

A Classification Approach to Fetal Cardiotocography Dataset using R

R. Hepzibah^a, A. Hepzibah Christinal^a, R. Jayanthi^a, S. Jebasingh^a, D. Abraham Chandy^b, Chandrajit Bajaj^c

^a Department of Mathematics, Karunya University, Coimbatore, India

^b Department of Electronics and Communication Engineering, Karunya University, Coimbatore, India

^c Computational applied Mathematics chair in visualization, Institute for Computational Engineering and Sciences, University of Texas, Austin

Abstract

Electronic fetal heart monitoring to check fetal status during pregnancy is common. Cardiotocography is a technique for assisting obstetricians in obtaining clear details during the time of childbirth as a method of monitoring the health condition, especially in pregnant women who are at risk of difficulties. This paper deals with the classification of the fetal cardiotocography dataset using R. The supervised machine learning-based approach is applied for the categorization of fetal datasets. It is classified as normal, suspect, and pathologic based on the random forest classifier. It produces an accuracy of 99.94% in training and 93.57% in testing which is found to be a better performance. It also provides the best results in terms of sensitivity, and specificity in the classification of normal, suspect, and pathology in both training and testing datasets. It is found that this method provides a greater accuracy compared to all other methods.

Keywords

Machine learning, Random Forest classifier, cardiotocography, fetal heart rate

1. Introduction

Machine learning is an advancing field in the research of Engineering and computer science and various algorithms of machine learning play the main role in the medical field. It also helps in the computation of the image features which helps in the classification and better detection of diseases [1]. It helps in learning the empirical data and making decisions accurately using complex algorithms [2]. Supervised learning which includes regression, classification, and reinforcement learning is the general classification of machine learning. The clustering, blind source estimation, and density estimation come under supervised learning. and the information systems and the semi-supervised classification are part of semi-supervised learning [3]. In medical image processing, pixel-based machine learning is the evolving field that deals directly with the pixels or voxels of the images. It performs best in preventing the loss of information caused by improper segmentation or feature computations [4]. Machine learning libraries such as Torch is a freely available software library. There are different algorithms of machine learning such as support vector machine, Parzen windows, Adaboost K nearest Neighbours, Hidden Markov models, multi-layer perceptron, Bagging, Bayes classifiers, etc [5]. The Linear classifiers include Logical regression, Quadratic classifiers, Naive Bayes classifier, Perceptron, Quadratic classifiers, support vector machine, Boosting, Decision tree which aggregate random forest, Bayesian and Neural Networks that deal with classification [6]. To diagnose a human body mathematical algorithms are used in Artificial intelligence along with data points [7]. It is very much useful to develop

CVMLH-2022: Workshop on Computer Vision and Machine Learning for Healthcare, April 22 – 24, 2022, Chennai, India.

EMAIL: hepzia@yahoo.com (A. Hepzibah Christinal)

ORCID: 0000-0003-3965-3183 (A. Hepzibah Christinal)



