Early Risk Detection of Pre-eclampsia for Pregnant women using Artificial Neural Network

https://doi.org/10.3991/ijoe.v15i02.9680

Endah Purwanti (✉), Ichroom Septa Preswari, Ernawati
Airlangga University, Kota SBY, Indonesia
endah-p-1@fst.unair.ac.id

Abstract—Pre-eclampsia still dominates maternal mortality cases in Indonesia. One effort that can be done is to establish early detection of the risk of pre-eclampsia in pregnant women. Automated devices with high accuracy are needed to detect the risk of pre-eclampsia so that the maternal mortality ratio can be reduced. This study aims to design an early detection system for the risk of pre-eclampsia based on artificial neural networks. The system is designed with 11 input parameters in the form of risk factors and output in the form of positive or negative risk of pre-eclampsia. The classification tool used in this study is backpropagation neural network with cross validation scenario at the training stage. The advantage of this system is the weighting of risk factor parameters by obstetric and gynecology specialists so that the results of testing the device show high accuracy. In addition, the device for early detection of pre-eclampsia was also conducted by user acceptance tests for a number of pregnant women.

Keywords—Pre-eclampsia, pregnant women, early detection, artificial neural network.

1 Introduction

The case of maternal death is the biggest problem in the world. The World Organization of Maternity Health mentions that the world's maternal deaths in 2015 reached 303,000 people with a Maternal Mortality Rate of 216 per 100,000 live births and an estimated 830 women die every day due to complications during pregnancy and childbirth. Maternal deaths occur in developing countries higher than in developed countries. The maternal mortality ratio in developing countries by 2015 is 239 per 100,000 live births, whereas in developed countries it is 12 per 100,000 live births [1].

Indonesia is one of the developing countries with high maternal mortality rate. Based on the Indonesian Demographic Health Survey in 2015, there is a maternal mortality rate of 305 per 100,000 live births [2]. The maternal mortality rate in East Java in 2014 is based on data from Local Area Monitoring of Maternal and Child Health of East Java Province of 93.53 per 100,000 live births. Pre-eclampsia and eclampsia were the dominant factor of 31.04% of maternal deaths in East Java [3].

Pre-eclampsia is a major cause of maternal and perinatal morbidity and mortality and is considered a consequence of placental impairment [4]. Pre-eclampsia is characterized by hypertension and an organ disorder caused by pregnancy or affected by current pregnancy. This disease generally occurs in the 3rd quarter of pregnancy, and it gets worse with increasing gestational age. However pre-eclampsia can occur even without showing symptoms of elevated blood pressure and presence of protein in urine [5]. Among deaths of pre-eclampsia, about 42% due to delay in seeking medical treatment, about 39% due to the health condition of patients, and about 39% due to lack of patient. knowledge about the severity of symptoms or conditions experienced [6]. The etiology of peeklampsia is not known. There have been many theories that suggest the occurrence of preeclampsia (hypertension in pregnancy), but none of these theories can explain the various symptoms that arise. So pre-eclampsia is referred to as "the disease of theory" [7].

Every pregnant woman is at risk for pregnancy-related illness. Pre-eclampsia is one of the complications of pregnancy that can not be ascertained the cause. Pre-eclampsia occurs during pregnancy with a higher risk of first-time pregnancy, adolescence, containing more than one baby [8]. Based on data that has been reported to show that pre-eclampsia events still dominate the cause of maternal death so it needs to be done more seriously, one of them by enforcing early detection of risk of pre-eclampsia occurrence in pregnant women. Therefore this research is proposed so that the results can be used as a tool for pregnant women to find out as early as possible the potential for pre-eclampsia that may occur so that they can take preventive measures and consult a doctor. In addition, it is expected to provide information regarding risk factors for the occurrence of pre-eclampsia so that pregnant women will be more concerned about the health of themselves and their babies.

The development of science and technology today has made use of artificial intelligence to solve problems, one of which is a classification tool. One study that utilizes artificial intelligence entitled "Expert System for Early Detection of Risk Levels for Pregnant Women against Pre-eclampsia with Fuzzy Logic" [9]. They built a pre-eclampsia detection system with 7 input variables namely systolic blood pressure, diastolic blood pressure, weight gain, gestational age, maternal age, edema and proteinuria. The results of the system evaluation obtained 85% accuracy. The re sults of the study still need to be improved, especially in terms of improving system accuracy.

