

Research Article

Automatic detection of fetal health status from cardiotocography data using machine learning algorithms

Md. Tamjid Rayhan, A S M Shamsul Arefin* and Sabbir Ahmed Chowdhury¹

Department of Biomedical Physics & Technology, Faculty of Science, University of Dhaka, Dhaka, Bangladesh

ARTICLE INFO	ABSTRACT
<p>Article History</p> <p>Received: 22 June 2021 Revised: 26 August 2021 Accepted: 22 November 2021</p> <p>Keywords: Cardiotocography, Machine Learning, Fetal health, SDG, Stillbirth.</p>	<p>A method for the automatic determination of the fetus health status using Cardiotocography (CTG) and computer-based machine learning algorithms was developed. Five computation friendly machine learning algorithms were used to create multiclass classification models to predict the fetus health status from secondary CTG dataset containing normal, suspected and pathologic data available at University California Irvine Machine Learning Repository. Furthermore, a comparative analysis among the built models was executed. According to the comparative analysis, the best model to automatically detect fetal health was the extreme gradient boosting algorithm-based model with an accuracy of 96.7% and an F1-Score of 0.963 in the pathologic class. This finding thus has the potential to diagnose fetal heart conditions unsupervised, and more efficiently and effectively.</p>

Introduction

In 2019, approximately 2.0 million incidents of stillbirth took place worldwide and about 65% of the cases belonged to families of low and lower-middle income (Hug et al., 2020). Stillbirths have an enormous impact on mothers, families, health care professionals, and the community (Heazell et al., 2016). Previous studies have quantified the direct (Mistry et al., 2013) and indirect (Heazell et al., 2016) financial costs for parents after an experience of stillbirth. However, the psychological and social costs associated with stillbirth have been described as unquantifiable (Tarricone, 2006). Based on the health burden associated with stillbirth, sustainable development goal (SDG) 3.2 has focused on ending preventable stillbirths by 2030 (Frøen et al., 2016).

Unfortunately, the stillbirth rate is higher in rural areas than in urban areas (Abir et al., 2017; Fauveau et al., 1990). Ironically the medical facilities and doctors are less in rural areas (Abir et al., 2017). Hence, the knowledge of the health status of the fetus can help the doctors make effective decisions in emergency conditions. For example, if a fetus is

pathologic by a diagnostic device, the doctors can focus their attention on that patient and decide to start a surgical procedure. Against this backdrop, Cardiotocography (CTG) can be an effective tool to assess fetal health. CTG is the visual representation of fetal heart rate (FHR) and uterine contractions. FHR is an important indicator of fetal status. The fetal neurologic system controls the fetal heart through afferent and efferent networks. As the fetal neurologic system controls FHR, it directly indicates fetal well-being (Pettker and Campbell, 2018). CTG signal is used for fetal health status detection by carefully examining certain signal features. The normal baseline frequency for a healthy fetus is 110-160 beats/minute, whereas a baseline less than 110 beats/minute for 10 minutes will be considered Bradycardia. On the other hand, a baseline greater than 160 beats/minute for 10 minutes will be considered Tachycardia. Variability from baseline heart rate 6-25 beats/minute is considered normal. The presence of acceleration is a must in normal fetus, whereas deceleration is non-reassuring.

*Corresponding author: <arefin.bmpt@du.ac.bd>

¹Institute of Education and Research (IER), University of Dhaka, Dhaka, Bangladesh