© 2022 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

the prediction accuracy in cancer and related death [8], also involves in predicting cardiac risk [9], and also helps in the diagnostic accuracy of magnetic resonance imaging [10], computerized tomography scan in radiological investigations. To decrease the inconveniences in classification outcomes, CTG interpretation is automated by professionals in medical and engineering [11]. The use of electronic fetal heart monitoring to check fetal status during labor is common. Despite the lack of evidence for its usefulness, this method is nonetheless widely utilized in every current labor and delivery hospital in industrialized countries. To maximize the safety and outcomes of patients, all the contributors of health care to the woman in labour and her new born must have a comprehensive awareness of the underlying pathogenesis of monitoring the heart of fetus as well as an recognition for the labor course and issues as they develop. [12]. In gynaecology, fetal abnormalities are the most likely cause of pregnancy complications. If the fetus's environment inside the womb is unsuitable, the fetus's health is likely to worsen. The fetal heart rate and uterine contractions are recorded at the corresponding time using the cardiotocography technique. Decision Tree, Support Vector Machine, and R - Studio approach for Naive Bayes have been utilized in the research. The datasets are extracted from UCI Machine Learning Repository and categorized into fetal stages as a normal, suspect, and pathological class that is trained, and by using algorithms it is tested, and compared by the use of different performance measurements [13]. To prevent intrapartum hypoxic-ischaemic injury, examining the heart rate of fetus with a cardiotocograph is used to identify variations in the heart rate of fetus during labor [14]. The classification is required to predict the health of newborns, especially in urgent circumstances. Cardiotocography is a technique for assisting obstetricians in obtaining precise details during gestation as a method of monitoring fetal health, especially in women who are pregnant and under great risk . CTG is a continuous electronic record of the baby's heart rate taken from the mother's belly, according to obstetricians. The information obtained is important to visualize the embryo's healthiness and allows for early intervention before the embryo suffers a permanent impairment. The intention of machine learning methods is to make use of the qualities of data collected from the data to solve problems. In this study, they compared the classification capabilities of eight various methods of machine-learning using antepartum cardiotocography data [15]. The most important technique to detect fetal distress is to check the fetal heart rate as this distress leads to complications. The main diagnosing tool to measure FHR is cardiotocography. The wrong results of CTG's graph could result in a significant loss. Decision Tree, K-Nearest Neighbours, Logistic Regression, Support Vector Machine, Random Forest, and Naive Bayes are the six algorithms presented for classification in that study for the categorization of CTG data. A feature selection methodology that is based on classification is used to remove the unnecessary features from the dataset to improve the performance of the classifiers. The evaluating metrics are used to measure the precision, accuracy, and recall of classification algorithms [16,17]. To bring out the difference between normal and abnormal fetal heart rate signals, the Bagging ensemble machine learning technique was used. The F-measure, ROC area and accuracy are used as evaluating indicators to evaluate the classifiers' success. The Bagging ensemble classifier generated favorable results in experiments, and Bagging plus Random Forest produced favorable results having an accuracy of 99.02 percent [18]. An open-access software with MATLAB is introduced to detect the fetal heart rate signals. It is freely available software for research and the software details are given. In addition to the non-linear, linear, morphological, and time-frequency characteristics, the software uses a new approach called image-based time-frequency features for analyzing the fetal heart rate signals. In addition, CTG-OAS was used in an experimental investigation using the CTU-UHB database which is publicly available to test the dependability of the software. The accuracy was 77.81 percent, the sensitivity was 76.83 percent, the specificity was 78.27 percent, and the geometric mean was 77.29 percent in the experimental investigation. [19]. The least-squares support vector machine with a binary decision tree is used to evaluate the fetal state for cardiotocography classification. Particle swarm optimization is used to enhance the LS-SVM parameters. The method's robustness is tested using a 10-fold cross-validation procedure. The method's performance is assessed in terms of accuracy. To examine and display the method's performance, cobweb representation along with receiver characteristic analysis is presented. This method achieves an incredible accuracy rate of 91.62 percent in classification, according to experimental results [20]. This work used genetic algorithms and support vector machines (SVM) to provide a new technique for evaluating fetal well-being from cardiotocograph (CTG) data (GA). Obstetricians commonly employ CTG recordings to determine fetal well-being because they contain rate of heart and uterine contraction of fetus. An SVM-based classifier was constructed using features

collected from normal and abnormal Uterine contraction and signals from Fetal heart rate. After that, the GA is exploited to identify the right characteristic subset for the classifier to classify based on normal and pathology in this data. The production of the novel system was estimated using comprehensive CTG data classified by three professional obstetricians. [11]

