Neural Networks Algorithm to Inquire Previous Preeclampsia Factors in Women with Chronic Hypertension During Pregnancy in Childbirth Process

Muhlis Tahir, Tessy Badriyah, and Iwan Syarif
Department of Information and Computer Engineering
Graduate Program of Engineering Technology
Politeknik Elektronika Negeri Surabaya
Surabaya, Indonesia muhlistahir@pasca.student.pens.ac.id,
{tessy.badriyah,iwanarif}@pens.ac.id

***Abstract*— Preeclampsia and eclampsia is one of the complications of pregnancy caused directly by the pregnancy itself. The success of handling preeclampsia is determined by maternity compliance in pregnancy care. Pregnant women who do not have their pregnancies cause no detectable high-risk factors experienced during pregnancy. We investigated the detection of preeclampsia by using neural network and also investigated the importance of the Previous PE Case attribute on the classification results. Preeclampsia dataset obtained from Haji General Hospital Surabaya with 17 parameters that are considered to affect the risk characteristics of the occurrence of preeclampsia, including the Previous PE Case parameters are positive or negative. This study uses Neural Network to classify Preeclampsia data. We compare with other algorithms like Naive Bayes, K-Nearest Neighbors, Linear Regression, Logistic Regression and Support Vector Machine. This experiment showed that the neural network algorithm achieved the best accuracy with three validation tests, 92,46% split data, 10-folds cross validation 94,23% and LOO validation 96.66%. The same data is applied to the neural network after excluding information about the previous PE case, for the learning process. The result is the correct classification as high as 96.66% of cases of case preeclampsia using all parameters, in the test set. Predicted cases of preeclampsia, for the total results of the unknown verification test, is 90%. And if the information on the previous PE case is not used, the result will decrease significantly.**

Keywords—Neural Network; Preeclampsia; Classification

I. Introduction

Hypertension (high blood pressure) is the most familiar medical complication knowing by mother during pregnancy. Preeclampsia (more written as pre-eclampsia) issue most serious pregnancy complication, breathtaking 4-8% of all pregnancies. It is repeated (is different to earlier disorders), hypertension syndrome, whichever, if it is not prepared, can advise into eclampsia, a very fatal and usually cause a serious condition identified by mass of thickened blood and convulsion. When we look at separating the above assertion, preeclampsia is a principal subscriber to maternal and fetal morbidity and mortality general. Preeclampsia is described to high blood pressure (hypertension), fluid retention (edema) and excess protein rate in the urine (proteinuria). This evidence is nope seen in the beginning level from pregnancy and therefore preeclampsia may be hard to identify. This can simply be detected to the routine antenatal hearing of maternal and urine blood tension, and because such female without entryway to satisfy health services are especially at risk [1].

Currently, there are no at all to cure preeclampsia. In heavy matter, the simple how to ease maternal indication is to cause artificial labor whether in premature delivery by cesarean section. Premature birth could own heavy consistency and each year four million babies are born by fetal growing limitation for a consistency of preeclampsia. Even though preeclampsia has been the point of many years of scientifically study, the correct etiology of conditions is anonymous. The recent study suggests that needy placental establishment will preclude the relocation of nutrients of mom to a baby is very important for the development of health. Through we study, our purpose, especially which lead preeclampsia and to recognize how for advance diagnosis of the condition, which allows we to supervise it over to good purpose. We trust for expanding safe and effective therapies for preeclampsia and lower the risks that inflict on mothers and children around the world [2]. So is necessary to make predictive models employ risk factors by data mining techniques to interventions relative to the expansion of preeclampsia. Health data record could it be to create a resolve for support systems and recognize diseases preeclampsia. any of these is by using data mining method that aims extract and finds a scheme of an aggregation of precious information. within the method of data mining, there are several methods of learning so that can be used to balance two methods of preeclampsia patient data, so lest in the health sectors can be used to predict Preeclampsia, each method also has a different feature model.