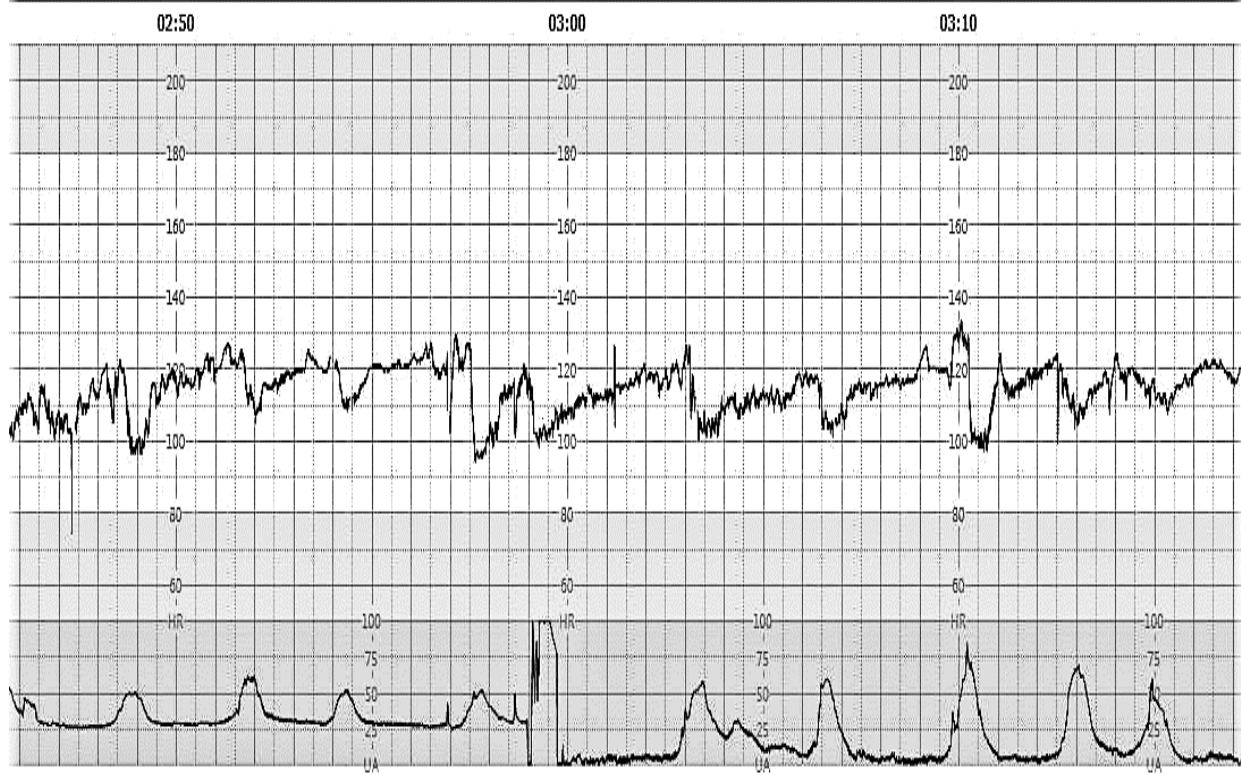


Fig. 1. Normal cardiocotography signal with early decelerations (CTG Case 19, 2012).