2. Methods

In this paper, we used supervised classification on the basis of ML method. The main task of supervised learning is the classification where various techniques are used to create a function that matches the input to the appropriate output. Here the learner learns a function that matches the vector to different classes with the help of the other examples of the input-output function [28]. There are various classifiers which include the Multinomial Logistic Regression, Support Vector Machine, Multilayer perceptron, Random forests and so on which are involved in the classification process [27]. An ensemble technique that is useful to increase the robustness is Bootstrap aggregating. Random forest is found to be a favorable method for decision trees and bagging. Here we used the random forest classifier for classification [31]. We used R studio for implementation. The package of the random forest has some additional details such as the importance of variables and the measure of proximity. The classification is done by the random forest if the response is a factor. If it is not a factor it performs regression [29]. We used the random forest classifier to categorize the fetal stage as normal, suspect, and pathology. The dataset used is the cardiocography dataset taken from the generally accessible ML repository. The features based on the assessment of the heart rate of fetus and uterine contraction which are organized by obstetricians are available in this dataset. It consists of 2126 fetal cardiocograms (CTGs) which are processed automatically and measured. We select the attributes for our method. The attributes used in this paper are described in table 2. The classification is mainly based on the 3 class experiments which give the fetal state as normal, suspect and pathology, and also based on 10 class experiments that involve the morphological patterns [21]. The selected subset of the training dataset helps to build a group of decision trees by the random forest classifier. The votes from distinct trees are gathered to make the final decision. The individual trees are grown in the following way:

- i. Consider N samples to be used as the training set for the development of the tree.
- ii. Consider Q input variables and $q \ll Q$ denotes that for individual nodes, q variables are selected randomly out of Q , and this q helps in splitting the node. When the forest is growing, the value of m is considered to be a constant.
- iii. Without pruning the trees grow the maximum [30]

2.1. Random Forest Classification

Here, we first divide the entire dataset to test and train. Here 80% of dataset made use for training and 20% for testing is used. The data type is changed from numeric into factors for accurate classification. The random forest classifier is utilized in the regression and classification. It is a collaboration of tree predictors where the individual tree relies on the value of random vector with the equal distribution of all the trees in forest. The generalization error of this classifier mainly anticipated on the firmness of the discrete trees and the association among them [22]. It is a classifier containing huge decision trees. The single decision tree may lead to overfitting while the large number of trees leads to number of predictions [23]. The classification of three class is done by the use of random forest classifier. Let Q be the data set comprising of M points of data and s features having V classes, K_i , $i=1, 2, \dots, V$. A subset of independent dataset Φ_t chosen randomly from the dataset Q , so that $\Phi_t \subseteq Q$, having d features set $d \subseteq s$. A tree is trained $h(y, \Phi_t)$ as classifier which is weak for the training set where y is the input. In the random forest classifier, combining various trees helps to predict the class of a specific feature vector by dominant voting. In this paper, the random forest helps in the classification of normal, suspect and pathology images [24]. In ensemble learning, the increased number of classification trees in random classifier helps to increase accuracy and attains a greater generalization. It uses the base classifiers where multiple prediction models are used by combining a group of classifiers which

are independent. Multiple trees are built with the subspace of features randomly. Majority voting is used to select the forest classification which is often combined by the base classifier. The error rate of majority voting is given by $\epsilon_{mv} = \sum_{i=\lfloor P/2 \rfloor + 1}^P \binom{P}{i} \epsilon^i (1 - \epsilon)^{P-i}$ Where P represents the total base classifier and ϵ represents the identical error rate for all base classifiers [25]. The majority of votes by these trees gives the output. To make the accuracy best pruning can be performed. As the number of features enlarge thenumber of trees also enlarge. Hence, this algorithm is great for dealing with higher dimensional data [26]. Here we find the prediction accuracy for classification in account of the testing and training datasetusing the random forest classifier. The figure 1 illustrates the outline of the random forest classifier.

