Research Article

Modeling fetal morphologic patterns through cardiotocography data: Decision tree-based approach

R. S. Kamath[1]*, R. K. Kamat[2]

ABSTRACT

Objective: The present research aims at decision tree (DT) modeling of fetal morphologic patterns by exploring cardiotocography (CTG). CTG consists of fetal heart rate and topographic measurements and is used for the verification of fetal health. **Materials and Methods:** In the present study, we have carried out DT modeling for CTGs data classification based on fetal morphologic patterns. Decision tree model is the most commonly used data mining technique for classification and prediction. The dataset employed in the present study comprises ten classes of morphologic patterns with a sample size of 2126 records. The optimum decision tree model is derived by tuning parameters such as min split, min bucket, max depth, and complexity. This model entails recursive partitioning approach implemented in the "rpart" package of R. The performance of the model is evaluated in terms of mean square error estimate of error rate. **Results:** Thus, derived decision tree model leads to values for tuning parameters such as min split, min bucket, max depth, and complexity are 20, 7, 30, and 0.01, respectively. The 1488 observations from the inputted dataset are considered for the construction of the tree. Root node error is 0.7211. Thus, derived DT model efficiently classifies validation data with very less error. **Conclusion:** The result suggests that the DT modeling has the potential to exhibit as the best tool for modeling of CTG data.

KEY WORDS: Cardiotocography, Classification, Decision tree, Fetal morphologic pattern, Rattle

INTRODUCTION

Cardiotocography (CTG) indicates fetal health in terms of fetal heart rate (FHR), uterine contraction, and fetal movement and is taken from 27 weeks of pregnancy.[11] The CTG analysis done by obstetricians during FHR pattern observation helps in recognizing fetal state such as physiological, suspect, and pathological.[10] Thus, prosperity of embryo can be visualized and taken care in advance.[12]

Literature review divulges that there are a few reported researches of utilizing the machine learning approaches for the study of CTG data.[7-9] Kamath and Kamat have presented random forest (RF) modeling of fetal morphologic patterns for CTG data and derived optimum RF model by tuning different properties.[2] A study by Karabulut and Ibrikci explains machine learning techniques for analyzing CTG data.[3] Their study explains Decision tree system with accuracy 95.01%.[3] Yet another paper by Tomas *et al*. have explained RF model for automatic recognition of three assorted fetal states.[4] This framework supports decision system as a part of pre-birth care. Sundar *et al*. explained CTG data classification by designing artificial neural network model.[5] This classifier was capable of classifying fetal states with less error. The performance of aforesaid model was measured in terms of precision, recall, F-score, and rand index. Yet another paper by Kamath and Kamat reports decision tree modeling of proteins expression levels for down syndrome.[6]

In the background of the research portrayed above, the present study demonstrates decision tree modeling of fetal morphologic pattern using CTG Data. The dataset with 2126 observations of CTGs is selected for the analysis.[1] These data contain FHR, uterine contraction, and fetal movement measurements. The present work is carried out in Rattle. Rattle is a graphical data mining application built on the statistical language R. The study derives DT model that classifies CTG data with very less error.

Access this article online	
Website: jprsolutions.info	**ISSN:** 0974-6943

[1]Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research, Kolhapur, Maharashtra, India, [2]Department of Electronics, Shivaji University, Kolhapur, Maharashtra, India

***Corresponding author:** R. S. Kamath, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research, University Road, Kolhapur, Maharashtra, India. Phone: +91-9096004515. E-mail: rskamath@siberindia.edu.in

Received on: 19-09-2017; Revised on: 22-10-2017; Accepted on: 25-11-2017

The rest of paper is designed as follows: Introduction followed by the materials and methods utilized in the present study. Then, the third section summarizes the computational details, results, and discussions of the DT model. The conclusion at the end justifies the suitability of DT model for fetal morphologic patterns.

MATERIALS AND METHODS

The CTG dataset for the present study is taken from UCI data repository.[1] It consists of FHR, uterine contraction, and fetal movement measurements. This dataset contains 2126 samples of fetal CTGs classified into ten classes of morphologic patterns and three classes of fetal state.[1] This classification was done by expert obstetricians. Proposed research reveals decision tree-based classification of CTGs data into ten classes of morphologic patterns. Figure 1 shows a number of observations corresponding to these classes described in the dataset.