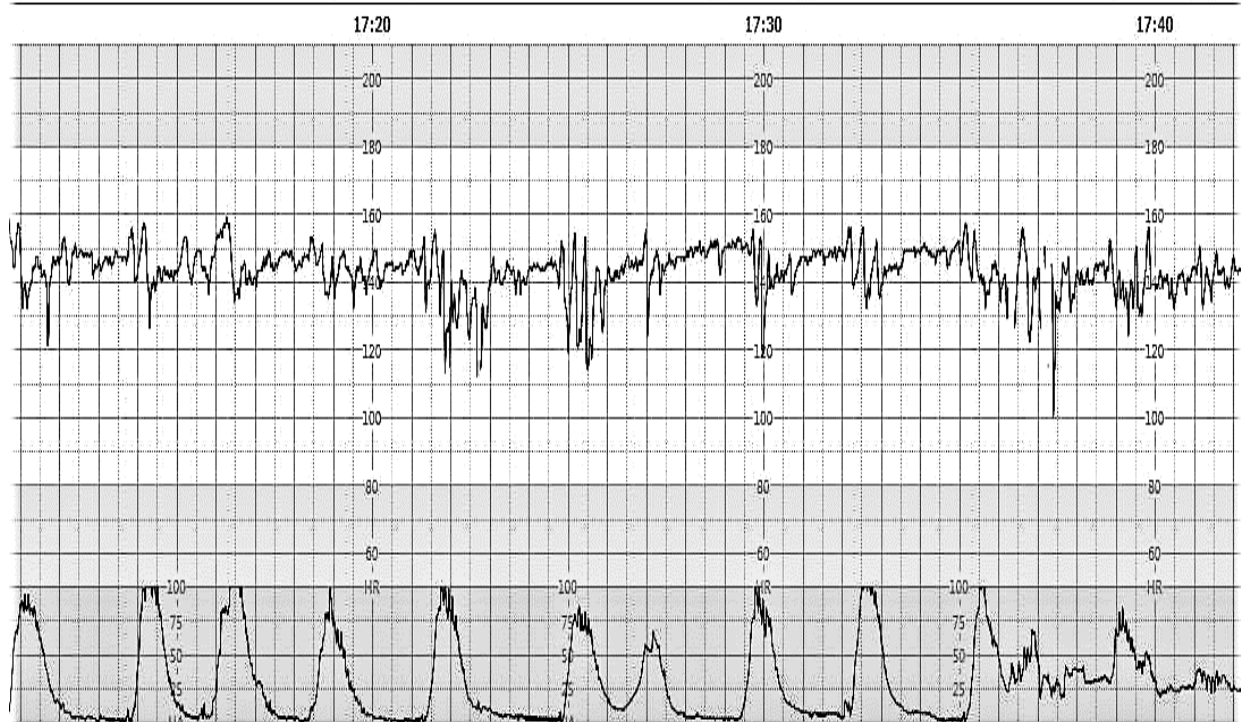


Fig. 2. Pathologic signal with repetitive late decelerations (CTG Case 20, 2012).

While deceleration is non-reassuring, the presence of early decelerations concerning uterus contraction is not a sign of pathology. The signal in Fig. 1 has early decelerations and a normal baseline heart rate. Although this signal does not indicate pathology, the physician decided to perform forceps delivery resulting in an unnecessary intervention (CTG Case 19, 2012). But in the case of Fig. 2, there is a presence of late repetitive decelerations and increasing the baseline heart rate, which is a clear sign of pathology. The physician decided to perform a caesarean section which was a necessary intervention (CTG Case 20, 2012). Thus, CTG signal interpretation is a complex process involving examining several different signal features. Manually performing this task is hard, time-consuming, and error-prone. Misinterpreting the signal can lead to fetal death or unnecessary intervention. Hence, automating the process of CTG signal interpretation will benefit better decision-making.

In this context, to obtain the SDG of ending preventable stillbirths by 2030 for Bangladesh, proper use of technology can be of great help. The study aimed to combine the power of biomedical instruments such as Cardiotocography or fetal heart monitor and the inferential ability of computer-based machine learning algorithms to determine the health status of the fetus automatically. Overall, the potential goals that can be attained using the results of this study are:

1. Automatic fetal health prediction from raw CTG data.
2. Effective and efficient decision making by the doctors in times of pregnancy emergency. For example, situations like labor or complications related to pregnancy and ruling out most healthy cases as healthy and facilitating the doctor to focus on the really risky patients.

Materials and Methods

This study explored the optimal machine learning models that can accurately detect the suspected and pathologic fetus from a normal fetus. Secondary CTG data were collected from the University of California Irvine (UCI) machine learning repository (Dua & Graff, 2017). Collected data were processed for outlier removal, and following that feature engineering algorithm was applied. After that, multiple machines learning algorithm-based model development and evaluation processes were performed. The details are discussed step by step in the following subsections.

Experimental Paradigm

Each row of the dataset represents one instance of CTG recording. Features of these recordings, which are useful in predicting the fetus's health, are listed in each row. Each row of this dataset is labeled with three classes – Normal, Suspected, and Pathologic. The experimental protocol contained two key stages: the data preprocessing and the machine learning model development stages. The data preprocessing stage is shown in the top four boxes in Figure 3. There are three key steps in this data preprocessing stage: i) removing unnecessary data columns such as patient names, ii) removing outlier data to obtain a clean dataset, and iii) selecting important features using feature selection algorithms.

After preprocessing the dataset, the machine learning algorithm-based models were built. The parallel dark 5 boxes in the middle of Fig. 3 describe the models considered for this classification problem. Logistic Regression, Random Forest Classifier, Support Vector Machine, K-Nearest Neighbor, and Extreme Gradient Boosting were the five different machine learning algorithms applied in this study.