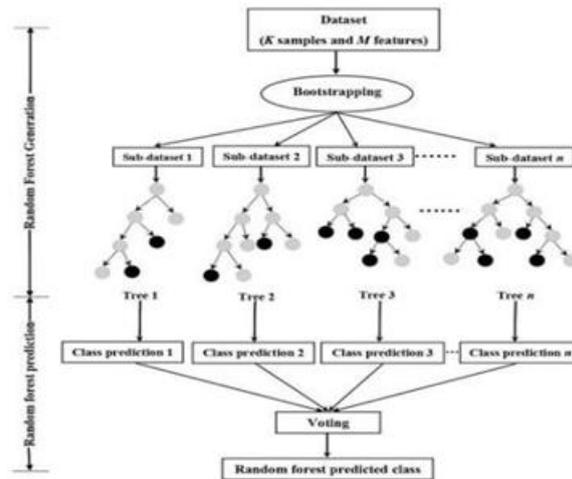


Figure 1: Representation of the Random Forest classifier

3. Experimental Results and Discussion

In this paper, we used the fetal cardiotocography dataset for experimentation. The software implementation of this method is done using R using an intel core i3 64- bit processor on the Windows 10 operating system. The capability of the model is predicted through the confusion matrix. The constituent of the confusion matrix are used to find the important metrics like sensitivity, accuracy, and specificity. Table 1 gives the performance of the training dataset and testing dataset based on sensitivity, positive and negative predicted value, and specificity of class I (normal), class II (suspect), and class III (pathology) datasets.

Table 1: Performance Measures of Dataset Using RFC

Metrics	TRAINING DATA			TESTING DATA		
	Class I	Class II	Class III	Class I	Class II	Class III
Sensitivity	100	99.59	100	99.09	67.30	82.50
Specificity	99.74	100	100	79.35	99.18	98.68
Positive predicted value	99.92	100	100	94.48	92.10	86.84
Negative predicted value	100	99.93	100	96.05	95.55	98.16
Accuracy	99.94%			93.57%		
Kappa	99.84%			81.12%		
Error	0.0006			0.0643		

The RFC achieves the following results in terms of the training dataset. In the case of sensitivity, it achieves 100 % in class 1 and class 3 and 99.595 in class 2. The specificity rate is 100% in class 2 and class 3. And 99.74% in class 1. The general accuracy based on training is 99.94% and the value of kappa is 99.84%. The average accuracy of the testing is 93.57% and the kappa value is 81.12%. It is found that the accuracy is greater in training compared to testing. But overall, it has a very good accuracy rate. In the testing dataset, the random forest classifier achieves sensitivity of 99.09%, 67.30% 82.50% and specificity of 79.35% 99.18%, 98.68% respectively in class 1(normal), class2(suspect) and class 3(pathology) datasets. It is found that the sensitivity rate is much higher in normal and pathology and in suspect, it is similar to 100% in the training dataset, and in the testing, it is found that in the case of suspect it is only 67% while in normal it is 99% and in pathology 82%. Hence, it is found that the sensitivity rate has some more difference in the classification with respect to suspect. In terms of specificity, it is found that it achieves 100 % in suspect and pathology and in normal it achieves 99.74%. in training and in testing it achieves only 79.35% in normal and achieves 99% and 98% in suspect and pathology. The specificity rate is higher in suspect when compared to other metrics in the testing dataset. In testing pathology, it achieves the best results in specificity than others. The overall terms of sensitivity are higher in terms of normal images. The classification error is computed as 1- (accuracy) in both training and testing datasets. The figure 2 illustrates the classification error rate of the RFC. The RFC provides a promising accuracy rate in comparison with other classifiers.

3.1. Evaluating Metrics

To evaluate we used the following metrics in this paper

- **Accuracy**

The correlation of prediction with the actual classification is the accuracy.

$$\text{Acc} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$$

where Acc represents accuracy, TP is the true positives, TN is the true negative, FN is the false negative and FP is false positive

- **Sensitivity**

The ratio of the actual prediction to ground truth is called as the sensitivity. It is assessed as

$$\text{Sens} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

Where Sens represents sensitivity, FN is the false negative and TP is the True positive.

- **Specificity**

It is the opposite to recall

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}}$$

where FP is the False positive and TN is the true negative.

