

## Use of Machine Learning Algorithms for Prediction of Fetal Risk using Cardiotocographic Data

### Abstract

**Background:** A major contributor to under-five mortality is the death of children in the 1<sup>st</sup> month of life. Intrapartum complications are one of the major causes of perinatal mortality. Fetal cardiotocograph (CTGs) can be used as a monitoring tool to identify high-risk women during labor. **Aim:** The objective of this study was to study the precision of machine learning algorithm techniques on CTG data in identifying high-risk fetuses. **Methods:** CTG data of 2126 pregnant women were obtained from the University of California Irvine Machine Learning Repository. Ten different machine learning classification models were trained using CTG data. Sensitivity, precision, and F1 score for each class and overall accuracy of each model were obtained to predict normal, suspect, and pathological fetal states. Model with best performance on specified metrics was then identified. **Results:** Determined by obstetricians' interpretation of CTGs as gold standard, 70% of them were normal, 20% were suspect, and 10% had a pathological fetal state. On training data, the classification models generated by XGBoost, decision tree, and random forest had high precision (>96%) to predict the suspect and pathological state of the fetus based on the CTG tracings. However, on testing data, XGBoost model had the highest precision to predict a pathological fetal state (>92%). **Conclusion:** The classification model developed using XGBoost technique had the highest prediction accuracy for an adverse fetal outcome. Lay health-care workers in low- and middle-income countries can use this model to triage pregnant women in remote areas for early referral and further management.

**Keywords:** Fetal cardiotocography, machine learning, perinatal risk

### Introduction

The Millennium Development Goal 4, which aimed to reduce under-five mortality by two-thirds globally, was not able to meet its target.<sup>[1]</sup> In 2015, 45% of all under-five deaths occurred in the neonatal period.<sup>[2]</sup> The leading causes of deaths among this group are preterm birth complications (35%), intrapartum events (25%), and infections (e.g., sepsis or meningitis in 15%).<sup>[3]</sup> According to the UNICEF 2018 report, Pakistan has one of the highest newborn mortality rates of 46/1000 live births.<sup>[3]</sup>

According to the International Federation of Gynaecology and Obstetrics (FIGO) guidelines, cardiotocograph (CTG) can be classified as normal, suspect, or pathological based on the fetal heart rate (FHR), heart rate variability, accelerations, and decelerations.<sup>[4]</sup> This interpretation can be done by skilled health personnel

(e.g., obstetricians) or computerized software.<sup>[5]</sup> A recent Cochrane review by Grivell *et al.* reported a significant reduction in perinatal mortality with computerized CTG (relative risk: 0.20, 95% confidence interval [CI]: 0.04–0.88) as compared to traditional CTG.<sup>[5]</sup> However, since the studies were of moderate quality evidence, further work to assess the impact of CTG on perinatal outcomes needs to be conducted.<sup>[5]</sup>

Artificial intelligence (AI) uses mathematical algorithms and several data points from the human body to generate a diagnosis.<sup>[6]</sup> These models have been used to improve the accuracy of predicting cancer recurrence and mortality,<sup>[7]</sup> cardiovascular risk prediction,<sup>[8]</sup> and the diagnostic accuracy of radiological investigations such as computerized tomography scan and magnetic resonance imaging.<sup>[9]</sup> Medical and engineering professionals have been working to automate CTG interpretation, hence decreasing inconsistencies in classification of outcomes.<sup>[10]</sup> The existing

Zahra Hoodbhoy,  
Mohammad  
Noman<sup>1</sup>,  
Ayesha Shafique<sup>1</sup>,  
Ali Nasim<sup>1</sup>,  
Devyani  
Chowdhury<sup>2</sup>,  
Babar Hasan

Department of Paediatrics and Child Health, The Aga Khan University, <sup>1</sup>Department of Artificial Intelligence, Ephlux Pvt Ltd., Karachi, Pakistan, <sup>2</sup>Cardiology Care for Children, Pennsylvania, USA

**Received:** 11 November, 2018.

**Revision:** 29 March, 2019.

**Accepted:** 01 August, 2019.

**Published:** 11 October, 2019.

**Address for correspondence:**

Dr. Babar Hasan,

Department of Paediatrics and Child Health, The Aga Khan University, Karachi, Pakistan.

