

PAPER • OPEN ACCESS

The Application of Machine Learning Models in Fetal State Auto-Classification Based on Cardiotocograms

To cite this article: Gongao Xue 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **310** 052007

View the [article online](#) for updates and enhancements.

The Application of Machine Learning Models in Fetal State Auto-Classification Based on Cardiotocograms

Gongao Xue¹

¹Sino-German College of Technology, East China University of Science and Technology, Shanghai, 200237, P. R. China

Email: smofl7@163.com

Abstract.Cardiotocography (CTG) is widely used by obstetricians in accessing the physical condition of a fetus during pregnancy, for it provides obstetricians with the data regarding the fetal heartbeat and the uterine contractions which helps determine whether the fetus is pathologic or not. Traditionally obstetricians analyze data from CTG artificially, which is both time consuming and lack of reliability. For this reason, developing a fetal state auto-classification model is necessary, for it can not only reduce the time for diagnosing but also save medical resources. With machine learning developing rapidly nowadays, it has been widely applied in areas like biology and medicine to solve various problems. In the condition of fetal state classification, we apply neural network and random forest to analyze the cardiotocographic data from the UCI Repository. Since there is high imbalance in our data, method of weighing has also been applied to optimize our model. Random forest outperforms neural network in terms of accuracy in classifying types of fetuses, which achieves 88.84% and 91.85% accuracy on the training and testing set respectively.

1. Introduction

Cardiotocography (CTG), also called electronic fetal monitoring (EFM), is a continuous record of fetal heartbeat and the uterine contractions obtained via an ultrasound transducer placed on the mother's abdomen. It is widely used in pregnancies to provide the obstetricians with information about the health conditions of fetuses, for example, by accessing cardiotocography an obstetrician can tell whether a fetus is suffering from lack of oxygen. In this way fetal intervention is given in time to prevent the fetus from death or other neurological diseases [1]. However, the interpretation of cardiotocography is traditionally done by obstetricians and proved to be deficient sometimes, which might lead to unnecessary surgical intervention [2]. Other studies show that there is no significant difference in perinatal mortality or potentially preventable deaths between traditional CTG and no CTG, but computerized CTG showed a significant reduction in perinatal mortality when comparing with traditional CTG [3]. These facts suggest that standardized interpretation given by computers is necessary. Only in this way, the information from CTG can be used most efficiently.

Nowadays, with human intelligence booming, it has been applied in extensive areas. It is defined as an ability of a system to interpret and learn from external data, and achieve specific goals and tasks set by human through adaptations [4]. Machine learning is a subset of artificial intelligence and a study of algorithms and statistical models. Instead of requiring explicit programs to accomplish specific tasks, machine learning demands training data to build mathematic models so it can make predictions and solve problems [5]. Since it can handle a mass of data efficiently, it has been used as an alternative to traditional solutions of many problems. Models such as random forest and neural network are both



classic and effective machine learning models that have been used in this area, which will also be applied in our essay.

Neural network is an imitation of biological nerve system. In the case of biological nerve system, nerve cells are connected to each other. When a nerve cell is excited, it will release packets of chemicals to other cells via synapses, so the electric potentials of other cells are changed. Once the electric potential of a nerve cell is higher than its threshold, it will also be excited, and repeat the progress mentioned above [6]. In this way, a signal is spread in a biological nerve system. Neural network is inspired by the net of nerve cells inside a brain of a creature, it contains a group of nodes connected to each other, as presented in Figure 1. Each of the nodes represents an artificial neuron, while the line between two nodes represents the value passed from one artificial neuron to the other. Each of the nerve cell receives input signals with weight, and the total value will be compared the threshold to create an outcome [7]. Criticisms exist due to it requires too much training and computing resources, however, for its powerful function of modeling none linear progress, it has still been widely applied in medical diagnosis, object recognition, finances and many other fields [8].

Random forest, an ensemble learning method, is first developed by Tin Kam Housed. Nowadays, it has been widely used for classification, regression and other tasks [9]. The first step of building a random forest is to take a random sampling with replacement from the original data set to build sub-data sets. These sub-data sets will then be used to build decision trees. Once there is new data needed to be predicted, predictions will be given from these decision trees, and the result of ensemble learning will be produced by voting, which means the minority is subordinate to the majority. By constructing a large number of decision trees at training time, the final result is given by finding the mean prediction, in this way, accuracy of the model is improved substantially. Random forest model has many advantages comparing with other models: it is extremely simple and easy to realize, and do not cost much computing resource. It is also the reason why we apply it to build our auto-classification model.