Some of the results of previous research have shown good accuracy, but still need to be improved so that the performance of an algorithm can produce high predictions and accuracy. In this study we explain the prediction of preeclampsia in pregnant women with a history of hypertension in previous deliveries who experienced preeclampsia, occurring on average, pregnant women could not know about the symptoms of preeclampsia experienced. Contributions to this study can help doctors and midwives to predict the onset of preeclampsia after the first pregnancy because it is very dangerous when doctors and midwives do not know exactly whether these pregnant women have experienced preeclampsia in previous pregnancies Using the NN algorithm with three validation methods, namely, data split, 10 Folds cross-validation and LOO validation to optimize NN performance so as to produce high predictive accuracy. After obtaining information about the prediction of the report, he also conducted an experiment by deleting one attribute, namely the history of the previous preeclampsia factor attributes that could no longer be used, whether the accuracy was increased, fixed or decreased. we will also explain in this study.

II. Related Works

Mario W.L. Moreira et al [3] has made research on predicted preeclampsia using data mining. The modeling is the Bayesian network. In his research Bayesian network proposed an intelligent decision support system in predicting preeclampsia to help doctors or experts in the care of pregnant women. The main contribution of this paper is a decision-making system during pregnancy care with Bayesian network modeling. A further study in classifying the severity of hypertension in pregnant women has done the comparison of Naïve Bayes modeling and AODE (Average One-Dependence Estimators Classifier) in conclusion Naïve Bayes has excellent performance with a precision of 0.400 and F-measure of 0.397 than AODE modeling results with 0.275 and 0.295 as Precision value and F-Measure value. In the subsequent research modeling conducted on the research is Random Forest applied to pregnant women to know early disorders in pregnancy. From this modeling, the result obtained F-Measure 0.431 and ROC curve 0.731 and the value of Kappa 0.2505 than the previous modeling of preeclampsia prediction.

Eduardo Tejera, et al [4] has made his research on preeclampsia using artificial neural networks with a heart rate variability index (HRV) in women with normal pregnancies, hypertension, and preeclampsia. In addition to the HRV index, researchers also considered other factors such as maternal history either before pregnancy or during pregnancy and the result of blood pressure, so as to obtain a sensitivity value of about 80% in pregnant women who have hypertension and normal. And the other sensitivity value is about 85-90%. The conclusion shows the HRV index and Artificial Neural Network can be useful for research and characterization of pregnant women.

Souvik Saha, et al. [5], has conducted research with the K-Nearest Neighbor algorithm and Meta-Heuristic Algorithms with the process of completing genes in the preeclampsia dataset. From the data collection used at least normal data or the response of 25,000 genes (rows) that have 75 samples (columns) and researchers have chosen 30 genes as critical genes for preeclampsia. By optimizing using the Simulate Annealing Algorithm and Particle Swarm Optimization Algorithm as an optimization of the sample used the results of 150 samples (75 with normal categories and 75 Pre-eclampsia categories) showed 80-90 samples correctly. PSO outperformed SA with an average score of 150 samples.

Wenshuai Cheng, et al. [6] In this study, focused on gestational hypertension, in 412 pregnant women who presented risk factors to patients. The results, taking into consideration the estimated accuracy of the model, of the relatively small factors obtained by gestational hypertension and preeclampsia groups were 13.3%, 8%, and 14.3.

III. Purpose

This study will predict preeclampsia with a history of preeclampsia before there is much that cannot be used for preeclampsia if it is not routinely examined by doctors and midwives. Most pregnant women do not tell the history of preeclampsia in previous pregnancies in doctors and midwives at the time of the examination, so we will predict this with the aim of helping doctors and midwives to make preliminary predictions both in subsequent pregnancies or first pregnancies because by predicting in advance, doctor and field very helpful first. By using the NN algorithm with three validation methods namely, Split Data, 10 Folds cross-validation and LOO validation and compare with other algorithms so as to produce a high level of predictive accuracy.