E-mail: babar.hasan@aku.edu

Access this article online

**Website:**

www.ijabmr.org

**DOI:**

10.4103/ijabmr.IJABMR\_370\_18

**Quick Response Code:**



**How to cite this article:** Hoodbhoy Z, Noman M, Shafique A, Nasim A, Chowdhury D, Hasan B. Use of machine learning algorithms for prediction of fetal risk using cardiotocographic data. *Int J App Basic Med Res* 2019;9:226-30.

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

algorithms have high accuracy to predict the pathological state of the fetus but did not perform well on predicting the suspicious state.<sup>[11,12]</sup>

The objective of this study was to develop a machine learning model that can identify high-risk fetuses (suspicious as well as pathological state) as accurately as highly trained medical professionals.

## Methods

The dataset was obtained from the University of California Irvine Machine Learning Repository.<sup>[13]</sup> It comprised of 2126 pregnant women who were in the third trimester of pregnancy. The dataset consisted of 21 attributes used in the measurements of FHR and uterine contractions (UCs) on CTG [Table 1]. According to the standards and concord of the National Institute of Child Health and Human Development, the core risk variable used to derive the state of fetus includes qualitative and quantitative descriptions of FHR (i.e., baseline heart rate; baseline variability; number of accelerations per second; number of early, late, and variable decelerations per second; number of prolonged decelerations per second; and sinusoidal pattern) and UCs (i.e., baseline uterine tone, contraction frequency, duration, and strength).<sup>[14]</sup> The CTG of pregnant women were classified by three experts who were specialized in obstetrics, with their interpretation being considered to be the gold standard. The fetal CTGs were generated

from SisPorto 2.0 software (Speculum, Lisbon, Portugal), a program for automated analysis of CTG. The machine learning algorithms used in this study were multilayer perceptron, support vector machine with linear and radial basis function kernel, K-nearest neighbors, XGBoost classifier, AdaBoost classifier, random forest, logistic regression, Gaussian Naïve Bayes, and decision tree. The current dataset was split into training and testing folds using K-Fold Cross Validation technique to test the performance of each machine learning model in the training phase.<sup>[15]</sup> Classifiers were compared on the bases of the highest average accuracy across all the folds and sensitivity value of each class to obtain the classifier with best generalization on given dataset. As the dataset was imbalanced, Synthetic Minority Over-sampling Technique (SMOTE) balancing technique was used.<sup>[16]</sup> SMOTE is used to avoid overfitting of the machine learning model on skewed classes. This technique was only applied on training folds and was then tested on real, intact, and unseen data.

The key outcome of this study was to compare major machine learning algorithms (listed above) with regard to their precision accuracy and sensitivity to predict normal, suspect, or pathologic fetal state based on CTG attributes.

Various statistical techniques were used to compare the performance of the algorithms. These included precision, sensitivity or recall, F1 score, and overall accuracy ( $[\text{true positive} + \text{true negative}] / [\text{true positive} + \text{true negative} + \text{false positive} + \text{false negative}]$ ).

**Table 1: Essential cardiogram attributes used in the models**

Variable symbol	Variable description
LB	Fetal heart rate baseline (beats per minute)
AC	Number of accelerations per second
FM	Number of fetal movements per second
UC	Number of uterine contractions per second
DL	Number of light decelerations per second
DS	Number of severe decelerations per second
DP	Number of prolonged decelerations per second
ASTV	Percentage of time with abnormal short-term variability
MSTV	Mean value of short-term variability
ALTV	Percentage of time with abnormal long-term variability
MLTV	Mean value of long-term variability
Width	Width of FHR histogram
Min	Minimum of FHR histogram
Max	Maximum of FHR histogram
Nmax	Number of histogram peaks
Nzeros	Number of histogram zeroes
Mode	Histogram mode
Median	Histogram median
Variance	Histogram variance
Tendency	Histogram tendency
NSP	Fetal state class code (N=Normal, S=Suspected, P=Pathological)

FHR: Fetal heart rate

## Results

The CTG data of 2126 pregnant women were classified into the normal, suspect, or pathologic state by three obstetricians. The CTG data comprised of 70% normal fetal state, 20% suspect state, and 10% pathologic state as determined by the obstetrician.