- **Kappa**

The kappa coefficient helps to measure the rate between classification and the ground truth.

$$\text{Kappa} = \frac{\text{TA} - \text{RA}}{1 - \text{RA}}$$

where RA is the random accuracy and TA is the total accuracy.

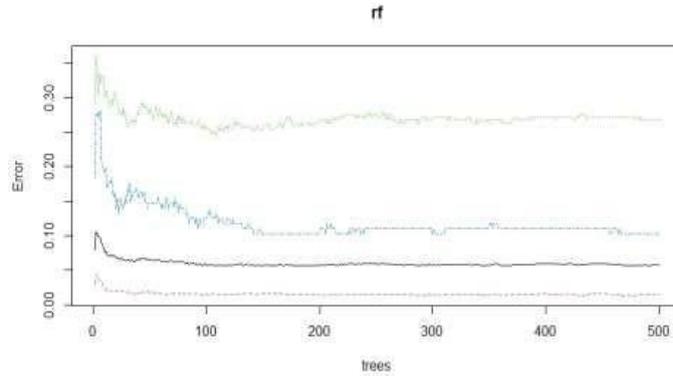


Figure 2: Error rate of random forest

Table 2: Attributes for the process of classification

S.No	Features	Description
1.	B	Start_instant
2.	E	End_instant
3.	LB	Value of baseline
4.	AC	No. of Accelerations
5.	FM	Gesture of fetus
6.	UC	Uterine contractions
7.	DL	Light decelerations
8.	DS	Severe decelerations
9.	DP	Prolonged decelerations
10.	ASTV	Abnormal short-term variability with respect to percentage of time
11.	MSTV	short-term variability with respect to mean
12.	ALTV	Abnormal long-term variability with respect to percentage of time
13.	MLTV	Long-term variability with value of mean
14.	Width	Width of Histogram
15.	Min	Lower frequency
16.	Max	Higher frequency
17.	Nmax	Count of histogram peaks
18.	Nzeros	count of histogram Zeros
19.	Mode	Mode of Histogram
20.	Mean	Mean of Histogram
21.	Median	Median of Histogram
22.	Variance	Variance of Histogram
23.	Tendency	Tendency of Histogram

4. Conclusion

Cardiotocography is a technique for assisting obstetricians in obtaining precise details during gestation as a method of monitoring health of fetus. Here we use the fetal cardiotocography dataset to classify the fetal stage as normal, pathology and suspect. This paper is about the classification of the fetal cardiotocography dataset implemented in R studio. We used 23 attributes for the process of classification. In this classification is done using the random forest classifier. The classification rate is predicted with training and test data and are assessed using the various performance metrics. It is found that whole accuracy rate of training is 99.94% and of testing is 93.57% which performs better and has a higher accuracy compared to other conventional algorithms. In the testing dataset, the random forest classifier achieves sensitivity of 99.09% in normal, 67.30% in suspect, 82.50% in pathology datasets

and specificity of 79.35% in normal 99.18%, in suspect 98.68% in pathology datasets. It is found that it achieves very high sensitivity in terms of normal and best specificity rate in terms of suspect and pathology images. This method is found to be best with astounding results in the classification of the normal, suspect and pathology in the classification of the cardiocography dataset. Hence, this method performs better in classification purpose and further it can be developed by improving the classifier accuracy by feature selection or other methods.