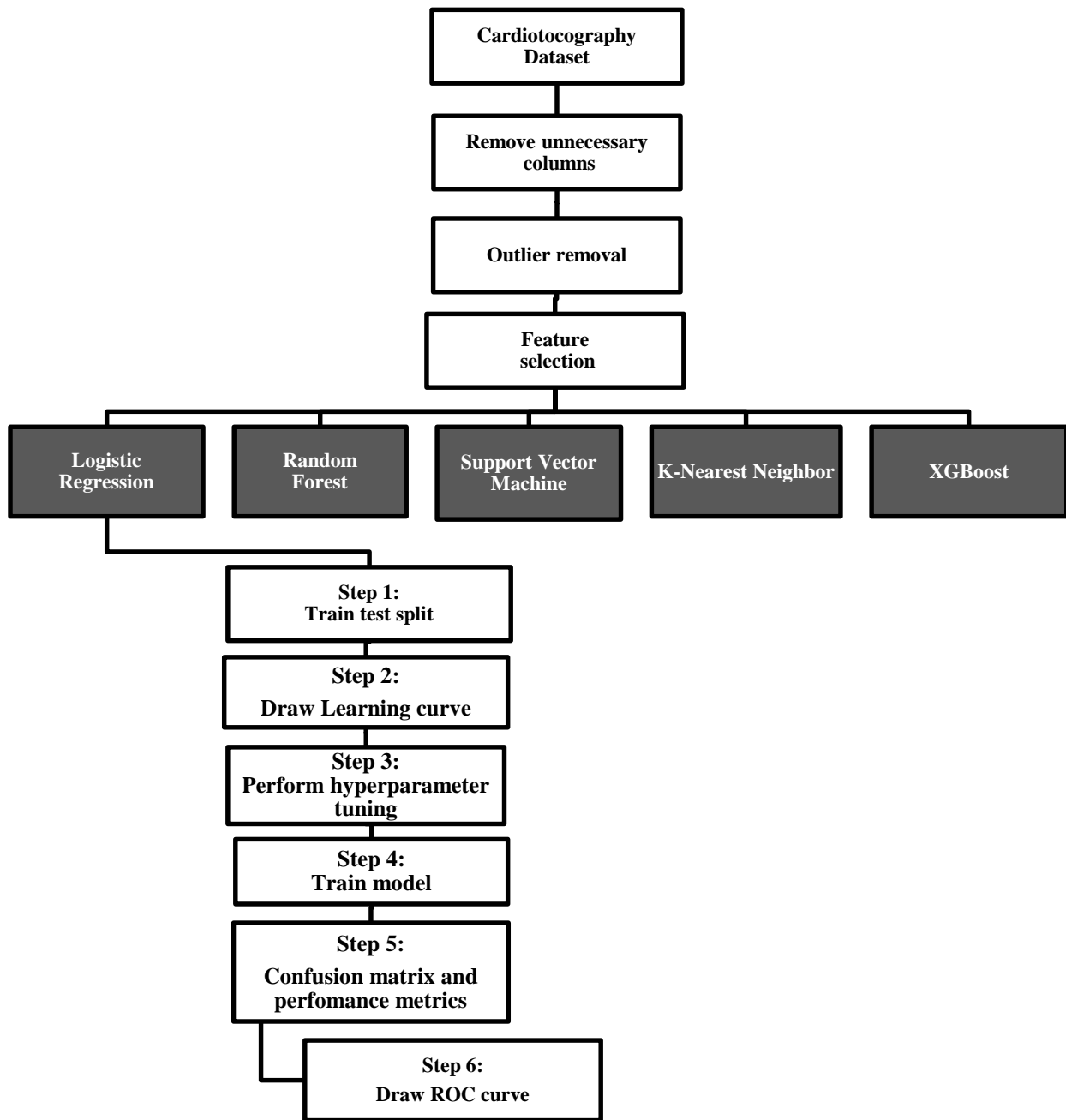


Fig. 3. Flowchart for building predictive models for the task of predicting fetus heart's state from Cardiocography dataset (Top four boxes – dataset preprocessing steps, middle five parallel dark boxes – different machine learning algorithms applicable to this dataset, bottom six boxes- steps to develop machine learning models).

They were chosen to avoid unnecessary computational complexity on a small, symmetrically distributed tabular dataset and consider the implementation capacity and interfacing suitability in the context of Bangladesh. The bottom six boxes represent the model building and model validation stage. These steps are similar for five different algorithms, and they are:

1. Train test split: Dividing the dataset into two sets—Training and Test sets.
2. Learning curve: Plotting learning curve to determine if enough data is available for training.
3. Hyperparameter tuning: Plotting validation curve to determine the optimal value of hyperparameter for a chosen algorithm.
4. Model Training: Training the model using the training dataset and derived hyperparameters for a chosen algorithm.
5. Confusion matrix and performance metrics: This is the first step in model validation to plot the confusion matrix and determine performance metrics such as accuracy, sensitivity, precision, F1 - score. These performance metrics show how accurately one model classifies each class.
6. Receiver operating characteristic (ROC) curve and area under the curve (AUC) value: This is the second step in model validation. Plotting the ROC curve gives us a visualization of how well a model performs on a given class. AUC value is a numeric measure of models' performances in a particular class, and the closer it is to 1, the better.