Similar to other models, the most prominent risk factors depicted by our all ten machine learning models were percentage of time with abnormal short-term variability, percentage of time with abnormal long-term variability, number of accelerations per second (AC), mean value of short-term variability, and UCs. These five factors were seen to have the highest weight in predicting the fetal state. The performance metrics of all the ten machine learning models on training and testing data are shown in Tables 2 and 3, respectively.

On the training dataset, it was seen that the machine learning model generated by XGBoost technique, decision tree, and random forest had high precision and sensitivity (>96% and >99%, respectively) to predict the suspect and pathological state of the fetus based on the CTG tracings. However, when this algorithm was applied on the testing dataset, the model developed using XGBoost

**Table 2: Comparison of machine learning models on training data**

ML model	Precision			Recall			F1 score		
	N	S	P	N	S	P	N	S	P
MLP	0.88	0.87	0.94	0.87	0.84	0.96	0.87	0.85	0.95
XGBoost classifier	0.99	0.96	0.996	0.97	0.987	0.992	0.976	0.975	0.994
Decision tree	0.998	1	1	1	0.998	1	0.999	0.999	1
Random forest	0.992	0.989	0.997	0.989	0.992	0.996	0.99	0.991	0.997
Logistic regression	0.87	0.77	0.88	0.84	0.79	0.88	0.86	0.79	0.88
SVM linear kernel	0.9	0.8	0.89	0.85	0.83	0.91	0.87	0.81	0.9
SVM RBF kernel	0.98	0.92	0.99	0.92	0.97	0.99	0.95	0.94	0.984
KNN	0.995	0.95	0.99	0.95	0.993	0.995	0.97	0.97	0.993
Naïve Bayes	0.88	0.66	0.86	0.76	0.88	0.68	0.82	0.75	0.76
AdaBoost	0.86	0.88	0.988	0.89	0.88	0.95	0.87	0.88	0.97

N: Normal state; S: Suspect state; P: Pathological state; MLP: Multilayer perceptron; SVM: Support vector machine; RBF: Radial basis function; KNN: K-nearest neighbors; ML: Machine learning

**Table 3: Comparison of machine learning models on testing data**

ML model	Precision			Recall			F1-Score		
	N	S	P	N	S	P	N	S	P
MLP	0.96	0.52	0.7	0.85	0.72	0.89	0.9	0.6	0.77
XGBoost classifier	0.98	0.73	0.92	0.94	0.88	0.92	0.96	0.8	0.92
Decision tree	0.96	0.74	0.87	0.95	0.74	0.92	0.95	0.74	0.89
Random forest	0.96	0.73	0.86	0.95	0.78	0.88	0.95	0.75	0.87
Logistic regression	0.96	0.48	0.64	0.84	0.75	0.84	0.9	0.58	0.72
SVM linear kernel	0.97	0.49	0.68	0.84	0.79	0.88	0.9	0.6	0.76
SVM RBF kernel	0.98	0.62	0.84	0.91	0.82	0.88	0.94	0.7	0.86
KNN	0.96	0.6	0.82	0.9	0.76	0.87	0.93	0.66	0.84
Naïve Bayes	0.97	0.42	0.46	0.76	0.85	0.67	0.85	0.56	0.54
AdaBoost	0.96	0.58	0.88	0.89	0.81	0.87	0.92	0.67	0.87

N: Normal state; S: Suspect state; P: Pathological state; MLP: Multilayer perceptron; SVM: Support vector machine; RBF: Radial basis function; KNN: K-nearest neighbors; ML: Machine learning

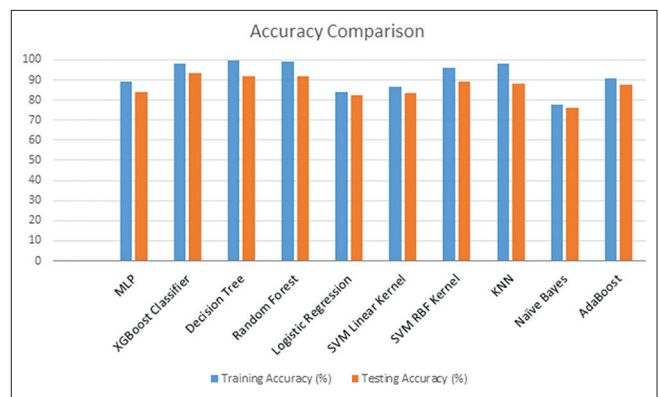
had high sensitivity (92%) to predict a pathological fetal state as compared to the other models, but the sensitivity for the suspect state dropped to 73%.