Artificial neural networks are also one of the fields of artificial intelligence that have reliable ability to recognize data patterns . Artificial neural networks can change their structure to solve problems based on external and internal information flowing through the network. Therefore artificial neural networks can be used as a classification device. Another study, entitled "Neural Networks to Estimate The Risk For Preeclampsia Occurrence" utilizes artificial neural networks in diagnosing pre-eclampsia [10]. They used a multilayer Neural Network to estimate the risk of pre-eclampsia in pregnant women. The data used consisted of 6838 cases of pregnant women in England obtained from the Harris Birthright Research Center for Fetal Medicine, London. Each of the 24 data parameters were selected to be 15 parameters considered to characterize the risk of pre-eclampsia including ethnic origin, smoking during pregnancy, alcohol intake during pregnancy, drug abuse during pregnancy, medical history, drugs, gestational age (in day), pre-eclampsia history of previous pregnancy, family history of preeclampsia (sister, mother, or both), maternal weight and height, mean blood pressure, uterine pulsatility index, and crown rump length. The result of the research has prediction percentage of pre-eclampsia case in training set equal to 83,6% and test set equal to 93,8%.

Based on the background above, researchers are interested in designing automated pre-eclampsia detection systems based on artificial neural networks. This research begins with determining the input system and its weighting. This stage is assisted by obstetrics and gynecology experts. The system input consists of 11 parameters in the form of risk factors for pre-eclampsia, namely maternal age, body mass index before pregnancy, body mass index during pregnancy, pregnancy, history of pre-eclampsia / eclampsia, history of abortion more than 2 times, history of hypertension , history of diabetes mellitus, family history of hypertension, family history of diabetes mellitus and pregnancy in couples now. We use medical record data for system training. Data classification uses backpropagation artificial neural networks.

2 Materials and Methods

The research conducted at Medical Instrumentation Laboratory, Department of Physics, Faculty of Science and Technology, Airlangga University, RSUD dr. Mohamad Soewandhie Surabaya, East Java, and Kholifah Midwifery Clinic, Srengat, Blitar. Dataset used in this study using information relating to the incidence of pre-eclampsia from medical record data of pre-eclampsia patients at the RSUD dr. Mohamad Soewandhie Surabaya. The research stage consists of the stages of planning, designing, training and testing the system.

2.1 Planning stage

Data collection is in the form of medical record and interview data to collect information about the occurrence and risk factors in pre-eclampsia by means of question and answer and consult directly to obstetric and gynecology specialists. Selection of data was performed to select some data input required in the early detection of risk of pre-eclampsia. The selected data is divided into 12 categories including maternal age, pre-pregnancy weight, body weight during pregnancy, height, parity, pregnancy in the current partner, history of pre-eclampsia / eclampsia, history of miscarriage more than 2 times, history of hypertension, Diabetes mellitus, family history of hypertension, family history of diabetes mellitus. Each variable has a range of values ranging from 0 to 1. Giving weight is based on an expert's assessment of obstetric and gynecology specialists. The system output is in the form of information on the risk of pre-eclampsia or no risk of pre-eclampsia. The risk parameters for pre-eclampsia and their weights are shown in Table 1.

Table 1. The risk parameters for pre-eclampsia and their weights

Risk Factors	Category	Weight
Maternal age (X1)	Less than 20 years	0,6
	20 – 35 years	0,5
	More than 35 years	0,6
Body mass index before pregnancy (X2)	Less Than 18,49 kg/m^2 (Less Weight)	0,5
	18,5 – 24,9 kg/m^2 (Normal)	0,5
	25,0 – 29,9 kg/m^2 (More Weight)	0,6
	More than 30,0 kg/m^2 (Obesity)	0,8
Body mass index during pregnancy (X3)	18,5 – 24,9 kg/m^2 (Normal)	0,5
	25,0 – 29,9 kg/m^2 (More Weight)	0,6
	More than 30,0 kg/m^2 (Obesity)	0,8
Pregnancy (X4)	First	0,8
	More than 1 times	0,5
History of pre-eclampsia / eclampsia (X5)	Yes	0,9
	No.	0,2
History of abortion more than 2 times (X6)	Yes	0,7
	No.	0,2
History of hypertension (X7)	Yes	0,9
	No.	0,2
History of diabetes mellitus (X8)	Yes	0,8
	No.	0,2
Family history of hypertension (X9)	Yes	0,6
	No.	0,2
Family history of diabetes mellitus (X10)	Yes	0,6
	No.	0,2
Pregnancy in couples now (X11)	First	0,8
	More than 1 times	0,4

2.2 System design

In this study, the classification method used is backpropagation artificial neural network.The system design uses 11 neurons (according to table 1, namely 11 risk factors) in the input layer and 1 neuron in the output layer. The number of neurons in the hidden layer will be varied to get optimal performance. Backpropagation algorithm consists of two parts : backpropagation training algorithm and backpropagation testing algorithm. Architecture of artificial neural network backpropagation in the research can be seen in Figure 1.

