge Management. The conceptual stucture of DSS is shown in Figure 1. The data management includes the database(s), which contains relevant data for the situation and is managed by software called database management systems (DBMS). Model Management may include financial, statistical, management science, or other quantitative models that provide the system's analytical capabilities, and an appropriate software management. The user can communicate with and command the DSS through the Communication (dialog subsystem). It provides the *user interface.* The optional subsystem, Knowledge Management, can support any of the other subsystems and act as an independent component [13].

The model management subsystem of the DSS is composed of the following elements: model base, model base management system, modeling language, model directory, model execution, integration, and command.

Figure 1: Conceptual structure of DSS [13]

A model base contains routine and special statistical, financial, management science, artificial intelligence approach and other quantitative models that provide the analysis capabilities in a DSS. The ability to invoke, run, change, combine, and inspect models is a key capability in DSS that differentiates it from the traditional computer-based information system (CEIS).

One of the models in the model base discussed here is a Kohonen artificial neural network. The model is used to predict whether a patient suffer conorary artery disease or not based on risk factors are conditions that belong to patient.

Kohonen is self-organizing network (KSON), also known as the Kohonen self-organizing map or self-organizing map, belongs to the class of unsupervised learning networks. This means that the network, unlike other forms of supervised learning-based networks, updates its weighting features without the need for a performances feedback from a teacher or a network trainer. One major feature of this network is that the nodes distribute themselves across the input space to recognize groups of similar input vectors, while the output nodes compete among themselves to be fired one at time in response to a particular input vector [3,4].

A Kohonen self organizing network (KSON) is used to cluster a set of input vectors which each consists of the values representing the risk factors of coronary artery disease. The risk factors include high blood pressure, high blood cholesterol levels, diabetes condition, smoking condition, overweight level, and level of physical activity. The structure of the Kohonen neural network for this purpose is shown in Figure 2.

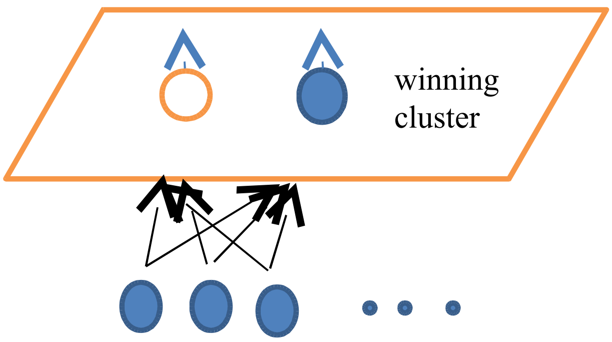
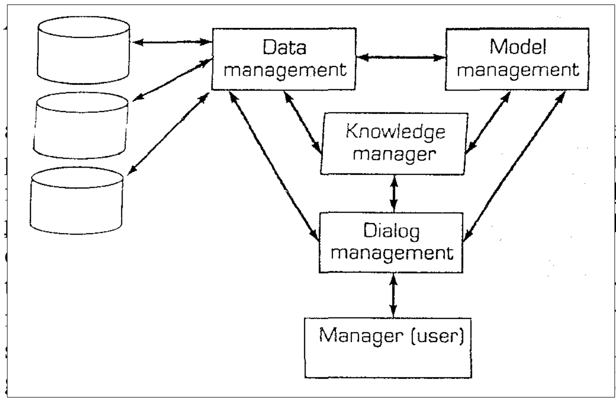


Figure 2: Structure of the Kohonen neural network



and 1
enting
t data
blood
l total
1 or
e any
tions,
were
karta,

Indonesia.

These features are assigned values between 0 and 1, according to the patient's data, such as tabulated in Table 1. The categorization of these features follows references [6,7,18] . They are assigned as vector input to KSON, the number of cluster output is two. One of them represent the prediction of coronary artery disease from which patient can suffer, other is vise versa.

Table 1: Features to confirm a prediction of conorary artery disease.

	Features	Physical sign	Value		Parameters	Physical sign	Value
1	Age	years		6	Total cholesterol	mg/dL	
		20-34	0.2			<200	0.2
		35-49	0.4			200-239	0.5
		50-64	0.6			>=240	1
		65-79	0.8				
		85-95	1				
2	Sys Tolic	mmHg		7	Triglycerides	mg/dL	
		< 120	0.25			<150	0.2
						150-200	0.5
		120–139	0.5			>=200	1
		140–159	0.75				
		>160	1				
3	Diastolic	mmHg		8	Smoking	not	0
		< 80	0.25			smoking	1
		80–89	0.5	9	Diabetes	not	0
		90–99	0.75			type 1	0.5
		>100	1			type 2	1
4	LDL cholesterol	mg/dL		10	Physical activity	vigorous	0.2
		< 100	0.2				
		100-129	0.4				
						moderate	0.5
						sedentary	1
		130-159	0.6	11	Overweight level	BMI	
		160-189	0.8				
		>=190	1				
					Under Weight	<18.5	0.2
					normal	18.5-24.9	0.4
5	HDL cholesterol	<40	1		Over weight	25.0-29.9	0.6
		40-59	0.6		obesity	30.0-39.9	0.8
		>=60	0.2		extreme obesity	>40	1