When the training and testing dataset was compared for overall accuracy, the model developed by XGBoost technique had the highest overall accuracy (93%) as compared to other machine learning models [Figure 1].

### Discussion

In this study, ten different machine learning models were applied on the CTG recordings of 2126 pregnant women to predict an adverse fetal outcome. XGBoost technique was found to have the most accuracy for an adverse fetal outcome, i.e., suspected and pathological states. CTG interpretation relies heavily on obstetrician’s analysis of the tracing and leads to subjectivity in the interpretation of the data.<sup>[17]</sup> Interobserver agreement between trained obstetricians using the FIGO guidelines was fair (kappa statistic – 0.48).<sup>[18]</sup> AI systems may thus be the solution where different parameters can be assessed by the machine with high reliability.<sup>[19]</sup>

Costa *et al.* reported a high intraclass correlation coefficient 71% for accelerations (95% CI: 69%–73%)



**Figure 1: Overall accuracy of the different models on the training and testing data**

and 68% for decelerations (95% CI: 66%–70%) between the automated models and trained clinicians.<sup>[20]</sup> The CTG Open Access Software used 552 raw CTG samples to obtain a prediction accuracy of 87.9%.<sup>[21]</sup> The model generated by Cömert *et al.* on the same dataset reported an accuracy of classification of the artificial neural network model and extreme learning machine as 91.8% and 93.4%, respectively.<sup>[11]</sup> Although these models

had a high accuracy for predicting a pathological fetal state (97%), the accuracy dropped significantly (59%) for prediction of suspect fetal states (59%).<sup>[12]</sup> The model generated by this study using XGBoost technique had a similar overall accuracy (96%) of predicting pathological state, but the prediction accuracy for a suspect state was higher (73%). Both suspect and pathological fetal states may be a sign of fetal hypoxia due to conditions such as excessive uterine activity, aortocaval compression, or maternal hypotension.<sup>[4]</sup> It is hence important to use an algorithm that has a high accuracy for both these fetal states (suspect and pathological), thus determining the optimal time and mode of delivery, avoiding prolonged fetal hypoxia at the same time preventing unnecessary obstetric interventions.<sup>[4]</sup>

Several computerized algorithms have thus been developed with varying accuracy to help in analyzing CTG data; however, none of them have been universally adopted.<sup>[12]</sup> One of the reasons for this could be that automated programs such as SisPorto software may help in clinical decision making but are compatible with only certain brands of machines.<sup>[19]</sup> This restricts their universal utility, especially in low-resource countries where access to technologically advanced machines may be limited. Vendor independence using the machine learning model such as the one developed in this study may be possible but would require open interfacing of vendor-specific raw data with the machine learning algorithm.

CTG may play a valuable role in identifying high-risk fetal states during the early stages of labor and if appropriately managed may prevent birth asphyxia and fetal deaths.<sup>[22]</sup> As CTG requires expert interpretation, it limits its applicability in remote areas where skilled health professionals are scarce. The role of eHealth technology in enhancing health-care utilization and improving the quality of antenatal and postpartum care in low- and middle-income countries (LMICs) has been established.<sup>[23]</sup> Using technological advancements, incorporation of an automated machine learning CTG model with high accuracy (such as the one developed in this study) to predict suspect as well as pathological fetal state would enable task sharing with lay health-care providers to ensure timely referral and management of women in labor, hence improving perinatal outcomes.<sup>[24]</sup>

The strength of this study was that it used ten different machine learning techniques on the CTG dataset and proposed the one with the highest accuracy on both suspect and pathological fetal states. It also used SMOTE balancing technique to avoid the bias of the model toward skewed data, hence improving prediction accuracy of the machine learning algorithm. However, a major limitation of this work is that this dataset was obtained from a repository in the developed world. Due to the differences in sociodemographic characteristics of pregnant women

in LMICs, the machine learning algorithm may report a different accuracy. Further, this dataset did not include any information on participants' sociodemographic data or other relevant clinical characteristics, such as primiparity, maternal nutritional status, and anemia, gestational age, fetal well-being, etc. which may affect the intrapartum course of events and could potentially contribute toward further refinement of the AI model.