CTG Dataset

The CTG dataset was obtained from the UCI machine learning repository (Dua & Graff, 2017). The CTG dataset used in this study contains the data of 2126 fetal cardiocograms. This CTG data were automatically processed, and the respective diagnostic features were measured. Three expert obstetricians also classified the CTGS, and a consensus classification label was

assigned to each of them (Ayres-de-campos et al., 2000). The classification was concerning a morphologic pattern (A, B, C, ...) and a fetal state: Normal, Suspected, and Pathologic in Table 1 (Ayres-de-campos et al., 2000). This is a highly skewed dataset where most data points are from the normal class.

Table 1. Number of data points in different classes of the CTG dataset.

Fetal state	Number of data points (support)
Normal	1655
Suspected	295
Pathologic	176

A list of the features extracted in the CTG dataset (Ayres-de-campos et al., 2000) are as follows:

1. LB-Fetal Heart Rate baseline (beats per minute).
2. AC - number of accelerations per second.
3. FM - number of fetal movements per second.
4. UC - number of uterine contractions per second.
5. DL - number of light decelerations per second.
6. DS - number of severe decelerations per second.
7. DP - number of prolonged decelerations per second.
8. ASTV - the percentage of time with abnormal short-term variability.
9. MSTV - mean value of short-term variability.
10. ALTV - the percentage of time with abnormal long-term variability.
11. MLTV - mean value of long-term variability.
12. Width - width of FHR histogram.
13. Min - minimum of FHR histogram.
14. Max - maximum of FHR histogram.
15. Nmax - number of histogram peaks.
16. Nzeros - number of histogram zeros.
17. Mode - histogram mode.

18. Mean - histogram mean.
19. Median - histogram median.
20. Variance - histogram variance.
21. Tendency - histogram tendency.
22. CLASS - FHR pattern class code (1 to 10).
23. NSP - fetal state class code (N=normal; S=suspect; P=pathologic).

Outlier Removal

Outlier removal was done in our work on the three classes separately instead of removing them from the whole dataset at once using the Isolation Forest Algorithm (Liu et al., 2008). This was done to prevent detecting the pathologic data points as outliers. A number of estimators used by this Isolation Forest algorithm was set as 20.

Feature Engineering

This study used Minimum Redundancy Maximum Relevance (MrMr) model to select the important features. This algorithm selects good features according to the maximal statistical dependency criterion based on mutual information (Peng et al., 2005). MrMr was applied to the CTG dataset to find the 10 most important features to predict the fetus’s health status. The selected features were - ALTV, MSTV, ASTV, MLTV, Tendency left asymmetric, Tendency right asymmetric, Mean, Variance, Mode, and Min.

Machine Learning Algorithms

Logistic Regression (LR)

To verify the applicability of LR algorithm (Walker & Duncan, 1967), a learning curve was drawn, which is shown in Figure 4. As we can see, as the number of training sets increases, the validation score improves, which indicates that LR is a suitable algorithm for this problem. The optimum values for the hyperparameters were derived by drawing validation curves. The optimum number of iterations for LR was found to be 50 and the optimum value for the regularization constant was 0.8.

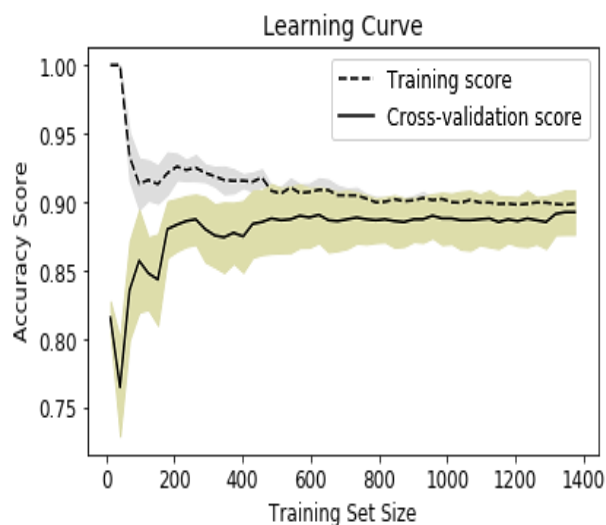


Fig. 4. The learning curve for logistic regression

The main bottleneck in training the model for better performance was the skewness of the dataset. From Table 1, we can see that the frequency of suspected and pathologic classes is much less than the normal class. If each class has equal regularization constant while training, the pathologic and suspected class sensitivity becomes very poor. But the goal of this study is to improve the sensitivity of pathologic and suspected classes even if we have to trade off the precision of normal class. That way, we may have more false positives in the normal class, but in return, we will reduce the number of false negatives in the pathologic and suspected class. To achieve that, balanced class weight for three classes should be used while training the model. When using balanced class weights, the classes with lower frequency will have higher class weight and thus, they will give a higher penalty for misclassification. The formula used for balanced class weight is:

Class weight for class n =

$$\frac{\text{Number of samples}}{\text{Number of classes} \times \text{Frequency of class}_n}$$

The parameters used while training the models were: maximum iterations = 50 (found by hyperparameter tuning), regularization constant = 0.8 (found by hyperparameter tuning), and chosen class weight so that each class has a balanced weight.

Random Forest (RF)

Plotting the learning curve for the RF algorithm showed that the algorithm yields a better validation score with more data (Ho, 1995). Validation curves were drawn for hyperparameter tuning. The chosen parameters for training random forest models were: maximum depth of trees = 10 (found by hyperparameter tuning), number of estimators = 500, and chosen class weight so that each class has a balanced weight.

Support Vector Machine (SVM)

Again plotting the learning curve for the SVM showed that the algorithm yields better validation scores with more data (Cortes & Vapnik, 1995). Validation curves were drawn for hyperparameter tuning. The chosen parameters for training SVM models were: maximum iterations = infinity (found by hyperparameter tuning), regularization constant = 0.1 (found by hyperparameter tuning), chosen class weight was so that each class has balanced weight, and kernel was set to linear (insight gained from the learning curve).

Extreme Gradient Boosting (XGB)

XGB differs from the RF by the method of training the trees. Each iteration of XGB greedily adds a new tree that improves the ensemble's performance, minimizing this objective function (Chen & Guestrin, 2016; Friedman, 2001). The optimum learning rate of XGB was found to be 0.1 by trial and error. The number of estimators was initially set to a very high value at 20000. Then Early Stopping Technique was used to detect the optimum number of estimators as the XGB models learn from the dataset. This method continuously checks an error metric of the validation set after adding each estimator to the ensemble. If the validation error keeps increasing for a certain number of consecutive iterations of the algorithm, the algorithm stops training and keeps only the estimators that reduce the error. The number of estimators to check until the training is stopped

was set to 20 as it yielded reasonable accuracy. Hence, the chosen hyperparameters were: learning rate = 0.1, early stopping rounds = 20, and error metric for early stopping method = Multiclass log loss.

K-Nearest Neighbors (KNN)

KNN algorithm can be used for supervised learning classification problems. This study used a machine learning library called 'scikit-learn' to apply KNN to the dataset (Pedregosa et al., 2011).

The optimum number of nearest neighbors based on which a data point will be predicted was set as 5 by trial and error. Also, the algorithm was set to take a weighted approach. The closer neighbors will have more influence on a query point than the distant points. So, the following hyperparameters were decided: several nearest points to be considered = 5, and nature of weights = Distance (closer neighbors of a query point will have a greater influence than neighbors further away).

Model Evaluation

Several performance metrics like accuracy, precision, sensitivity/recall, and F1-score/F-measure were used in this study to validate the machine learning models developed. These metrics are defined based on True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN).

Results

Ten models were constructed for comparison. Among them, five models were built using each of the five machine learning algorithms, which had all the dataset features as a descriptive variable.

The remaining five models were built using each of the five machine learning algorithms again but using the features selected by the MrMr algorithm. All of these models were built upon preprocessed dataset, and tuned hyperparameters. Five machine learning algorithms, namely LR, RF, SVM, XGB and KNN, were used to build these models. Performance metrics of these ten models are presented in Fig. 5.