5. References

- [1] Erickson, B. J., Korfiatis, P., Akkus, Z., & Kline, T. L. (2017). Machine learning for medical imaging. *Radiographics*, 37(2), 505-515.
- [2] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
- [3] Christakou, C., Lefakis, L., Vrettos, S., & Stafylopatis, A. (2005, November). A movie recommender system based on semi-supervised clustering. In *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)* (Vol. 2, pp. 897-903). IEEE.
- [4] Suzuki, K. (2012). Pixel-based machine learning in medical imaging. *International Journal of Biomedical Imaging*, 2012.
- [5] Collobert, R., Bengio, S., & Mariéthoz, J. (2002). Torch: a modular machine learning software library (No. REP_WORK). Idiap.
- [6] Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*, 3, 19-48.
- [7] Yu, K. H., Beam, A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature biomedical engineering*, 2(10), 719-731.
- [8] Cruz, J. A., & Wishart, D. S. (2006). Applications of machine learning in cancer prediction and prognosis. *Cancer informatics*, 2, 117693510600200030.
- [9] Weng, S. F., Reys, J., Kai, J., Garibaldi, J. M., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. *PloS one*, 12(4), e0174944.
- [10] Wang, S., & Summers, R. M. (2012). Machine learning and radiology. *Medical image analysis*, 16(5), 933-951.
- [11] Ocak, H. (2013). A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *Journal of medical systems*, 37(2), 1-9.
- [12] Freeman, R. K., Garite, T. J., Nageotte, M. P., & Miller, L. A. (2012). Fetal heart rate monitoring. *Lippincott Williams & Wilkins*.
- [13] Agrawal, K., & Mohan, H. (2019, January). Cardiotocography analysis for fetal state classification using machine learning algorithms. In *2019 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.
- [14] Pinas, A., & Chandharan, E. (2016). Continuous cardiotocography during labour: Analysis, classification and management. *Best practice & research Clinical obstetrics & gynaecology*, 30, 33-47.
- [15] Sahin, H., & Subasi, A. (2015). Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*, 33, 231-238.
- [16] Afridi, R., Iqbal, Z., Khan, M., Ahmad, A., & Naseem, R. (2019). Fetal heart rate classification and comparative analysis using cardiotocography data and KNOWN classifiers. *International Journal of Grid and Distributed Computing (IJGDC)*, 12, 31-42.
- [17] Huang, M. L., & Hsu, Y. Y. (2012). Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network.
- [18] Subasi, A., Kadasa, B., & Kremic, E. (2020). Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier. *Procedia Computer Science*, 168, 34-39.
- [19] Cömert, Z., & Kocamaz, A. F. (2018). Open-access software for analysis of fetal heart rate signals. *Biomedical Signal Processing and Control*, 45, 98-108.

- [20] Yılmaz, E., & Kılıkçier, Ç. (2013). Determination of fetal state from cardiocogram using LS-SVM with particle swarm optimization and binary decision tree. *Computational and mathematical methods in medicine*, 2013.
- [21] Bache, K., & Lichman, M. (2010). *Cardiotocography data set*. UCI Machine Learning Repository.
- [22] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- [23] Devetyarov, D., & Nouretdinov, I. (2010, October). Prediction with confidence based on a random forest classifier. In *IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 37-44). Springer, Berlin, Heidelberg. Arif, M. (2015).
- [24] Classification of cardiocograms using random forest classifier and selection of important features from cardiocogram signal. *Biomaterials and Biomechanics in Bioengineering*, 2(3), 173-183.
- [25] Comert, Z., & Kocamaz, A. F. (2017). Comparison of machine learning techniques for fetal heart rate classification.
- [26] Nagendra, V., Gude, H., Sampath, D., Corns, S., & Long, S. (2017, August). Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiocography data. In *2017 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)* (pp. 1-6). IEEE.
- [27] Dahiya, S., Tyagi, R., & Gaba, N. (2020). Comparison of ML classifiers for Image Data (No. 3815). EasyChair.
- [28] Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*. b, 4, 51- 62.
- [29] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18- 22
- [30] Imran Molla, M. M., Jui, J. J., Bari, B. S., Rashid, M., & Hasan, M. J. (2021). *Cardiotocogram Data Classification Using Random Forest Based Machine Learning Algorithm*. In *Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019* (pp. 357-369). Springer, Singapore.
- [31] Lee, T. H., Ullah, A., & Wang, R. (2020). Bootstrap aggregating and random forest. In *Macroeconomic Forecasting in the Era of Big Data* (pp. 389-429). Springer, Cham.