IV. The proposed system

A. Dataset

Before taking the demographic and laboratory data we will use, we bring references in the form of papers and journals related to the attributes used in the prediction. At the time of submission of research at Surabaya Haji Hospital because of several articles and journals that discuss preeclampsia, the attributes used are very diverse, finally from these references obstetrics and gynecology specialists' doctor one by one the features that will be used in predicting preeclampsia with the reason that. After that the researchers obtained real data from the hospital, this attribute that can be used to provide predictions in preeclampsia. The preeclampsia data set consisted of 17 attributes, 239 samples, two classes, and data for the last 2 years (2016-2017 from Surabaya Hajj Hospital as many as 239 medical records). Preeclampsia dataset usually has two classes of patients, consisting of mild and severe preeclampsia.

TABLE I. PARAMETERS THAT WERE USED FOR PREECLAMPSIA PREDICTION

1.	RM
2.	First Pregnancy
3.	Age
4.	Mean Arterial Pressure (MAP)
5.	Systolic Blood Pressure
6.	Dystolic Blood Pressure
7.	BMI (Body Mass Index)
8.	Classification BMI
9.	Previous PE?
10.	Personal History of Hypertension?
11.	Personal Pregnancy History of Hypertension?
12.	Diabetes Pragestasional?
13.	Glukosa mg/dl
14.	Proteinurine
15.	Childbirth Process
16.	Pre-exiting Medical Conditions
17.	Mother had Mild Preeclampsia or Severe Preeclampsia?

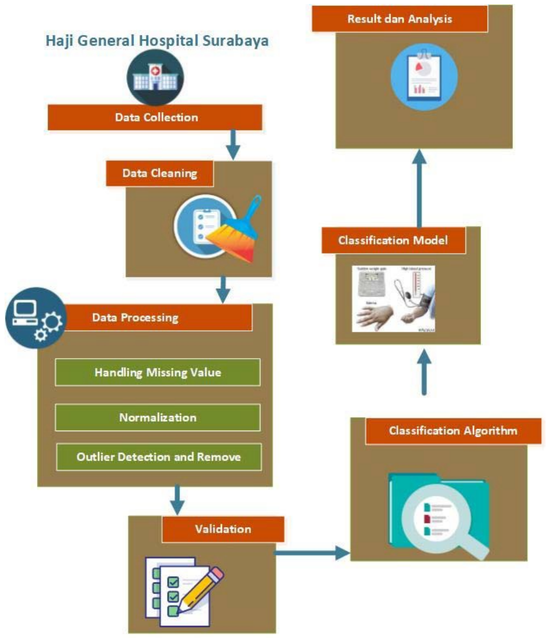

Fig.1 System Design

B. Data Cleaning

The data to be analyzed by data mining techniques is sometimes incomplete (less than the value of a particular attribute or attribute of interest, or contains only aggregate data), contains (contains errors, or outlier values deviate from expected), and is inconsistent (e.g., contains the difference in the departmental code used to categorize the goods). Data Cleaning device to "clear" the data to contents in the lost value, flattening the noise data, recognize or erase outliers, and resolving discrepancy [7].

C. Preprocessing

The process of preprocessing data is to prepare data into fixed data before data becomes training data. It assigned relies at data mining experts to improve data quality, improve the precision and influence of data mining processes. The pre-processing assignment would require 60% from the data mining process attempts [8]. I follow in main effort at preprocessing data which consists Handle missing values, normalization and outlier and delete detection.

D. Neural Network

Neural networks or commonly called Artificial Neural Networks (ANN) is a performance imitating the network of nerve cells in the brain. The computational trick in this way lies in the relationship between neurons. There are at least two layers consisting of the input layer (including the input value node) and the output layer (including the output node value after processing the activation function) [9]. Neuron modeling in this method is as follows:

$$Z = a_1w_1 + a_2w_2 + \dots + a_k w_k + b \qquad (1)$$

Where:

