Proc. of the 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE)
12-13 June 2020, Istanbul, Turkey

Application of Naive Bayes Method for IUGR (Intra Uterine Growth Restriction) Diagnosis on The Pregnancy

Tessy Badriyah
Politeknik Elektronika Negeri Surabaya, INDONESIA
tessy@pens.ac.id

Nadia Ayu Savitri
Politeknik Elektronika Negeri Surabaya, INDONESIA
nadiaayusavitri95@gmail.com

Umi Sa'adah
Politeknik Elektronika Negeri Surabaya, INDONESIA
umi@pens.ac.id

Iwan Syarif
Politeknik Elektronika Negeri Surabaya, INDONESIA
iwanarif@pens.ac.id

Abstract— **Intra Uterine Growth Restriction is a condition when the infant fails to grow as expected (smaller) than normal size and can be affected to the safety of pregnant women and the fetus. Data mining can be used as a tool to make prediction and diagnosis of IUGR. This study deals with IUGR dataset from the hospital and applied Naïve Bayes method as one of the potential method in the Data Mining. The diagnosis can be classified based on existing aspects. In the experiment result, the average accuracy value obtained from testing of Naive Bayes is 84%, with precision 52.5%, recall 86.7% and specificity 83.8%, which results show a pretty good performance.**

Keywords— Intra Uterine Growth Restriction, classification, Data Mining, Naive Bayes Classifier

I. Introduction

IUGR (Intra Uterine Growth Restriction) or a small baby during pregnancy is a condition where the baby fails to grow according to expectations (smaller) than normal size when in the womb. In many cases IUGR is caused by a lack of nutritional intake during pregnancy so that maternal weight does not increase according to gestational age. As a result the fetus gets a little extra nutrition which results in a fetal body weight being smaller than normal. Other causes of IUGR are diseases that pregnant women suffer from such as hypertension (high blood pressure), heart, kidney, lung etc. Mothers who have been diagnosed with IUGR will experience a high-risk pregnancy, the baby can be born small and premature so the risk of death is high and also the risk of infection, neurological disorders and heart disease can arise.

There is no knowledge that teaches midwives to do ultrasound (ultrasonography) and there is no permission to do ultrasound at a health center or hospital so that to calculate TFU (High Fundus Uteri) midwives only use manual calculations using fingers. This is one of the causes of the lack of early detection of pregnant women who suffer. IUGR has an effect on the slow handling of pregnant women.

From the above problems, a breakthrough is needed with the application of IUGR diagnosis. The diagnosis can use data mining techniques with the Naive Bayes Classifier. This application serves as a supporting medium in the diagnosis process for pregnant women conducted by midwives where midwives can do the diagnosis process and print the diagnosis results from the examination process of pregnant women. With the application of the IUGR diagnosis, it is expected that it can help midwives to provide diagnosis and early warning of pregnant women with IUGR cases so that a solution or prevention can be prevented from happening worse

II. Related Works

Many researchers have conducted a study about IUGR [1][2][3][4][5][6][7][8][9][10] In the process of developing this diagnosis of IUGR, two studies among those can be explained as follows:

Sharma, D. et.al.[11] discusses the Diagnosis of Pregnancy Complications Using Data Mining Prediction. In their paper said that in Sri Lanka there were 16.3% of babies born under normal weight which could be caused by several factors ranging from maternal illness, lifestyle, and complications in pregnancy. Complications during pregnancy are the most important cause in increasing the risk of maternal and infant death, and are strongly associated with poor outcomes such as miscarriage, death in childbirth, and premature birth. Therefore it is important to assess the level of risk of pregnancy to control pregnancy complications. The authors performed a diagnosis of pregnancy complications by combining 3 (three) data mining techniques including Artificial Neural Network (ANN), Naive Bayes (NB) and Novel Hybrid Algorithm where ANN and Naive Bayes had diagnosis accuracy of 80% and 70%, while Novel Hybrid has 86% accuracy. The diagnosis of pregnancy complications using data mining predictions using ANN, NB and hybrid algorithms, the doctor can only see the results without any detailed results from the diagnosis process.

978-1-7281-7116-6/20/$31.00 ©2020 IEEE

Nagarajan, S. et. al.[12] discusses Application of Data Mining for Diabetes Gestational Diagnosis. In their journal explaining that in India there were 17% of pregnant women detected by GDM (Gestational Diabetes Miletus) caused by hormones that cause an increase in blood sugar content although pregnant women do not have a history of diabetes before. The number of pregnant women who do not know that they have GDM can have an impact on the physical condition of the fetus that becomes bigger and later dangerous in the birthing process which can endanger the lives of pregnant women and the fetus. They applies several data mining techniques namely ID3, Naïve Bayes, C4.5 and Random Tree. And based on his research, the Random Tree became one of the techniques that had high accuracy results and the least amount of error.

In this study, we develop a system that is not only a diagnostic answer to IUGR disease method that can facilitate midwives before determining the diagnosis of IUGR suffered by patients. We also conducted an experiment to test how well this Naïve Bayes method was applied to diagnose IUGR disease in pregnancy by using performance measurement accuracy, precision and recall.