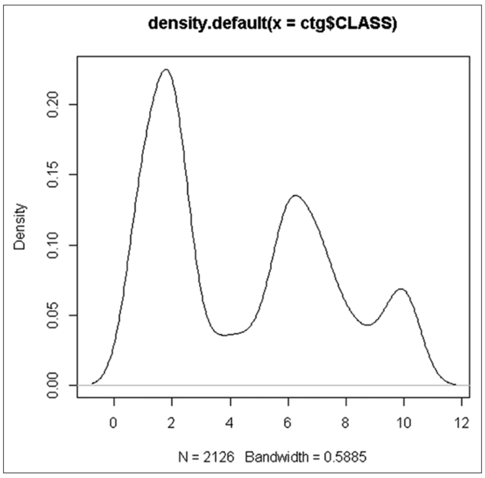

Figure 1: Cardiotocography data projection

In the present study, we have carried out DT modeling for CTGs data classification into ten classes of fetal morphologic patterns. The decision tree is an advantageous and proficient representation of information. It begins with a solitary root node that part into different branches, prompting to further nodes, each of which may additionally part or else end as a leaf node.[14] Connected with each non-leaf node will be a test or question that figures out which branch to take after. The leaf nodes contain the choices.

The reported work is carried out in rattle environment.[13] The model is of type multi-input and single-output arrangement. The model has 21 inputs, namely, values of FHR and UC features, whereas morphologic pattern class is an output variable. Figure 2 shows that decision tree derived in the present investigation represents fetal morphologic pattern class for FHR and UC features. That the root node of decision tree tests AC value ≥0.08543558 continues down to the left side of the tree, otherwise right side of the tree. The next test down this right side of the tree is DL value. Thus, it proceeds and will be able retrieve class value for fetal morphologic pattern class. Table 1 gives details of tuning parameters varied in rattle to obtain optimized decision tree model for classification. The performance of the model is measured in terms of mean square error between predicted output and actual output.

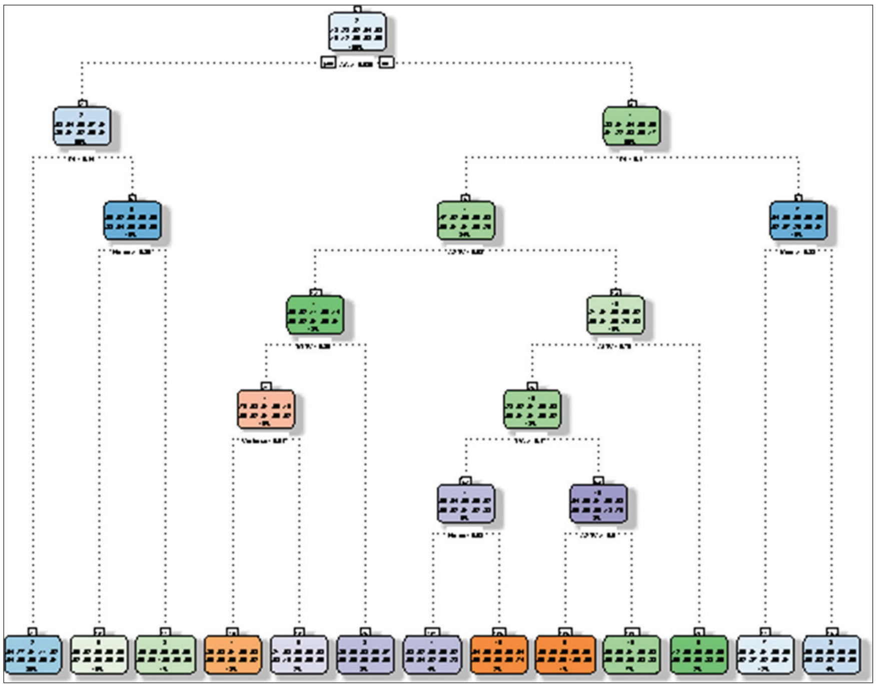

Figure 2: Decision tree for fetal morphologic pattern classification

COMPUTATIONAL DETAILS, RESULTS, AND DISCUSSIONS

The experiment is carried out in Rattle data mining platform to obtain optimum model structure by varying configuring tuning parameters of DT model.[13] Figure 3 shows DT model thus obtained for classification. This textual view highlights the key interface widgets of decision tree construction. Node and split details of corresponding decision tree are shown in Figure 4. The DT shown in Figure 2 translates to the rules, where each rule corresponds to one path through the tree, starting at the root node and ending at a leaf node.

