Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018)
IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2

A Boosting Approach for Maternal Hypertensive Disorder Detection

Neethu Mathew
Dept. of Computer Science and Engineering
Jyothi Engineering College
Thrissur, India
neethumathew425@gmail.com

***Abstract*— Pregnancy is the delicate stage in every woman's life cycle. The changes in the health during this period may lead to risk in pregnancy. That is a high risk pregnancy is one that endangers the women's or baby's health or life. Hypertensive disorders are the most common cause of these pregnancy related complications. In the present era, it's a major challenge in health care because their results in increasing maternal and fetal death. There exist various techniques widely used in data mining for the early identification of health diseases. Better treatment can be provided in the early stage by the early identification of hypertensive disorders. The main intent of this paper is to pinpoint the present complications of a pregnant woman by applying a boosted random forest approach for predicting hypertensive disorders. This classification method helps to predict the risks, diagnose and thereby can decrease maternal and fetal mortality, which remains as a large problem in much of the developing nation.**

Keywords—classification; hypertensive disorder; data mining; pregnancy

I. INTRODUCTION

Pregnancy is an experience of motherhood, which needs much more medical care and attention. Women who were healthy and normal before getting pregnant may experience some risk of problems with them. Hypertension occurs when women have high blood pressure. Having hypertension at any point can lead to health problems, but especially during the cycle of gestation. It could have dangerous outcomes for mother and child, if the adequate care is not taken to control it effectively. The mortality rate in the world has reduced. Women are still dying due to pregnancy complications. Hypertension and its disorders are a leading cause of concern and its prevalence is increasing in developing countries. Therefore, getting identification at an early stage and regular pregnancy care can help to decrease the risk for problems before they become more serious.

There are wide varieties of factors that complicate pregnancy. The maternal and fetal conditions include maternal smoking, blood pressure, mean arterial pressure, body mass index, excessive protein in urine, gestational age, etc. Also women face higher risk due to obesity, woman over the age 40 or under 20, multiple births, woman with diabetes and kidney disease, etc. The maternal hypertensive disorders in pregnancy can be classified into various classes like preeclampsia-eclampsia, gestational hypertension, chronic hypertension and preeclampsia superimposed on chronic hypertension. Preeclampsia (PE) is a condition that pregnant women develop with high blood pressure accompanied by proteinuria. This condition generally appears after the twentieth week of gestation. It may develop into the more severe state called eclampsia. Development of high blood pressure in the second half of pregnancy without proteinuria or other symptoms of preeclampsia is Gestational hypertension (GH). A medical complication developed before pregnancy or before the gestation week 20 is called chronic hypertension. Women with Chronic high blood pressure diagnosed before pregnancy develops a rapid increase in excess protein in the urine or blood pressure results in chronic hypertension with superimposed preeclampsia. The effect of high blood pressure in pregnancy varies based on these disorders and factors.

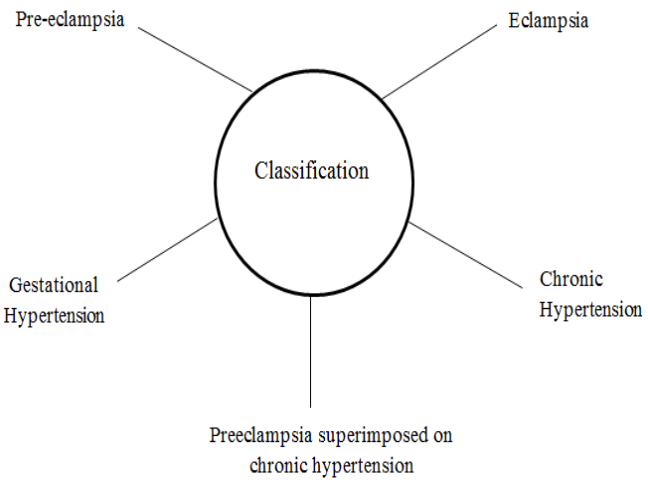

Fig.1. Classification of hypertensive disorders in pregnancy

Data mining techniques helps for the extraction of medical data and their analyzing. Its main aspect is pattern discovery. Classification is a basic task in this field of knowledge discovery in database. The classification in medical diagnostic problem assigns a disease label to the instance. Many researches are going on to find a long term health effect of maternal hypertensive disorders. So it is needed to develop an efficient and better method for identifying, diagnosing and treating women at risk.

This paper is outlined as follows: section II focuses on literature survey. Section III involves the proposed system where the model is explained.

II. LITERATURE SURVEY

This section survey about existing work which make use of data mining methods for finding reliable information about pregnant women having hypertensive diseases. Some of these studies find solutions for healthcare limitations in maternal care domain.

M. W. Moreira, J. J. Rodrigues, A. M. Oliveira and K. Saleem [1] proposed an inference mechanism using data mining technique based on two Bayes based classifiers. And to classify the severity of disorder from its symptoms they compare two Bayesian classifiers named Naïve Bayes and Averaged one-dependence estimators classifier (AODE).