a_1 = inputs

w_1 = weight

b = bias

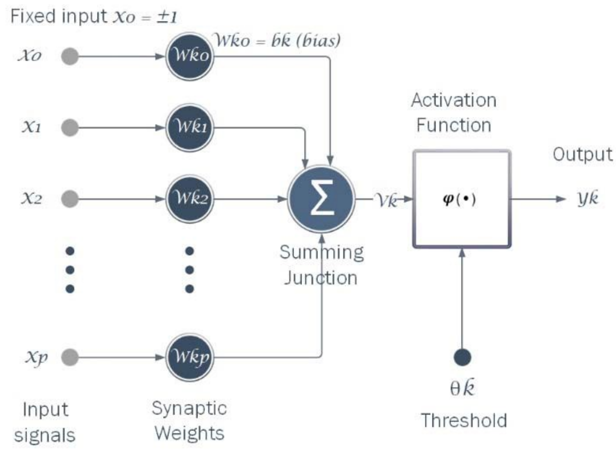

Fig.2 Ilustration of a Neural Network

The general formula used as Activation Function (AF) is Sigmoid, TanH, and Linear Rectifier. In this example the sigmoid used is as follows.

$$\sigma(z) = \frac{1}{1+e^{-z}} \qquad (2)$$

In nodes in the output, layer represent the existing classes. The proceeds of AF are a representation of the value of closeness to a class. The highest value will be used as the class determiner. The range of sigmoid function is 0 to 1. Each result of AF is the final output or it can also be the input value for the next neuron. The learning process can be illustrated if the input value is a, b and c, then the output must be m. And if the input value is d, a, and f, then the output must be k. And so on until the whole iteration is reached. In the learning process, when the desired results are not appropriate, it is necessary to update the weight update is as follows.

$$\mathbf{w_i = w_i + error * \alpha_i * \mu} \qquad (3)$$

This learning process is similar to regression. If the regression is looking for coefficient value at the time of learning, the neural network looks for the weight value [9].

V. EXPERIMENT AND RESULT

A. Classification Neural Prediction

We validate using three approaches. The first approach is to divide the training data by 70% and test the data by 30% so-called split data, the second by taking 10% of the dataset as the data by the 10-way cross-validation method in the data set into two sets of training data and test data.

The third validation is performed using Leave-One-Out Cross-Validation, which uses each data record as test data. In this study, results from Neural Network (NN) were compared with Naive Bayes, K-Nearest Neighbor (KNN), Linear Regression, Logistic Regression and Support Vector Machine (SVM). We use machine learning tool collaboration with Rapidminer. In the current operator using Rapidminer.

B. Neuron testing model

The preeclampsia data obtained is then preprocessed to obtain quality data by applying data cleaning using RapidMiner, which is data whose attribute values will be empty so that the data becomes more accurate. The preeclampsia dataset is 239 records, after the data cleaning process does not get data and data, with complete attributes of 239 records that will be used in the experimental process. Preeclampsia datasets before and after the data cleaning process will be testing against the Neural Network to get a good level of accuracy, with results comparison can be seen in table II.

TABLE II. COMPARISON OF DATA ACCURACY BEFORE AND AFTER THE DATA CLEANING PROCESSS

Number of datasets	Accuracy	AUC
239	90,02%	0,952
239	90,22%	0,952

The next step is to test the neuron model using one-layer layer with the number of neurons 3 to 20 and determine the error value using the smallest RMSE. The best results with one hidden layer are found in the hidden layer with the number of neurons 17 and RMSE generated by 0.104. The results can be seen in table III.

TABLE III. EXPERIMENT DETERMINATION OF THE NUMBER OF HIDDEN LAYER SIZE.

RMSE	Number of neurons
0,255	3
0,217	4
0,195	6
0,188	7
0,182	9
0,172	10
0,162	11
0,183	12
0,154	13
0,121	14
0,119	15
0,104	17
0,139	19
0,142	20

Next we determine learning values and momentum by including learning values between 0.10 to 0.50 and momentum values in the range of 0.50 to 0.95 using one layer of hiden and 17 neuron sizes. The results are found in table IV.

TABLE IV. RESULTS FROM NEURON MODELS WITH STRUCTURES.