Thus, derived decision tree model leads to values for tuning parameters such as min split, min bucket, max depth, and complexity are 20, 7, 30, and 0.01, respectively. Figure 3 explains performance evaluation of the model in terms of iterations and associated change in the accuracy of the model as new levels are added to the tree. The complexity parameter value reveals that

Table 1: Decision tree configuration

Tuning parameter	Description	Value
Min split	Lowest number of observations for a node resulting from a split before a performing spilt	20
Min Bucket	Minimum number of observations allowable in any leaf node of the decision tree	7
Max depth	Maximum deepness of any node of the final tree	30
Complexity	Controlling parameter to manage the size of the decision tree	0.01

```
Classification tree:
rpart(formula = CLASS ~ ., data = crs$dataset[crs$train, c(crs$input,
    crs$target)], method = "class", parms = list(split = "information"),
    control = rpart.control(usesurrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:
[1] AC       ALTV     ASTV     DL       Mean     MLTV     Nmax     UC
[9] Variance

Root node error: 1073/1488 = 0.7211

n= 1488

         CP nsplit rel error  xerror     xstd
1  0.220876      0   1.00000 1.00000 0.016122
2  0.185461      1   0.77912 0.77912 0.017837
3  0.137931      2   0.59366 0.59832 0.017805
4  0.068034      3   0.45573 0.46132 0.016939
5  0.046598      4   0.38770 0.39422 0.016216
6  0.027959      5   0.34110 0.34856 0.015595
7  0.013979      6   0.31314 0.33178 0.015337
8  0.012116      7   0.29916 0.32339 0.015202
9  0.011184      9   0.27493 0.31500 0.015063
10 0.010252     10   0.26375 0.31314 0.015031
11 0.010000     12   0.24324 0.30289 0.014854
```

Figure 3: Summary of the decision tree model for classification (built using “rpart”).

```
Summary of the Decision Tree model for Classification (built using 'rpart'):

n= 1488

node), split, n, loss, yval, (yprob)
      * denotes terminal node

  1) root 1488 1073 2 (0.18 0.28 0.024 0.038 0.034 0.16 0.12 0.051 0.032 0.089)
    2) AC>=0.08543558 749  344 2 (0.029 0.54 0.004 0.075 0.011 0.3 0.013 0.019 0 0.0053)
      4) DL< 0.1448468 517  117 2 (0.043 0.77 0.0058 0.11 0.015 0.044 0 0.0039 0 0.0077) *
      5) DL>=0.1448468 232   28 6 (0 0.022 0 0.0043 0 0.88 0.043 0.052 0 0)
       10) Nmax>=0.2598425 221   17 6 (0 0.023 0 0.0045 0 0.92 0.045 0.0045 0 0) *
       11) Nmax< 0.2598425 11    0 8 (0 0 0 0 0 0 0 1 0 0) *
    3) AC< 0.08543558 739  492 1 (0.33 0.014 0.045 0 0.057 0.0068 0.22 0.084 0.064 0.17)
      6) DL< 0.1043382 504  267 1 (0.47 0.018 0.063 0 0.083 0.002 0.014 0.006 0.093 0.25)
       12) ASTV< 0.6333333 274   86 1 (0.69 0.022 0.11 0 0.14 0.0036 0.018 0.0073 0 0.015)
         24) MLTV< 0.2889546 228   54 1 (0.76 0.026 0.0088 0 0.15 0.0044 0.022 0.0088 0 0.018)
           48) Variance< 0.01672862 194   27 1 (0.86 0.026 0.01 0 0.082 0 0 0 0 0.021) *
           49) Variance>=0.01672862 34   16 5 (0.21 0.029 0 0 0.53 0.029 0.15 0.059 0 0) *
         25) MLTV>=0.2889546 46   17 3 (0.3 0 0.63 0 0.065 0 0 0 0 0) *
       13) ASTV>=0.6333333 230  108 10 (0.21 0.013 0.0043 0 0.022 0 0.0087 0.0043 0.2 0.53)
         26) ALTV< 0.7527473 196   74 10 (0.23 0.015 0.0051 0 0.026 0 0.01 0.0051 0.087 0.62)
           52) UC>=0.1708328 80   40 1 (0.5 0.037 0 0 0.025 0 0.025 0.012 0.025 0.38)
            104) Nmax< 0.6811024 57   18 1 (0.68 0.018 0 0 0.018 0 0.035 0.018 0 0.23) *
            105) Nmax>=0.6811024 23    6 10 (0.043 0.087 0 0 0.043 0 0 0 0.087 0.74) *
           53) UC< 0.1708328 116   24 10 (0.043 0 0.0086 0 0.026 0 0 0 0.13 0.79)
            106) ASTV>=0.9 12    0 9 (0 0 0 0 0 0 0 0 1 0) *
            107) ASTV< 0.9 104   12 10 (0.048 0 0.0096 0 0.029 0 0 0 0.029 0.88) *
```