The rest of the article is organized as follows. In Section 2, Data from UCI Machine Learning Repository will be described in detail while in Section 3, two models including neural network and random forest will be used to build our auto-classification models. The results from both models will be presented respectively. These results will be compared and analyzed in Section 4. Finally, conclusions will be drawn based on all the information mentioned above, and also, advantages and disadvantages of both models will be given.

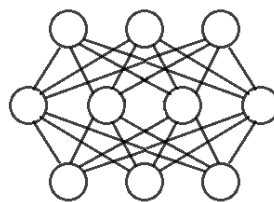


Figure 1

2. Data Description

The cardiotocographic data set are from UCI Machine Learning Repository, which is public available at <https://archive.ics.uci.edu/ml/datasets/cardiotocography>. It consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features of 2126 fetuses based on cardiotocograms. In this data set 23 attributes are measured, as presented in Table 1. Fetuses are classified by expert obstetricians into three types based on the features mentioned above. These three types are represented by fetal state class code N, S and P, which means that the fetus is normal, suspect and pathologic respectively. All the 2126 samples are divided into two parts as training and testing data, as presented in Table 2. From Table 2 we can also see that cardiotocographic data set is highly unbalanced, with fetuses of N type account for 77.84 percent of the total while fetuses of P type make up only 8.28 percent. But as a matter of fact, the accuracies of P type and S type are far more important than that of N type, so we are willing to sacrifice

the accuracy of N type to some extent to improve the accuracy of S and P types. Therefore, method of weighing will be applied to optimize our model, which will be presented in detail in the following section.

Table 1: Attributes of Cardiotocographic Data Set

Attribute		
1	LBE	baseline value
2	LB	baseline value
3	AC	accelerations
4	FM	foetal movement
5	UC	uterine contractions
6	ASTV	percentage of time with abnormal short-term variability
7	mSTV	mean value of short-term variability
8	ALTV	percentage of time with abnormal long-term variability
9	mLTV	mean value of long-term variability
10	DL	light decelerations
11	DS	severe decelerations
12	DP	prolonged decelerations
13	DR	repetitive decelerations
14	Width	histogram width
15	Min	low freq. of the histogram
16	Max	high freq. of the histogram
17	Nmax	number of histogram peaks
18	Nzeros	number of histogram zeros
19	Mode	histogram mode
20	Mean	histogram mean
21	Median	histogram median
22	Variance	histogram variance
23	Tendency	histogram tendency: -1=left asymmetric; 0=symmetric; 1=right asymmetric

Table 2: Constituent of Training and Testing Data

	N type	S type	P type	Total
Training	1153	205	130	1488
Testing	502	90	46	638
Total	1655	295	176	2126

3. Models

3.1. Neural Network

Neural network, an imitation of biological nerve system, is a group of nodes connected to each other. Like the working principle of nerve cells, when the sum of incoming signals is higher than the threshold of an artificial neuron, the artificial neuron will be activated and pass the signal down to other neurons [7]. For its capability of modeling non-linear progress, we apply it to build our auto-classification model.

A neural network consists of an input layer, several hidden layers and an output layer. The number of hidden layers and the nodes in each of the layers are both very important parameters in a neural network model. Too many nodes and layers might enable the model to fit the training data well, but it will perform badly when it comes to testing data, which is called overfitting [10]. While on the contrary, too few nodes and layers might lead to under fitting. After several attempts to minimize the unavoidable overfitting and under fitting phenomena, we finally come up with a neural network with 4 hidden layers,

which has relatively good generalization ability. There are 21 nodes in the input layer and 3 nodes in the output layer, and the 4 hidden layers consist of 32, 16, 8, 4 nodes respectively.

Batch gradient descent and least mean logarithm squared are applied when training our model. The loss of both training data and testing data with different iterations is calculated and plotted in Figure 2, and confusion matrix of testing data in neural network model is presented in Table 3. As which can be seen from Table 3, the error rate of predicting fetuses of N type is as low as 2.99%, but that of S and P type are unsatisfying which reach 41.1% and 43.48% respectively due to the imbalance of our data. In order to improve the accuracy of the model, data of the last two types have to be used more efficiently. Therefore, weighing method is introduced to improve our model.