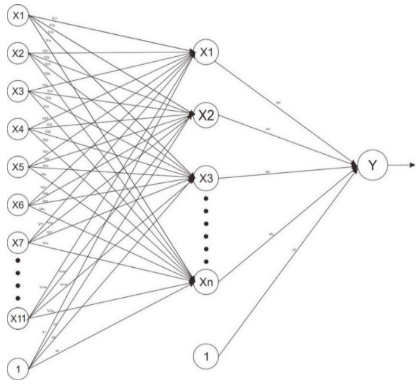

Fig. 1. Artificial Neural Network architecture backpropagation of early Pre-Eclampsia Detection system

2.3 Training

The data that has been given weighting then do the training to get the optimal weight value. This research will do variation of parameter to get optimal weight result. These parameters include hidden layer, maximum epoch, and the value of learning rate. The training process includes feedforward stage and backpropagation stage.

2.4 Testing

Backpropagation testing process used algorithm to stage feedforward. Inputs used for testing in the form of patient medical record data based on patient risk factors, the optimal weight value obtained from the backpropagation training process. Then the output value will be compared with the threshold value to determine the appropriate target. User Acceptance Test is done to find out whether the developed application is acceptable to users and the test results have satisfied the needs of the user. Aspects used in this test include aspects of software engineering, aspects of functionality and aspects of visual communication. Applications generated in this study will be tested to the respondent who is a pregnant women.

3 Results

In this experiments, data collected were 100 patient datas, covering 70 medical record data of pre-eclampsia patient and 30 medical record data of patients did not have pre-eclampsia in April until October 2017. Form of medical record data of patient of dr. Mohamad Soewandhie Local General Hospital.

Artificial neural networks are strongly influenced by the selection of initial weights in the rapidness of the training process toward convergence, as well as the minimum error value to be achieved. Small random numbers are used in assigning weight values and initial bias with a range of values from 0 to 1. The backpropagation parameters that affect the training process include maximum epoch, learning rate, number of neurons in the hidden layer and target error. Destination performance or target error is determined to compare with errors generated on the network during training. In this study determined the target error of 0.00001. In addition, variations of parameters for backpropagation training were conducted. Maximum variation of epoch values include 50, 100, 300, 500, 1000, 3000, 5000. Variations in the number of neurons in the hidden layer include 5 neurons, 10 neurons, and 15 neurons. Variations in learning rate with variations in values ranging from 0.1; 0.2; 0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9 and 1. The activation function used is the logsig activation function or the binary sigmoid. This activation function is chosen because the output unit is shown only 2 decisions, which has the risk of experiencing pre-eclampsia with a value of 1 or no risk with a value of 0.

Database is created to store tables containing medical record data and training results. In the data database consists of 4 tables including test tables, selection tables, training tables, and user tables. Interface design in pre-eclampsia risk detection application used backpropagation method consist of interface for user and interface for admin. The user interface consists of 5 menus: home menu, detection menu, article menu, help menu, and information menu. While the interface for the admin consists of 5 menus are the entry menu, home menu, training menu, test menu, and article menu.

The data used were 100 medical records of pre-eclampsia and not pre-eclampsia. This aspect of data sharing needs to be considered so that the artificial neural network used gets enough training data and get test data that is able to test the performance of the training conducted. Backpropagation training in this study aims to get the weight that is able to produce the overall calculation as close as possible to the target so as to produce high accuracy. Variations are performed on the training of data to obtain the desired optimal weight. Variations of backpropagation parameters in the study include maximum epoch, neurons in the hidden layer, and learning rate(α).

Cross validation to determine the effectiveness and reduce the overfitting of artificial neural network method backpropagation. This research is done cross validation by dividing 100 data into 20 fold (part) so that known 1 fold consist of 5 data. Data sharing for the training and testing process as in the original provision where 80% of the 20 fold is 16 fold as the training data set and 20% of the 20 fold is 4 fold as the test data set. Cross validation process is calculated by 1 fold sliding each scenario so that the maximum number of scenarios is 20 scenarios. For example, in the 1st scenario, the training data set consists of folds 1 through 16, and the remainder is used for test data sets. In the second scenario, the training data set shifts to fold 2 to 17 and the rest for the test data set. This process continues until the training data and test data are the same as the data in the 1st scenario.