Figure 4: Node and split details of decision tree

```
Error matrix for the Decision Tree model on normal_CTG.csv [validate] (counts):

      Predicted
Actual  1  2  3  4  5  6  7  8  9 10
     1 40  8  5  0  6  1  1  0  0  2
     2  1 80  0  0  1  2  0  1  0  1
     3  0  2  4  0  0  0  0  0  0  0
     4  0 11  0  0  0  0  0  0  0  0
     5  2  2  0  0  0  0  0  0  0  7
     6  0  7  0  0  0 38  1  0  0  0
     7  0  0  0  0  0  1 31  1  0  0
     8  1  2  0  0  0  1  2 10  0  0
     9  1  0  0  0  0  0  0  0  9  0
    10  3  1  0  0  1  0  2  0  0 29
```

Figure 5: Error matrix for validation data

as the tree splits into more nodes, the algorithm to stop partitioning since the error rate is not improving. It is also observed that cross–validation error is reduced. The 1488 observations from the inputted dataset are considered for the construction of tree. Root node error is 0.7211.

Thus, derived DT model efficiently classifies validation data with very less error. Figure 5 shows the error matrix for the decision tree model on validation data. Result concludes that DT modeling is a suitable technique since the result is more precise.

CONCLUSION

Decision tree model for fetal morphologic pattern classification through CTG data is explained. A decision tree model is a standout among the most widely recognized data mining models. The revealed investigation delineates ideal decision tree architecture accomplished by tuning parameters such as min split, min bucket, max depth, and complexity. DT model, consequently determined, is straightforward and involves recursive apportioning approach executed in the rpart package. Result presumes that DT expectation is an appropriate approach since the subsequent examination is more accurate.

REFERENCES

1. Lichman M. UCI Machine Learning Repository, School of Information and Computer Science. Irvine, CA: University of California; 2013. Available from: http://www.archive.ics.uci.edu/ml. [Last retrieved on 2016 Jul 01].
2. Kamath RS, Kamat RK. Modelling fetal morphologic patterns through cardiotocography data: A random forest based approach. J Pharm Biol Chem Sci 2016;7:2449.
3. Karabulut EM, Ibrikci T. Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach. J Comput Commun 2014;2:32-7.
4. Tomas P, Krohova J, Dohnalek P, Gajdos P. Classification of cardiotocography records by random forest. Telecommunications and Signal Processing (TSP), 36th International Conference; 2013. p. 620-923.
5. Sundar C, Chitradevi M, Geetharamani G. Classification of cardiotocogram data using neural network based machine learning technique. Int J Comput Appl 2012;47:19-25.
6. Kamath RS, Kamat RK. Modeling mice down syndrome through protein expression: A decision tree based approach. Res J Pharm Biol Chem Sci 2016;7:1193-9.
7. Magenes G, Signorini MG, Arduini D. Classification of Cardiotocographic Records by Neural Networks, Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks; 2000. p. 637-41
8. Sahin H, Subasi A. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. Appl Soft Comput 2015;33:231-8.
9. Czabanski R, Jezewski M, Wrobel J, Horoba K, Jezewski J. A neuro-fuzzy approach to the classification of fetal cardiotocograms. In: 14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics. Berlin, Heidelberg: Springer; 2008.
10. Alfirevic Z, Devane D, Gyte GM. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst Rev 2006;1:CD004116.
11. Ayres-de-Campos D, Bernardes J, Garrido A, Marques-de-Sa J, Pereira-Leite L. Sis Porto 2.0: A program for automated analysis of cardiotocograms. J Mater Fetal Neonatal Med 2000;9:311-8.
12. Martinez AM, Kak AC. PCA versus LDA, pattern analysis and machine intelligence. IEEE Trans 2001;23:228-33.
13. Graham W. Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery. New York: Springer; 2011.
14. Kamath R, Kamat R. Educational Data Mining with R and Rattle. Netherland: River Publishers Series in Information Science and Technology, River Publishers; 2016.