## Conclusion

Once validated, future implications for this model can include use by lay health-care workers to triage pregnant women in remote areas who may be at high risk for adverse perinatal outcomes based on CTG findings, for referral and further management.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## References

1. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, *et al.* Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: A systematic analysis by the UN inter-agency group for child mortality estimation. *Lancet* 2015;386:2275-86.
2. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, *et al.* Global, regional, and national causes of under-5 mortality in 2000-15: An updated systematic analysis with implications for the sustainable development goals. *Lancet* 2016;388:3027-35.
3. The United Nations Children's Fund. Every Child Alive: The Urgent Need to End Newborn Deaths. The United Nations Children's Fund; 2018.
4. Ayres-de-Campos D, Spong CY, Chandrachan E. FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *Int J Gynaecol Obstet* 2015;131:13-24.
5. Grivell RM, Alfirevic Z, Gyte GM, Devane D. Antenatal cardiotocography for fetal assessment. *Cochrane Libr* 2015;9:CD007863.
6. Karam A. Artificial Intelligence in Health Care; 2014. Available from: <http://azikar24.com/artificial-intelligence-in-health-care>. [Last accessed on 2018 Nov 04].
7. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform* 2007;2:59-77.
8. Weng SF, Reips J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017;12:e0174944.
9. Wang S, Summers RM. Machine learning and radiology. *Med Image Anal* 2012;16:933-51.
10. Ocak H. A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *J Med Syst* 2013;37:9913.
11. Cömert Z, Kocamaz AF, Güngör S, editors. Cardiotocography signals with artificial neural network and extreme learning machine. *Signal Processing and Communication Application*

- Conference (SIU). IEEE; 2016.
12. Sundar C, Chitradevi M, Geetharamani G. Classification of cardiotocogram data using neural network based machine learning technique. *Int J Comput Appl* 2012;47:14.
  13. Repository UIML. Available from: <https://archive.ics.uci.edu/ml/index.php>. [Last accessed on 2018 Jun 11].
  14. Macones GA, Hankins GD, Spong CY, Hauth J, Moore T. The 2008 national institute of child health and human development workshop report on electronic fetal monitoring: Update on definitions, interpretation, and research guidelines. *J Obstet Gynecol Neonatal Nurs* 2008;37:510-5.
  15. Rohani A, Taki M, Abdollahpour M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renew Energy* 2018;115:411-22.
  16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res* 2002;16:321-57.
  17. 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.
  18. Ayres-de-Campos D, Bernardes J, Costa-Pereira A, Pereira-Leite L. Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. *Br J Obstet Gynaecol* 1999;106:1307-10.
  19. Ayres-de Campos D, Bernardes J, Garrido A, Marques-de-Sá J, Pereira-Leite L. SisPorto 2.0: A program for automated analysis of cardiotocograms. *J Matern Fetal Med* 2000;9:311-8.
  20. Costa MA, Ayres-de-Campos D, Machado AP, Santos CC, Bernardes J. Comparison of a computer system evaluation of intrapartum cardiotocographic events and a consensus of clinicians. *J Perinat Med* 2010;38:191-5.
  21. Cömert Z, Kocamaz AF, editors. A novel software for comprehensive analysis of cardiotocography signals "CTG-OAS". *Artificial Intelligence and Data Processing Symposium (IDAP)*, 2017. International IEEE; 2017.
  22. Chandrharan E, Arulkumaran S. Prevention of birth asphyxia: Responding appropriately to cardiotocograph (CTG) traces. *Best Pract Res Clin Obstet Gynaecol* 2007;21:609-24.
  23. Lee SH, Nurmatov UB, Nwaru BI, Mukherjee M, Grant L, Pagliari C. NEffectiveness of mHealth interventions for maternal, newborn and child health in low – And middle-income countries: Systematic review and meta-analysis. *J Glob Health* 2016;6:010401.
  24. Dawson AJ, Buchan J, Duffield C, Homer CS, Wijewardena K. Task shifting and sharing in maternal and reproductive health in low-income countries: A narrative synthesis of current evidence. *Health Policy Plan* 2014;29:396-408.