In cross validation, each scenario is done by training with variation of backpropagation parameter, then selected the best parameter. The weights on the best parameters are used in the tests which are then known for their accuracy. Used also a threshold value of 0.5 specified in the backpropagation training process to know the number of conformity of the results of the application with the diagnosis. Table 2 shows the best model selection with the level of training accuracy as well as the test obtained through cross validation.

Table 2. of Backpropagation Training and Testing with Cross Validation

No.	Epoch	Neuron Hidden Layer	α	Training	MSE	Time	Testing
1	3000	5	1	100 %	0,0001732	11,0305	100 %
2	3000	5	1	100 %	0,0001187	11,5367	95 %
3	3000	5	1	100 %	0,0000997	11,8102	90 %
4	3000	5	1	100 %	0,0001073	10,9467	95 %
5	3000	5	1	100 %	0,0001689	10,7335	90 %
6	3000	5	1	100 %	0,0004221	11,0695	95 %
7	3000	5	1	100 %	0,0002018	10,9362	90 %
8	3000	5	1	100 %	0,0001763	10,6827	95 %
9	3000	5	1	100 %	0,0001361	10,8839	100 %
10	3000	5	1	100 %	0,0001246	10,7575	95 %
11	3000	5	0,8	100 %	0,0012168	10,8381	100 %
12	3000	5	0,9	100 %	0,0004367	12,4842	95 %
13	5000	5	1	100 %	0,0001268	17,1478	95 %
14	5000	5	1	100 %	0,0001022	17,9354	95 %
15	3000	5	1	100 %	0,0002010	10,8824	85 %
16	3000	5	1	100 %	0,0002470	10,6302	95 %
17	3000	5	1	100 %	0,0005210	10,6799	85 %
18	5000	5	1	100 %	0,0001101	17,6603	85 %
19	3000	5	1	100 %	0,0003534	10,9017	90 %
20	3000	5	1	100 %	0,0001591	10,8039	90 %

Based on information Table 2, the training accuracy rate for all scenarios is 100% while the accuracy of the test obtained is different. In the 1st scenario, the 9th and 11th scenarios have the highest level of training and testing accuracy. Judging from the parameters used, the 1st and 9th scenarios have the same backpropagation parameters as the maximum epoch 3000, the number of neurons in the hidden layer of 5 neurons, and the learning rate of 1. While the 11th scenario has a backpropagation parameter that is maximal epoch 3000, the number of neurons in the hidden layer of 5 neurons, and the learning rate of 0.8. The best model in the 9th scenario seen from the resulting MSE value is smaller than in the 1st and 11th scenarios. The value of MSE gained in training for the 9th scenario is 0.0001361. MSE graph backpropagation training in the 9th scenario to the maximum epoch can be seen in Figure 2.

Fig. 2. MSE Graph of Backpropagation Training on The 9th Scenario, The Number of Neurons in The Hidden Layer as Much as 5, Learning Rate of 1

The maximum requirement of epoch and the limit of MSE is used in the study so that the training process runs during both terms are met. The maximum epoch value can affect the accuracy and value of the generated MSE. The graph above shows that the value of MSE is getting smaller and closer to the value 0 along with the amount of epoch done.

Backpropagation testing in this study aims to determine the ability of backpropagation that has been built and trained in recognizing new patterns. Test results in the form of conformity with the target output which will then be calculated the level of accuracy testing. In knowing the conformity of the output results with the target in the research used a threshold value of 0.5 that has been determined in the backpropagation training process. The following is the result of testing on the 9th data subsection that can be seen in Table 3.

Test results in the 9th scenario show that all application detection results are in conformity with the expert diagnostic results so that 100% accuracy is obtained. In the 9th scenario the best parameters are obtained when the number of neurons in the hidden layer is 5, the learning rate of 1 and the maximum epoch of 3000 has the test result with 100% accuracy. This suggests that backpropagation has been able to recognize test data well and distinguish patient data that has a risk of pre-eclampsia and has no risk of pre-eclampsia.

Evaluation of User Acceptance Test conducted on early detection application of pre-eclampsia risk is by filling in questionnaire sheets to respondents who are pregnant women. This test was conducted at the midwife clinic Kholifah, Srengat, Blitar regency with the number of respondents as many as 10 people. In this test there are three aspects of assessment that is, software engineering aspects, functional aspects and visual communication aspects. Table 4 shows that the three aspects of the user acceptance test received a "good" response of 86.66%, 68% and 62.5%.