Momentum	Learning Rate	Sum of Square Error
0,50	0,10	0,030
0,60	0,10	0,061
0,70	0,10	0,039
0,80	0,10	0,031
0,90	0,10	0,041
0,95	0,10	0,465
0,50	0,20	0,040
0,60	0,20	0,032
0,70	0,20	0,041
0,80	0,20	0,041
0,90	0,20	0,219
0,95	0,20	0,271
0,50	0,30	0,030
0,60	0,30	0,021
0,70	0,30	0,039

TABLE IV. RESULTS FROM NEURON MODELS WITH STRUCTURES. (CONTINUED)

Momentum	Learning Rate	Sum of Square Error
0,80	0,30	0,031
0,90	0,30	0,021
0,95	0,30	0,195
0,50	0,40	0,354
0,60	0,40	0,491
0,70	0,40	0,339
0,80	0,40	0,019
0,90	0,40	0,421
0,95	0,40	0,195
0,50	0,50	0,280
0,60	0,50	0,471
0,70	0,50	0,029
0,80	0,50	0,071
0,90	0,50	0,051
0,95	0,50	0,119

From the results of table IV test, the NN parameters selected are one layer of hidden with the number of neurons 17, the value of momentum 0.80, learning rate 0.40 and RMSE 0.019.

C. Testing Validation Model

by using a number of neurons 17, the value of momentum is 0.80, the learning rate is 0.40 we try to compare with three validation models and some algorithm. Validation testing is done by comparing three methods, split validation, 10-Folds cross-validation and LOO validation and several algorithms such as Naive Bayes, KNN with K = 3, Linear Regression, Logistic Regression and LibSVM with Polynomial kernel. The residual configuration uses the standard parameters in RapidMiner. Test results are found in Table V.

TABLE V. VALIDATION COMPARISON TABLE AND SOME ALGORITHMS

Validation	Accuracy (%)					
	Naive Bayes	NN	K-NN	Lin.Reg	Log.Reg	SVM
Split Data	80,78	**92,46**	84,23	86,78	86,67	91,81
10-Folds Cross Validation	82,46	**94,23**	86,78	87,01	86,94	92,03
LOO	83,46	**96,66**	87,78	88	87,44	93,81

Table II shows the accuracy of Naive Bayes, NN, K-NN, Linear Regression, Logistic Regression and SVM with Split Data Validation, 10 Folds Cross Validation and LOO validation. Comparison of Neural Network (NN) value is better than others algorithms that is 92,46%, 94,23% and 96,66%. After all features are used to train and get the best out of comparison with neural networks, a set of attributes will be reduced and reused to test on the same neural network to build sensitivity to information about women with a history of labor with preeclampsia. whether pregnant women occur preeclampsia in the next birthing process or not.

TABLE VI. PREDICTED PREECLAMPSIA RESULTS WHERE THE INFORMATION "PREVIOUS POSITIVE PE CASE" AND "PREVIOUS NEGATIVE PE CASE"

Case Preeclampisa	Training Set	Test Tes	Verification Set
No case of pre-eclampsia was found	120	12	6
Cases of preeclampsia are predicted to be true	86	9	3
The percentage of cases of Preeclampsia predicted correctly.	85,7	65	55

TABLE VII. PREDICTED PREECLAMPSIA DEVELOP IF "NEGATIVE PREVIOUS PE" INFORMATION IS NOT USED

Case Preeclampisa	Training Set	Test Tes	Verification Set
No case of pre-eclampsia was found	120	12	6
Cases of preeclampsia are predicted to be true	87	8	5
The percentage of cases of Preeclampsia predicted correctly.	89,3	58,2	90

TABLE VIII. PREDICTED PREECLAMPSIA DEVELOP IF "POSITIVE PREVIOUS PE" INFORMATION IS NOT USED

Case Preeclampisa	Training Set	Test Tes	Verification Set
No case of pre-eclampsia was found	119	19	6
Cases of preeclampsia are predicted to be true	97	16	3
The percentage of cases of Preeclampsia predicted correctly.	90,7	89,6	55

VI. CONCLUSION AND FUTURE WORK

In this study, the dataset we used is a dataset that predicts preeclampsia which provides good performance on each accuracy test. The results of classification, Neural Network Algorithm in 3 validation methods are good Split Data, 10-folds cross validation and LOO Validation provide significant values that improve the learning process that is 92.46%, 94.23%, and 96.66%. That's better than the algorithm that is the comparison.