M. W. Moreira, J. J. Rodrigues, A. M. Oliveira and K. Saleem [2] proposed Naïve Bayes classifier and its performance is compared with a classifier called J48 classifier. These two classifiers are good predictors for decision making problems. Their result shows that the tree based classifier has more accuracy than the Naive Bayes classifier.

A probabilistic knowledge based system called Bayesian network has a significant role in healthcare. Diagnosis of the pregnancy disorder preeclampsia [3] is done using the Bayesian network. In this model they use network nodes risk factors, physiological mechanisms and symptoms/exams. And hypertension and proteinuria nodes provide more contribution for decision making.

Severe maternal morbidity can be predicted early using techniques like logistic regression [4], and they proposed tool for the identification and classification of patient's risk. Hidden Markov models [5] based studies are also done to classify the pregnancy disorders. They generated the sequence of observation using the symbolic dynamics for assessing cardiovascular control in hypertensive pregnancy disorder.

In [6], the authors propose a random forest approach for predicting the hypertensive disorders related to pregnancy. They also present the modeling and performance assessment. This tree based classifier shows a good performance and also comparing this with other classifiers.

Development of a predictive model is used for the classification of risk related to hypertension using decision tree [7]. The model was formed by two decision tree based algorithms.ID3 and C4.5 algorithms identified risk of hypertension using risk factors and the ID3 outperformed C4.5.

III. PROPOSED SYSTEM

In the proposed system we have put forward a novel method for detecting the pregnancy disorders. Women with maternal disorders have different changes in her body compared to normal one and this creates threat to her and infant life. In this model focus is on women suffering from maternal hypertensive disorders. Data mining techniques are widely employed to reach a decision or to classify or identify the disorder.

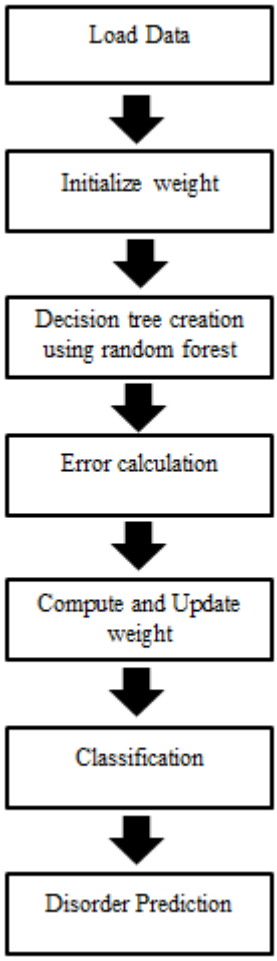

Fig 1: Proposed system architecture

A. Random Forest

The tree based classifier called random forest which is a multitude of decision trees is mainly used for several types of classification tasks. The prediction of several trees is

combined in it essentially. Random features and random inputs produce good results in classification. When we use random forest, a user has to determine only two parameters. They are the number of trees to be used and the number of variables to be randomly selected from the available set. Let the assumed number of training cases or observations in the training set is N. Then N observations are sampled at random, but with replacement. If there are K total input variables, the input variables which are used at random to find the decision at a node of the tree k should be less than K. They create best possible split to develop a decision tree model. Each tree is grown to the largest extent possible.

The process of splitting is performed with the measures like information gain or Gini index. The method of splitting affects trees accuracy. It splits the node into two or more sub nodes. The decision tree grows using a set of samples. Its growth starts from the root node and divides recursively by using splitting rules until it reach a stopping condition. Each individual tree gives a classification, a prediction. The leaves are outputting the prediction. Each of the tree is made with a sample of the data in the average of many decision trees. That is, once each of the trees is classified, the value that has majority among them is taken as the output.

B. Boosted Random Forest

In the proposed model we can apply a boosting technique for the maternal disease detection. It's a method used to boost single trees into strong learning algorithms. Finally it combines the classifiers by letting them vote on a final prediction. We can achieve accuracy by combining the random features with method of boosting. Boosted trees try to improve the model fit over different trees by considering past fits. The classifiers focus on the cases which are incorrectly classified in the last round.

Algorithm is as follows:

1. Input training sample S={X,C}
   Where $X: x_1, x_2 \dots x_N$ are features and
   $C: c_1, c_2 \dots c_M$ are the classes
2. Assign weight W for each features and initialize W=1/N
3. Create subset S_t from training sample S
4. For each feature in S_t calculate information gain IG
5. Find the best split and build the trees
6. Find the class label
7. Calculate the error rate of each decision tree
8. Compute weight based on error rate
9. Updating weights of decision trees
10. Do until stopping condition or reach a particular number of decision trees and weighted decision trees.
11. Output the classifier.