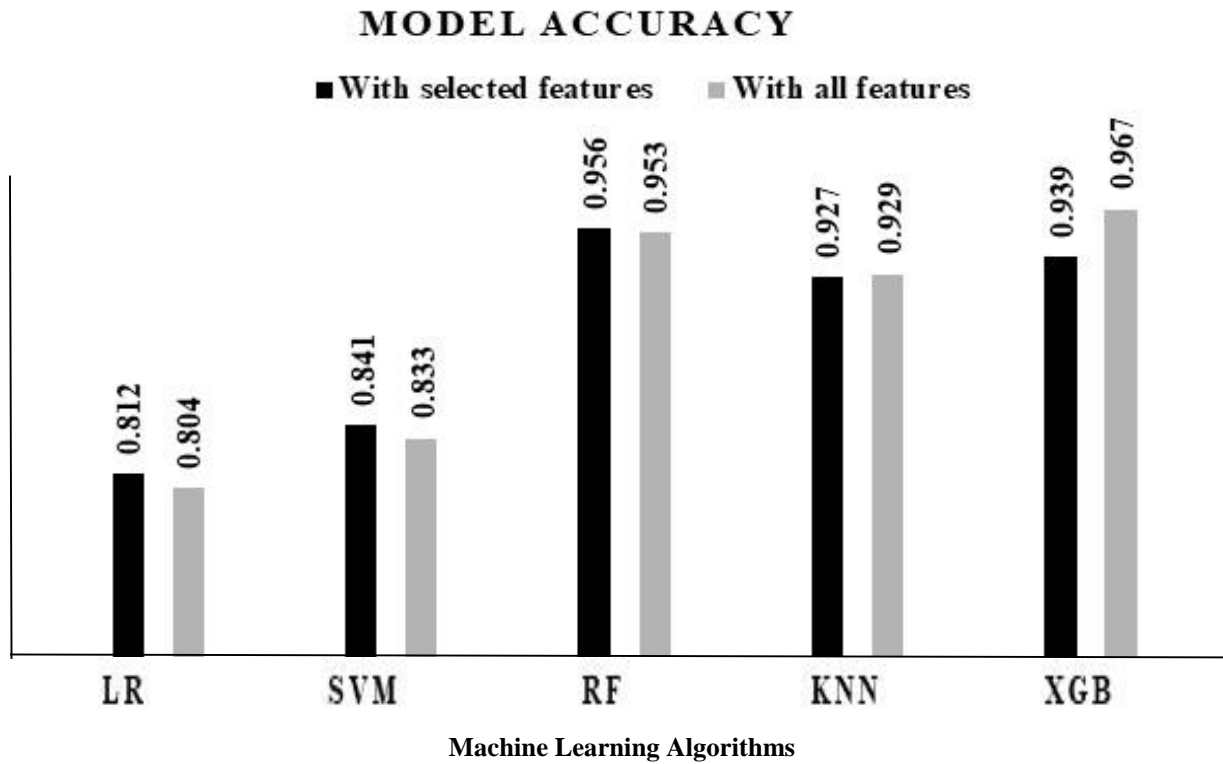


Fig. 5. Comparison of machine learning model accuracies based on features

Table 2. Performance metrics for Normal, Suspected and Pathologic classes.

CTG data Classes	Performance Metrics	Models with selected features					Models with all features				
		LR	SVM	RF	KNN	XGB	LR	SVM	RF	KNN	XGB
Normal Class	Precision	.976	.992	.983	.951	.960	.980	.984	.980	.960	.983
	Sensitivity	.802	.832	.970	.970	.964	.789	.822	.970	.970	.983
	F1-score	.880	.905	.977	.960	.962	.874	.896	.975	.965	.983
Suspected class	Precision	.407	.480	.766	.732	.724	.391	.427	.760	.725	.837
	Sensitivity	.767	.837	.878	.732	.724	.791	.814	.884	.707	.878
	F1-score	.532	.610	.818	.732	.724	.523	.560	.817	.716	.857
Pathologic class	Precision	.660	.630	1.00	.955	1.00	.673	.729	1.00	.885	1.00
	Sensitivity	.946	.919	.929	.740	.957	.946	.946	.892	.821	.929
	F1-score	.778	.747	.963	.840	.978	.787	.824	.943	.852	.963

According to Figure 5, the XGB model with all features has the highest accuracy of 0.967, and the RF model is second with 0.956 with selected features. Furthermore, the models examined in this study were multiclass classification models with three classes. Hence, precision, sensitivity, and F1-score were calculated for each class in all models. Table 2 displays the normal, suspected and pathologic class performance metrics.

From Table 2, the XGB model with all features performed very well with an F1-score of 0.983 in a normal class, 0.857 in suspected class, and 0.963 in pathologic class. RF model with selected features also performed exceptionally well with an F1-score of 0.977 in a normal class, 0.818 in suspected class, and 0.963 in pathologic class. However, we can see that all algorithms performed rather poorly for the suspected class. The suspected class contained data very similar to both normal and pathologic classes. Despite this drawback, we saw satisfactory performances from the XGB model with all features with an F1-score of 0.857. RF using only the selected features also had a similar F1-score of 0.818. Since reducing false positives in pathologic class and reducing false negatives in normal class is of paramount importance, we ruled out the poorly performing models developed with KNN, SVM and LR were not suitable for this work.