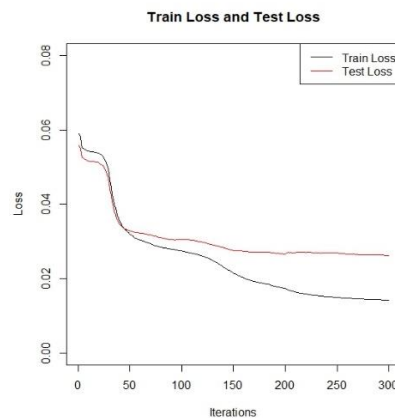


Figure 2

Table 3: Confusion Matrix of Testing Data in Neural Network Model

	N type	S type	P type	Error Rate
N type	487	14	1	2.988%
S type	33	53	4	41.111%
N type	9	11	26	43.478%

3.2. Weighted Neural Network

The accuracy of neural network model in Section 3.1 indicates that there is serious problem in our model. In real applications, we care more about determining fetuses of N type and P type correctly, so that fetal intervention can be given in time to prevent deaths or other unwilling conditions from happening. On the other hand, accuracy in predicting fetuses of N type is relatively unimportant. But as a matter of fact, our model is much more successful in predicting fetuses of N type due to the imbalance of our data. We handle this problem by assigning different weights to different types to improve our model. Through several attempts to improve the accuracy of the last two types, the weight of N, S and P type are set as 0.1, 0.5 and 0.4 respectively, so that more samples in S type and P type are involved. The loss of improved neural network model is plotted in Figure 3, and the confusion matrix of testing data is given in Table 4. Comparing with the result from the original neural network model, the loss of both training and testing data is higher due to the severe decrease in the accuracy of N type, but there is a substantial increase in the accuracy of the last two types, which meets the need in real life. However, even the accuracy of the improved neural network model is still not satisfying enough to be implemented. As a result, random forest model is introduced in the next section.

Table 4: Confusion Matrix of Testing Data in Weighted Neural Network Model

	N type	S type	P type	Error Rate
N type	454	45	3	9.561%
S type	19	69	2	23.333%
N type	5	12	29	36.956%

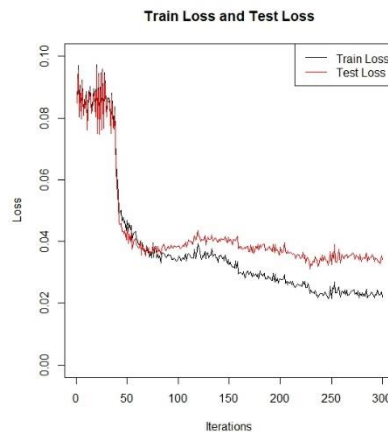


Figure 3

3.3. Random Forest

Random forest is a model commonly used for classification and regression, it constructs a number of trees and output mean prediction of individuals [9][11]. Trees are built based on the sub-data sets which are randomly chosen from the original data set. They vote to get the final result which enables the random forest model to achieve high accuracy in multiple tasks. For its high accuracy in predicting and good performance in handling big data, we also apply random forest model to improve our auto-classification model.

Overfitting phenomena does not exist in random forest [12], but this does not mean more trees lead to a model with better performance. Out-of-bag error (OOB error) is a commonly used method to evaluate the prediction error of random forest model utilizing bootstrap aggregating to training data [13]. OOB error of different number of trees is presented in Figure 4. Based on Figure 4, the number of trees is set as 100 for it does not cost much computing resource and achieves a relatively small OOB error. The confusion matrix of random forest model with 100 trees is presented in Table 5. It sharply reduces the error rate of testing data even comparing with weighted neural network model. However, similar to the conclusion drawn from Section 3.2, the model can still be improved by assigning weights to different types.

Table 5: Confusion Matrix of Testing Data in Random Forest Model

	N type	S type	P type	Error Rate
N type	492	9	1	1.992%
S type	25	64	1	28.889%
P type	4	1	41	10.870%

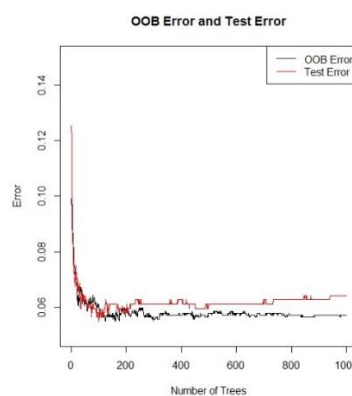


Figure 4

3.4. Weighted Random Forest

Conclusion from Section 3.2 indicates that in order to meet the need of real life, accuracy of N type can be slightly sacrificed to improve accuracy of the other two types. After several attempts to minimize the error rate of S type and P type, the weight of N, S and P type are set as 1, 2.05 and 1.3 respectively. The relation between number of trees and OOB error of weighted random forest model is plotted in Figure 5. For the purpose of reducing OOB error, the number of trees is set as 430. Based on all the information mentioned above, the confusion matrix is computed and presented in Table 6. From both Figure 5 and Table 6 we can clearly see that both OOB error and error rate of N type raised comparing with the original unweighted model, but the error rate of S type and P type reduced to 8.28% and 6.15% respectively, which means that weighted random forest model is a suitable model for fetal state auto-classification. Two other parameters are also closely related to out-of-bag samples: variable importance and Gini importance. Both of them evaluate the influence of a certain variable to the whole model, as presented in Figure 6. From it we can clearly see that AC, ASTV and ALTV are the three variables with highest variable importance, while ALTV, ASTV and Mean are the three variables with highest Gini importance, which indicates these variables has better predictive ability.