B. Learning Algorithm

The learning process permits the clustering of input data into a smaller set of elements having similar features. It is based on the competitive learning technique, known as winner take all stratey. This process is known as competitive learning. When suitably trained, the network produces a low dimension representation of the input spaces that preserves the ordering of the original structure of the network. This implies that two input vectors with similar pattern characteristic excite two physically close layer nodes. In other words, the nodes of the KSON can recognize groups of similar input vectors. This generates a topographic mapping of the input vectors to the output layer, which depends primarily on the pattern of the input space. The input pattern is given by the vector x and let us denote by w the weight vector connecting the pattern input elements to an output node with coordinate provided by indices i and j. In most algorithmhm implementations, the weight vectors are normalized to unit length. Let us also denote N as being the neighbourhood around the winning output candidate, which has its size decreasing at every iteration of the algorithm until convergence occurs[8]. The steps of the learning algorithm are summarized as follows:

Step 1 : Initialize all weights to small random values. Set a value for the initial learning rate and a value for the neighbourhood N

Step 2 : Chose an input pattern x from the input data set.

Step 3 : Select the winning unit c (the index of the best matching output unit) such that the performances index I given by the Euclidian distance from x to w_{ij} is minimized :

$$I = \left\| x - w_{ij} \right\| = \min_{ij} \left\| x - w_{ij} \right\|$$

Step 4 : Update the weights according to the global network updating phase from iteration k to iteration k + 1 as :

$$w_{ij}(k+1) = \begin{cases} w_{ij}(k) \propto (k)[x - w_{ij}(k)], if \ (i,j) \in N_c(k) \\ w_{ij}(k) \ otherwise \end{cases}$$

where $\propto (k)$ is the adaptive learning rate (strictly positive value smaller than unity) and $N_c(k)$is the neighbourhood of the unit c at iteration k.

Step 5 : The learning rate and the neighbourhood are decreased at every iteration according to an appropriate scheme. For instance, Kohonen suggested a shrinking function in the form of

$\propto (k) = \propto (0) \propto (1 - \frac{k}{T})$, with T being the total number of training cycles and α(0) the starring learning rate bounded by one.

Step 6 : The learning scheme continues until a sufficient number of iterations has been reached or until each output reaches a threshold of sensitivity with respect to a portion of the input space. The optimal weights have been found. *Grouping of input vectors into different type of classes is done by calculating distance between the optimal weight to input vector.*

III. Experimental Results

The KSON model was test experimentally to know its accuracy before integrated in the model based as an elemen of the model management subsystem of DSS. The KSON was used to group a number of patient data, at the beginning of the study, only 38 patient's data were used to assess whether a patient has a tendency to conorary artery disease or not. Each patient's data was characterized by some features as shown in table 1. An example of a real patient data is shown in Table 2. This patient's data is a pattern to predict conorary artery disease which patient may suffer. This pattern is represented by an input vector having element values shown in the fourth column of Table 2. Tbe input vector has elements as many as the number of features used to predict coronary heart disease. In this study, each pattern consists of 11 features, and total output is two patterns, which represent the prediction of possible diseases from which patient can suffer, then the weight vector used is in the form of 11 x 2 matrix. The 38 input vectors were trainned, and the training process was imposed to any input vector, to achieve convergence towards the optimal weights.

Table 2: Features to confirm a prediction of conorary heart disease

	Features	Physical sign	Value
1	Age	55 year	0.6
2	Systolic	150 mmHg	0.75
3	Diastolic	90 mmHg	0.75
4	LDL cholesterol	153 mg/dL	0.6
		< 100	0.2
5	HDL cholesterol	66 mg/dL	1
6	Total cholesterol	243 mg/dL	1
7	Triglycerides	170 mg/dL	0.5
8	Smoking	not	0
9	Diabetes	not	0
10	Physical activity	moderate	0.5
11	Overweight level	BMI	
		30-34.9	0.6

The prediction of possible diseases from which patient can suffer, was carried out by calculating the distance vector input with optimal weights. The model was tested using 20 data of patients suffering CAD, and 18 data that not suffering CAD. The experimental results showed that among 38 tested data, 4 data results in false predictions: 3 false prediction from 20 data, 1 false prediction from 18 data. The complete data was not shown here, one of them is shown in Table 2 as example of data of patient suffering CAD.