Neural networks have proven to be effective and reliable predictors for this preeclampsia dataset. Considering the other attributes of "Positive Previous PE Case" and "Negative Previous PE Case" as information that contributes significantly to predictions, the results in Tables VI, VII and VIII can be concluded that both are needed to achieve high accuracy and correct classification results. Likewise comparing the results in tables II and IV occurring that compile information from the "previous PE case" attribute done with Neural Network on unknown sets, the results are not very reduced in matters governing the holiday test decreasing to 96,66% to 58.2%. And by comparing the results in Tables V and VIII, see that the composition of "previous PE cases" in posts with neural networks, the results of correct predictions in predicting unrecognized preeclampsia fell from 90% to 55%. So, it can be concluded that previous PE Case attributes are needed to assist in generating high accuracy and in future work to produce good predictions with great accuracy will be selectively selected attributes to input as few attribute inputs as possible on the preeclampsia dataset and increase the quality of the dataset and improve the learning process.

ACKNOWLEDGMENT

Haji General Hospital Surabaya, East Java, Indonesia. We are grateful and pleased with Dr. Dwinanto Ananda Muttaqin, SpOG, and the Medical Record Team for their assistance and support in providing preeclampsia data set as well as the assignment given and their motivation.

REFERENCES

[1] W. Cheng, L. Fang, L. Yang, H. Zhao, P. Wang, and J. Yan, "Varying Coefficient Models for Analyzing the Effects of Risk Factors on Pregnant Women's Blood Pressure," *2014 13th Int. Conf. Mach. Learn. Appl.*, pp. 55–60, 2014.

[2] C. K. Neocleous, P. Anastasopoulos, K. H. Nikolaides, C. N. Schizas, and K. C. Neokleous, "Neural networks to estimate the risk for preeclampsia occurrence," *Proc. Int. Jt. Conf. Neural Networks*, pp. 2221–2225, 2009.

[3] M. W. L. Moreira, J. J. P. C. Rodrigues, A. M. B. Oliveira, R. F. Ramos, and K. Saleem, "A preeclampsia diagnosis approach using Bayesian networks," in *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–5.

[4] E.Tejera, areia Jose, Rodrigues M,Ramoa M,J. Manuel nieto-villar and I,Rebelo, "Artificial neural network for normal, hypertensive, and preeclamptic pregnancy classification using maternal heart rate variability indexes" *in The Journal of Maternal-Fetal & Neonatal Medicine*,(2011), pp. 1147-1151

[5] S. Saha, S. Biswas, and S. Acharyya, "Gene Selection by Sample Classification Using k Nearest Neighbor and Meta-heuristic Algorithms," *2016 IEEE 6th Int. Conf. Adv. Comput.*, pp. 250–255, 2016.

[6] M. W. L. Moreira, J. J. P. C. Rodrigues, A. M. B. Oliveira, K. Saleem, and A. J. V. Neto, "Predicting hypertensive disorders in high-risk pregnancy using the random forest approach," *IEEE Int. Conf. Commun.*, 2017.

[7] E. Rahm and H. Do Hong, "Data Cleaning: Problems and Current Approaches," *IEEE Data Eng. Bull.*, vol. 23, no. 23, pp. 3–13, 2000.

[8] Z. A. Othman, A. Abu Bakar, A. R. Hamdan, K. Omar, and N. L. Mohd Shuib, "Agent based preprocessing," *2007 Int. Conf. Intell. Adv. Syst. ICIAS 2007*, pp. 219–223, 2007.

[9] S. Haykin, *Neural networks: a comprehensive foundation*.1999.