When we introduce boosting algorithm into random forest, decision tree weighting and updating weights are considered. The training set weight is initialized into 1/N where N is taken as the size of training set. It also follows the same procedure of random forest in tree construction. The splitting process can be done with information gain. The information gain IG can be calculated as:

$$IG = E(S_n) - \frac{|S_l|}{|S_n|} E(S_l) - \frac{|S_r|}{|S_n|} E(S_r) \qquad (1)$$

Where the sample set S_n is splitting at node n into S_l andS_r. S_l And S_r are the sample sets at the left and right child node. The splitting process is continued recursively until the information gain is zero or until the stopping criteria .The probability distribution is stored at the leaf node. If the decision tree training is completed then we can estimate the class label. The misclassification rate is expressed as error rate. We compute the error rate of decision tree. Based on the error calculation ER we calculate the weight W.

$$W = \frac{1}{2} log \frac{(M-1)(1-ER)}{ER} \qquad (2)$$

Where ER is the error rate of the decision tree and M is the number of classes. If W > 0 the weight is updated, otherwise the tree is rejected. When an input is given its output is stored in the leaf nodes. That is, they produce the classification result. Class with highest probability is chosen. The hypertensive disorder class prediction will be the weighted majority vote.

The main intent of boosted random forest is to maintain the generality. The boosting method will help to attain good classification performance. With regard to standard input we are able to find individuals as having mild, moderate and severe hypertension and can classify in the correct disorder type.

IV. CONCLUSION

Data mining methods are essential in the medical field for prediction. The area of data mining has established itself as the major part of computer science and contributes a way to improve the decision making process. It has shown significant potential for future improvements. In this proposed work designing an approach called boosted random forest to classify the hypertensive pregnancy disorder for the early prediction. This work can bring a drastic change in medical fields, especially in the maternal care domain by effectively classifying the disorders. As future work we can improve the

performance and accuracy using other classification techniques and expanding size of datasets.

ACKNOWLEDGMENT

I take this opportunity to express my sincere gratitude to all respected personalities who had guided, inspired and helped me in the successful completion of this paper. First and foremost, I express my thanks to The Lord Almighty for guiding me in this endeavor and making it a success.

REFERENCES

[1] M. W. Moreira, J. J. Rodrigues, A. M. Oliveira, K. Saleem, "An inference mechanism using Bayes-based classifiers in pregnancy care," 2016 18th IEEE International Conference on, Munich, Germany, Sep. 14-17, 2016, pp. 1–5

[2] M. W. L. Moreira, J. J. P. C. Rodrigues, A. M. B. Oliveira, K. Saleem, "Performance evaluation of predictive classifiers for pregnancy care," in 2016 IEEE Global Communications Conference: Selected Areas in Communications: E-Health, Washington (DC), USA, Dec. 4-8, 2016, pp. 1–5.

[3] M. W. Moreira, J. J. Rodrigues, A. M. Oliveira, R. F. Ramos, and K. Saleem, "A preeclampsia diagnosis approach using Bayesian networks," in Communications (ICC), 2016 IEEE International Conference on, Kuala Lumpur, Malaysia, May 23-27, 2016, pp. 1–5

[4] Eugenia Arrieta Rodriguez1, Francisco Edna Estrada, William Caicedo, and Juan Carlos Martinez Santos," Early Prediction of Severe Maternal Morbidity Using Machine Learning Techniques", Advances in Artificial Intelligence, pp 259-270,*S*pringer, IBERAMIA 2016

[5] V. Baier, M. Baumert, P. Caminal, M. Vallverd´u, R. Faber, and A. Voss,"Hidden markov models based on symbolic dynamics for statistical modeling of cardiovascular control in hypertensive pregnancy disorders,"IEEE transactions on biomedical engineering, vol. 53, no. 1, pp. 140–143, 2006.

[6] M´ario W. L. Moreira Joel J. P. C. Rodrigues, Antonio M. B. Oliveira, Kashif Saleem, and Augusto J. Venancio Neto, " Predicting Hypertensive Disorders in High-risk Pregnancy Using the Random Forest Approach" ,2017 IEEE

[7] Idowu Peter Adebayo, "Predictive model for the classification of hypertension risk using decision tree algorithm",American Journal of Mathematical and Computer Modelling, vol. 2,no. 2, 2017, pp.48-59.

[8] Cheng-Ding Chang , Chien-Chih Wang , Bernard C. Jiang "Using data mining techniques for multi-diseases prediction modeling of hypertension and hyperlipidemia by common risk factors" , ELSEVIER , Expert Systems with Applications 38 (2011) 5507–5513

[9] Lakshmi.B.N, Dr.Indumathi.T.S, Dr.Nandini Ravi," A study on C.5 Decision Tree Classification Algorithm for Risk Predictions during Pregnancy" International Conference on Emerging Trends in Engineering, Science and Technology ICETEST, Science Direct**,** ELSEVIER,2015.

[10] P. Anastasopoulos, K. H. Nikolaides, C. N. Schizas, and K. C. Neokleous, "Neural networks to estimate the risk for preeclampsia occurrence," in 2009 International Joint Conference on Neural Networks, Atlanta (GA), USA, Jun. 14-19, 2009, pp. 2221–2225.

[11] Yohei Mishina,Masamitsu Tsuchiya and Hironobu Fujiyoshi,"Boosted random forest",International conference on computer vision theory and applications,VISAPP,2014

[12] Laura A. Magee, A. Pels, M. Helewa,E. Rey,P. von Dadelszen et al.,"Diagnosis,evaluation,and management of the hypertensive disorders of pregnancy",Pregnancy Hypertension: An international Journal of Women's Cardiovascular Health, vol.4, no. 2,pp. 105-145,2014