III. METHODOLOGY

In this section we will first explain the dataset used. Then after that the preprocessing is done before the data can be processed using the Naïve Bayes method. Later a discussion of the Bayes method itself and finally how to implement the Naïve Bayes method into IUGR diagnosis.

A. The Dataset Used

The data used in the study were IUGR diagnostic data in the form of pregnancy data for 2015-2016 which amounted to 205 data, which included age, height, weight, weight gain, maternal illness, family history, arm circumference, blood pressure, pulse , TFU (uterine fundus height), gestational age, and risk.

B. Data Preprocessing

The data cleaning process is done by finding and correcting errors or problems in data such as missing values and noisy data. There are several ways to overcome missing values including filling in the mode or average. The mode is used in data that is continuous in scale. Median / mode is used on categorical scale data, the disease of pregnant women. This is done to ensure data quality. In the Gending Health Center pregnancy data there are several data missing values so that the handling is done by finding the average value. The data used is 225 data and there are 25 data that cannot be used. Categorized as data that cannot be used if the number of missing values exceeds 2 attributes. There are 11 lines of data missing values on some attributes of arm circumference, and blood pressure.

Then the transformation process is carried out to convert numerical data into categorical data. In this process, the attributes for blood pressure labels that are still numerical are converted into categorical data with normal blood pressure as long as it is below 140/80 and the determination of the value of the target attribute whether labeled or not. The determination was determined by the midwife at the health centre.

C. Naïve Bayes Method

The process flow from the Naive Bayes Classifier algorithm starts from reading training data, then checking the attributes on the data, numeric type or not. If the attribute is numeric, then the attribute is calculated by the mean and standard deviation. If the attribute is not a numeric type, then the attribute is calculated by the number and probability.

Naïve Bayes is a classification method based on probabilistic theorems that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset.

The Naïve Bayes equation is:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \tag{1}$$

Where:

X : Data with the unknown classes
H : Data hypothesis is a specific class
P(H|X) : Probability hypothesis based on X condition (Posterior probabilities)
P(H) : Hypothesis probability of H (prior probability)
P(X|H) : X Probability based on the condition in the Hypothesis H
P(X) : Probabilities X

D. System Design to Implement Naïve Bayes into IUGR diagnosis

The picture below shows how the system design is designed to implement the Naïve Bayes method used for IUGR diagnosis.

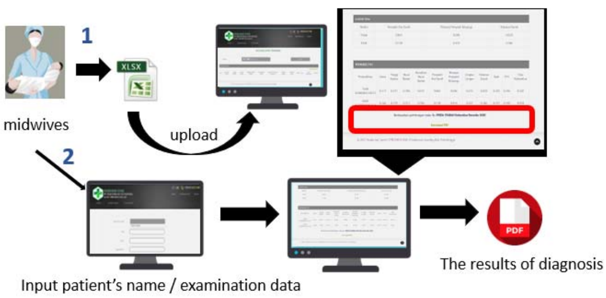

Figure 1. System Design

The system design can be explained as follows:

1) Midwives can upload patient data related to the condition of the patient pregnancy. If midwives want to update the data training she can modify the data and upload it again

2) Midwives input patient's name which was previously input by admin or midwives can input the examination data.

3) and finally, the IUGR or normal system will display diagnostics.

IV. Experiment and Analysis

In this section we will describe the experiments conducted under the implementation of the Naïve Bayes method to IUGR diagnosis. There are 2 (two) related matters, namely performance measurement used to calculate classification performance from the method used. And the second is the sampling technique used to conduct experiments.

A. Software Testing

The features available for use by midwives (medical staff) are as follows:

1. Login Form
2. Midwife's data update
3. Update Patient's data
4. Upload Data Pregnancy as Data Training
5. Patient Data Input (Data Testing)
6. Output Diagnosis (from Naïve Bayes method)
7. Result of Diagnosis (brief explain about patient condition and predictive analysis of IUGR)

From the above features, it is sufficient for users as midwives to use the application as a tool in decision making about IUGR diagnosis.

B. Classification Performance

The following will explain the classification performance used in this study by finding the value of accuracy, precision, and recall values whose formulas are obtained from the confusion matrix in table I.

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

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

$$Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum FP + \sum FN + \sum TN} \quad (4)$$

TABLE I. Precision, Recall and Accuracy

		actual values	
		TRUE	FALSE
prediction values	TRUE	TP (True Positive) *Corect result*	FP (False Positive) *Unexpected result*
	FALSE	FN (False Negative) *Missing result*	TN (True Negative) *Corect absence of result*

C. Classification Performance and Sampling Techniques

For the sampling technique used in this study using k-fold cross validation with the parameter k = 10 so using 10-fold, with test results as shown in Table II.