IV. Conclusions

A Kohonen self-organizing network proposed in this paper was studied to be one of models in DSS for prediction of a conorary artery disease (CAD) of patient. The proposed method consists of 6 steps computation for trainning process. The model was successfully implented and tested success ruate of 89.47%. Eventhough the model is still need to be tested with much more data, or need to be modified in the learning process. This model can be adopted as a model in DSS, which assists the patients and doctors in managing CAD with care and finally there may be implications for a positive social change by using the DSS tool.

Acknowledgment

The author would like to thank to the Computer Science Post Graduate Program, Universitas Gajah Mada for financial support of publication.

References

[1] S. Dretseitl, L.Ohno-Machado, H. Kitter, and S.Vinterbo,"A comparison of machine learning methods for the diagnosis of pigmented skin lesions". *Journal of Biomedical Information,* , 2001, 34: 28-36

[2] L. Fausett, Fundamental of neural networks architectures, algorithms and applications, New Jersey: Prentice Hall, 1994.

[3] S.Hartati, "Soft Computing Application Development for Diagnosing Heavy Protein Deficiency Using Neural Network Technique", Proc 2nd Jogja International Physics Conference, 2007, paper, E1014-E1019.

[4] S. Haykin, *Neural Network: A Comprehensive Foundation,* New Jersey, Prentice –Hall, 1999.

[5] K. Hoffman., and T. Gambichler, Epidermisiology and health services research diagnostic and neural analysis of skin cancer (DANAOS): A multicentre study for collection and computer-aided analysis of data from pigmented skin lesions using digital dermoscopy. *British Journal of Dermatology,* 149: 801-809, 2003

[6] (2010) The healthscout website. [Online]. http://www.healthscout.com/ency/1/165/

[7] Eurobarometer (2003), "Physical Activity" [Online]. Available: http://aepo-xdv-www.epo.cdc.gov/wonder/prevguid/p0000391/P0000391.asp

[8] K.Kim, and S.Cho, "Prediction of colon cancer using evolutionary neural network", *Physical Measurement*, 2003, 17:297-303.

[9] F.O. Karray and C. de Silva,"*Soft Computing and Intelligent Systems Design, Theory, Tools and Applications*, England: Pearson, Addison Wesley, 2004.

[10] V. Onnia, M.Tico, and J.Saarinen, "Feature selection method using neural network", *Proceedings, International Conference on Image Processing*, Volume, 2001, 1:513-516

[11] K.N Qureshi, F.C Hamdy , D.E. Neal, J.K. Mello, "Neural network analysis of clinic and molecular makers in bladder cancer", *Jurnal Urol* 163:630-633.

[12] P. Rubegni, and M. Buroni, "Digital dermoscopy analysis and ANN for the differentiation of clinical a typical pigmentation skin lesion: A retrospective study". *Journal of Biomedical Technology,* 2001, 51, 185-197S.

[13] E.Turban, *Decision Support System and Intelligent System*, 6th Ed, New York: Prentice Hall, 2001.

[14] Y. Zhu, and Z Yan, “Computerized tumor boundary detection using a Hopfield neural network”. *IEEE Transaction on Medical Imaging*, 1997, Volume 16, Issue 1: 55-67.

[15] V. P Wallace, J. C Bamber, and D. C. Crawford, 2000. Classification of reflectance spectra from pigmented skin lesions, a comparison of multivariate discriminate analysis and artificial neural network. *Journal Physical Medical Biology* ,45: 2859-2871.

[16] J. C Wyatt, D Spiegelhalter, “Field Trials of Medical Decision-Aids: Potential Problems and Solutions”. In: Clayton, P. (ed.): *Proc. 15th Symposium on Computer Applications in Medical Care*, Washington, 1991, McGraw Hill Inc, New York, 3-7.

[17] J.C Wyatt, “Knowledge for the Clinician Decision Support Systems”. *J. Roy.Soc. Med.* 2000, **93**, 629-633.

[18] F. Xavier Pi-Sunyer, D.Becker, C.Bouchard, R.A. Carleton, G. A. Colditz, W.Dietz, J.Foreyt, R. Garrison, S.Grundy, B. C. Hansen, Ph.D. M. Higgins, (1998) “Clinical Guidlines on the identification, evaluation, and treatment of overweight and obesity in adults”, The Evidence Report, N A, National Institutes of Health [Online] Available :http://journals.lww.com/acsm-msse/Abstract/1999/11001/Physical_activity_in_the_treatment_of_the.10.aspx

[19] M Yusoff, S.A. Rahman, S. Mutalib, and A. Mohammed, Diagnosing Application Development for Skin Disease Using Backpropagation Neural Network Technique, Journal of Information Technology, Vol 18, BIIL.2, December 2006.

.

[20]