Table 3. The result of testing on the 9th data subsection, positive means having a risk of pre-eclampsia, negative means not having a risk of pre-eclampsia

Testing Data (n)	Detection System	Expert Diagnosis	Result
1	Positive	Positive	Matched
2	Negative	Negative	Matched
3	Positive	Positive	Matched
4	Positive	Positive	Matched
5	Positive	Positive	Matched
6	Positive	Positive	Matched
7	Negative	Negative	Matched
8	Positive	Positive	Matched
9	Positive	Positive	Matched
10	Negative	Negative	Matched
11	Positive	Positive	Matched
12	Negative	Negative	Matched
13	Negative	Negative	Matched
14	Negative	Negative	Matched
15	Positive	Positive	Matched
16	Positive	Positive	Matched

Table 4. Result of User Acceptance Test

No.	Aspect of Assessment	Very Bad	Bad	Good Enough	Good	Very Good
1	Software Engineering Aspects	0 %	0 %	6,67 %	86,66 %	6,67%
2	Functional Aspects	0 %	0 %	10 %	68 %	22 %
3	Visual Communication Aspects	0 %	0 %	20 %	62,5 %	17,5 %

4 Conclusion

The results showed that the parameter weighting as a system input produces high accuracy both at the training and testing stages of artificial neural networks which is 100%. The cross validation procedure is also very helpful in finding the optimal parameters of backpropagation in recognizing training data patterns, namely the number of neurons in the hidden layer of 5 neurons and MSE, ie 0.0001361. Application user acceptance testing consisting of software engineering aspects, functional aspects and visual communication aspects shows a "good" response.

5 References

[1] WHO, 2015, *Tends in Maternal Mortality: 1990 to 2015,* http://apps.who.int, 12 Oktober 2016
[2] Kementrian Kesehatan Republik Indonesia, 2015, *Profil Kesehatan Indonesia 2015*, http://www.depkes.go.id, 12 Oktober 2016
[3] Dinas Kesehatan Provinsi Jawa Timur, 2015, *Profil Kesehatan Provinsi Jawa Timur* , http://www.depkes.go.od, 12 Oktober 2016
[4] Poon, Leona C., Nicolaides, Kypros H., 2014, Early Prediction of Preeclampsia, *Obstetrics and Gynecology International,* **2014**: 1-11 https://doi.org/10.1155/2014/297397
[5] Moreira, Mário W. L., Rodrigues, Joel J. P. C., Oliveira, Antonio M. B., Saleem, K., and Neto, A. V., 2016, A Preeclampsia Diagnosis Approach using Bayesian Networks, *IEEE International Conference on Communications,* 1-5 https://doi.org/10.1109/ICC.2016.7510893
[6] Brown, T., 2015, Maternal Mortality Largely Preventable, Large Study Shows, http://www.medscape.com, 12 Oktober 2016
[7] Manuaba, Ida B. G., Manuaba, Ida A. C., dan Manuaba, Ida B. G. F. 2007. *Pengantar Kuliah Obstetri.* Jakarta: EGC
[8] Cambell, Stuart. 2008. *Kehamilan Hari Demi Hari.* Jakarta: Erlangga
[9] Nugraheni, Adinda M.," Sistem Pakar Deteksi Dini Tingkat Resiko pada Ibu Hamil terhadap Pre-eklampsia dengan Logika Fuzzy", Universitas Airlangga, Surabaya, Indonesia, 2016.
[10] Neocleous, C.K., Nicolaides, K.H., Neokleous, K.C., Schizas, C.N. 2009. "Neural Networks to Estimate The Risk for Preeclampsia Occurrence". *International Joint Conference on Neural Networks*, 2221-2225
[11] Kusumadewi, Sri, 2004, Membangun Jaringan Saraf Tiruan menggunakan Matlab & Excel Link, Penerbit Graha Ilmu, Yogyakarta

6 Authors

Endah Purwanti is currently lecturer of Program of Biomedical Engineering at faculty of Science and Technology, Airlangga University, Indonesia. Her activities of research in the medical instrumentation laboratory focusing on software engineering, data mining, artificial intelligence and image processing.

Ichroom Septa Preswari received a bachelor's degree in biomedical engineering from Airlangga University, Indonesia, 2017. Her research interests include software engineering and artificial intelligence.

Ernawati is a lecturer at department of obstetrics and gynecology, faculty of medical, Airlangga University, Indonesia. She finished her Phd program with the topic of preterm preeclampsia on 2015.

Article submitted 10 October 2018. Resubmitted 21 October 2018. Final acceptance 17 November 2018. Final version published as resubmitted by the authors.