TABLE II. Sampling method using 10-Fold Cross Validation

K-	Accuracy	Precision	Recall	Specificity
1	65%	22	100	61
2	85%	50	100	82
3	75%	33	67	77
4	90%	67	67	94
5	90%	60	100	88
6	75%	33	67	77
7	95%	75	100	94
8	80%	43	100	77
9	90%	67	67	94
10	95%	75	100	94
95% confidence interval	84% ± 0.062	53 ± 11.8	87 ± 10.6	84 ± 6.84

In the experiment using 10-cross validation, average accuracy value obtained from testing of Naive Bayes method is 84% with 95% confidence interval of 84% ± 0.062. Accuracy itself is defined as the degree of closeness between the predicted value and the actual value. Whereas Precision is the accuracy between the information requested by the user and the answers given by the system. The small percentage of precision does not make the percentage of accuracy from the system small, the small precision value occurs because of many classifications of False Positive diagnosis results (in fact IUGR is not at risk but IUGR is diagnosed). Some midwives stated that such results were better because midwives could provide early warnings before the pregnant women were truly diagnosed with IUGR. Recall or sensitivity is the proportion of the number of diagnoses that are correctly predicted among all possible correct diagnoses. And the specificity is the proportion of the correct diagnosis among all diagnoses that are successfully predicted.

V. Conclusion

In this study, we present the Naïve Bayes method for predictive models for Intra Uterine Growth Restriction (IUGR) diagnosis on pregnancy. The application can be used as a tool for medical personnel in diagnosing abnormalities that occur in pregnant women in connection with the presence of IUGR where fetal failure to grow as expected (smaller) than normal size and can be affected to the safety of pregnant women and the fetus.

In the experiment result shows with 95% confidence the population mean is between 83.9 and 84.1, based on only 10 data from 10-cross validation, the value of accuracy, precision, recall and specificity are 84% ± 0.06, 53 ± 11.8, 87 ± 10.6, and 84 ± 6.84 respectively, which results show a pretty good performance from the Naïve Bayes method.

REFERENCES

[1] Bagi, K. S.; Shreedhara, K. S. Biometric measurement and classification of IUGR using neural networks. 2014 International Conference on Contemporary Computing and Informatics (IC3I), 2014, 27-29 Nov. 2014. p.157-161.

[2] Fayyad, A. M.; Harrington, K. F. Prediction and prevention of preeclampsia and IUGR. Early human development, v. 81, n. 11, p. 865-876, 2005/11// 2005. ISSN 0378-3782. Disponível em: <

[3] Flood, K. et al. The role of brain sparing in the prediction of adverse outcomes in intrauterine growth restriction: results of the multicenter PORTO Study. American Journal of Obstetrics and Gynecology, v. 211, n. 3, p. 288.e1-288.e5, 2014/09/01/ 2014. ISSN 0002-9378.

[4] Gadagkar, A. V.; Shreedhara, K. S. Features Based IUGR Diagnosis Using Variational Level Set Method and Classification Using Artificial Neural Networks. 2014 Fifth International Conference on Signal and Image Processing, 2014, 8-10 Jan. 2014. p.303-309.

[5] Haram, K.; Søfteland, E.; Bukowski, R. Intrauterine growth restriction. International Journal of Gynecology & Obstetrics, v. 93, n. 1, p. 5-12, 2006. ISSN 0020-7292.

[6] Longo, S. et al. IUGR and infections. Early Human Development, v. 90, p. S42-S44, 2014/03/01/ 2014. ISSN 0378-3782

[7] Owen, P. et al. Prediction of intrauterine growth restriction with customised estimated fetal weight centiles. BJOG: An International Journal of Obstetrics and Gynaecology, v. 110, n. 4, p. 411-415, 2003/04/01/ 2003. ISSN 1470-0328..

[8] Pillai, R. N. et al. Prediction of miscarriage in women with viable intrauterine pregnancy—A systematic review and diagnostic accuracy meta-analysis. European Journal of Obstetrics & Gynecology and Reproductive Biology, v. 220, p. 122-131, 2018/01/01/ 2018. ISSN 0301-2115.

[9] Rodriguez, A.; Tuuli, M. G.; Odibo, A. O. First-, Second-, and Third-Trimester Screening for Preeclampsia and Intrauterine Growth Restriction. Clinics in Laboratory Medicine, v. 36, n. 2, p. 331-351, 2016/06/01/ 2016. ISSN 0272-2712.

[10] Wosiak, A.; Zamecznik, A.; Niewiadomska-Jarosik, K. Supervised and unsupervised machine learning for improved identification of intrauterine growth restriction types. 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), 2016, 11-14 Sept. 2016. p.323-329.

[11] Sharma, D.; Shastri, S.; Sharma, P. Intrauterine Growth Restriction: Antenatal and Postnatal Aspects. Clinical Medicine Insights. Pediatrics, v. 10, p. 67-83

[12] Nagarajan, S.; Chandrasekaran, R.M.; Ramasubramanian, P., Data Mining Techniques for Performance Evaluation of Diagnosis in Gestational Diabetes, International Journal of Current Research and Academic Review. 2014.2(10): 91-98.