Table 6: Confusion Matrix of Testing Data in Weighted Random Forest Model

	N type	S type	P type	Error rate
N type	460	34	8	8.367%
S type	5	83	2	7.778%
P type	1	2	43	6.522%

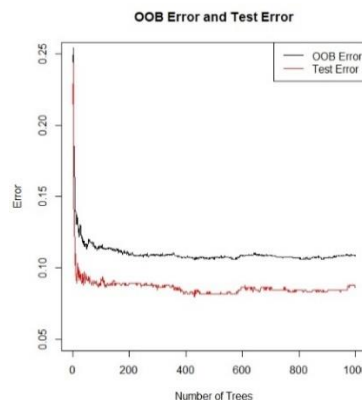


Figure 5

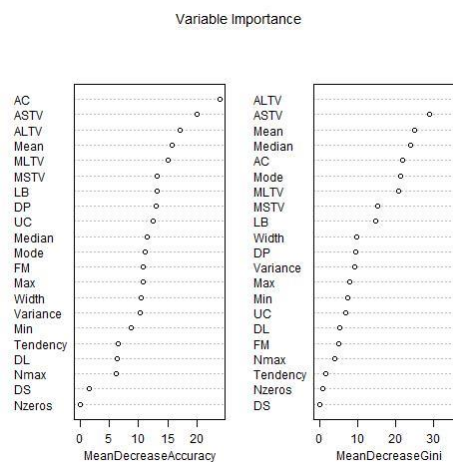


Figure 6

4. Summeries and Discussions

In order to get early detection of fetal hypoxia, cardiotocograms has been widely used in obstetrics practice. The accurate analysis of cardiotocograms is of great importance to further treatment, therefore, standardized machine learning method for analyzing is in great demand [14]. We established four models in the whole process, including neural network, weighted neural network, random forest and weighted random forest. Based on the results mentioned above, random forest outperforms neural network, while weighted model shows better performance comparing with unweighted one in this case. Finally, by applying weighted random forest model, we successfully reduce the error rate to 8.150% (the error rate of S type and P type are 7.778% and 6.522% respectively). From the weighted random forest model, we also come up with the variables with better predictive ability: AC, ASTV, ALTV and Mean are important variables that should be seriously taken into consideration when determining the health condition of a fetus. Although there is still large room for improvement, our model can work well as an alternative to artificial analysis, which will improve both efficiency and accuracy in diagnosing.

References

- [1] Esra, M.K., Turgay, I., (2014) Analysis of Cardiotocogram Data for Fetal Distress Determination by Decision Tree Based Adaptive Boosting Approach. *Journal of Computer and Communications*, 2:32-37.
- [2] Steer, P.J., (2008) Has electronic fetal heart rate monitoring made a difference. *Seminars in Fetal & Neonatal Medicine*, 13(1):2-7.
- [3] Grivell, R.M., Alfirevic, Z., Gyte, G.M., (2010) Antenatal cardiotocography for fetal assessment *Cochrane Database. Cochrane database of systematic reviews*, 12(1):CD007863.
- [4] Kaplan, A., Haenlein, M., (2018) Siri, siri, in my hand: who's the fairest in the land? on the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*.
- [5] John, R. K., Forrest, H. B., David, A., Martin, A. K., (1996) Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. *Artificial Intelligence in Design*, 96:151-170.
- [6] Marcel, V.G., Sander, B., (2017) Artificial Neural Networks as Models of Neural Information Processing. *J. Frontiers in Computational Neuroscience*, 11:114.
- [7] Zhou Z.H., (2016) *Machine Learning*. Tsinghua University Press, Beijing.
- [8] French, J., (2016) *The Time Traveler's CAPM*. Social Science Electronic Publishing, 1-16.
- [9] Ho, T.K., (1998) The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832-844.
- [10] Li, H., (2016) *Statistical learning method*. Tsinghua University Press, Beijing.
- [11] Ho, T.K., (1995) Random decision forests. In: *Proceedings of the 3rd International Conference on Document Analysis and Recognition*. Montreal. pp. 278-282.
- [12] Hastie, T., Friedman, J., Tibshirani, R., (2010) *The Elements of Statistical Learning*. *Technometrics*, 45(3):267-268.
- [13] James, G., Witten, D., Hastie, T., (2013) *An Introduction to Statistical Learning: with Applications in R*. Springer, New York.
- [14] Yilmaz, E., (2016) Fetal State Assessment from Cardiotocogram Data Using Artificial Neural Networks. *J. Journal of Medical and Biological Engineering*, 36:820-832.