According to these findings, one can note that the XGB model with all features has the highest accuracy score of 96.7%, and the highest F1 scores for all three classes. Hence, the XGB model with all features is the best for classifying this CTG dataset. XGB with all features performed the best in all evaluation metrics as it combines the power of ensemble learning with gradient boosting, while RF only uses the concept of ensemble classifier.

Figure 6 shows the confusion matrix for the XGB model where XGB with all features has 0 false positives and only 2 false negatives in the pathologic class, while the support was 28. Only 5 false negatives were in the normal class while the support was 300.

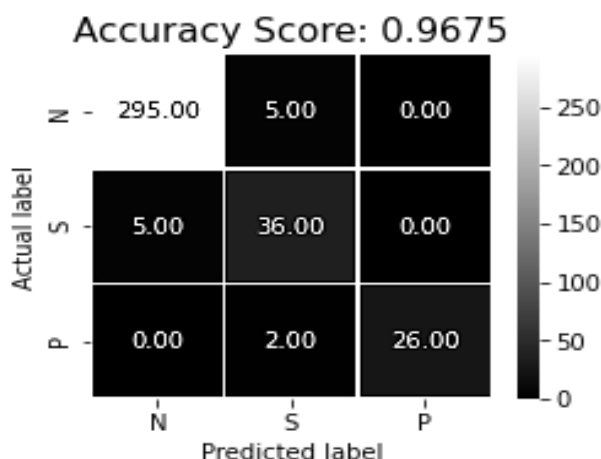


Fig. 6. Confusion matrix for XGB model with all features. Here, the labels N, S, and P represent normal, suspected, and pathologic classes.

RF model with selected features performed notably close enough to the best performing model with an accuracy score of 95.6%. The F1 scores in all three classes were similar to the best performing model. It is worth mentioning that the RF model with selected features used only 10 features instead of 23.

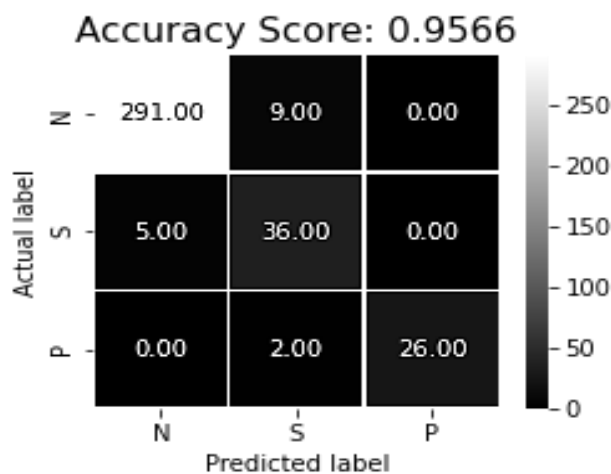


Fig. 7. Confusion matrix for RF model with selected features. Here, the labels N, S, and P represent normal, suspected, and pathologic classes.

Figure 7 shows the confusion matrix for this model. There were 9 false negatives in the Normal class with the support of 300. In the pathologic class, 2 false negatives were misclassified as suspected and no false positives.

ROC curves are drawn following one vs. all strategy. For instance, while drawing ROC curve for the normal class, the data labeled as usual are considered to be a positive class. In contrast, the other two classes (suspected and pathologic) combined are considered the negative class. ROC curves for the XGB model with all features and RF with selected features are given in Figure 8 and Figure 9 respectively. Figure 8 shows the ROC curves for the three classes while predicting with the XGB model with all features. From Figure 8, the AUC value for the normal class was calculated as 0.9878, which is satisfactory because it is close to 1. Furthermore, the AUC value for the suspected class was 0.9788, which is the worst among the three classes but still close to 1, i.e.,

XGB with all features gives satisfactory performance even in the suspected class. Finally, the XGB model had an almost perfect ROC curve for pathologic class with an AUC value of 0.9996, which is highly desirable as performance in the pathologic class is highly important.

If a pathologic fetus is misclassified as normal, there is a high risk of fetal death. On the other hand, if a normal fetus is classified as pathologic, there might be unnecessary medical intervention on a healthy fetus. Similarly, Figure 9 shows the ROC curve for the three classes while predicting the RF model with selected features. The AUC values for this model in normal, suspected, and pathologic classes are 0.9906, 0.9822, and 0.9991, respectively.

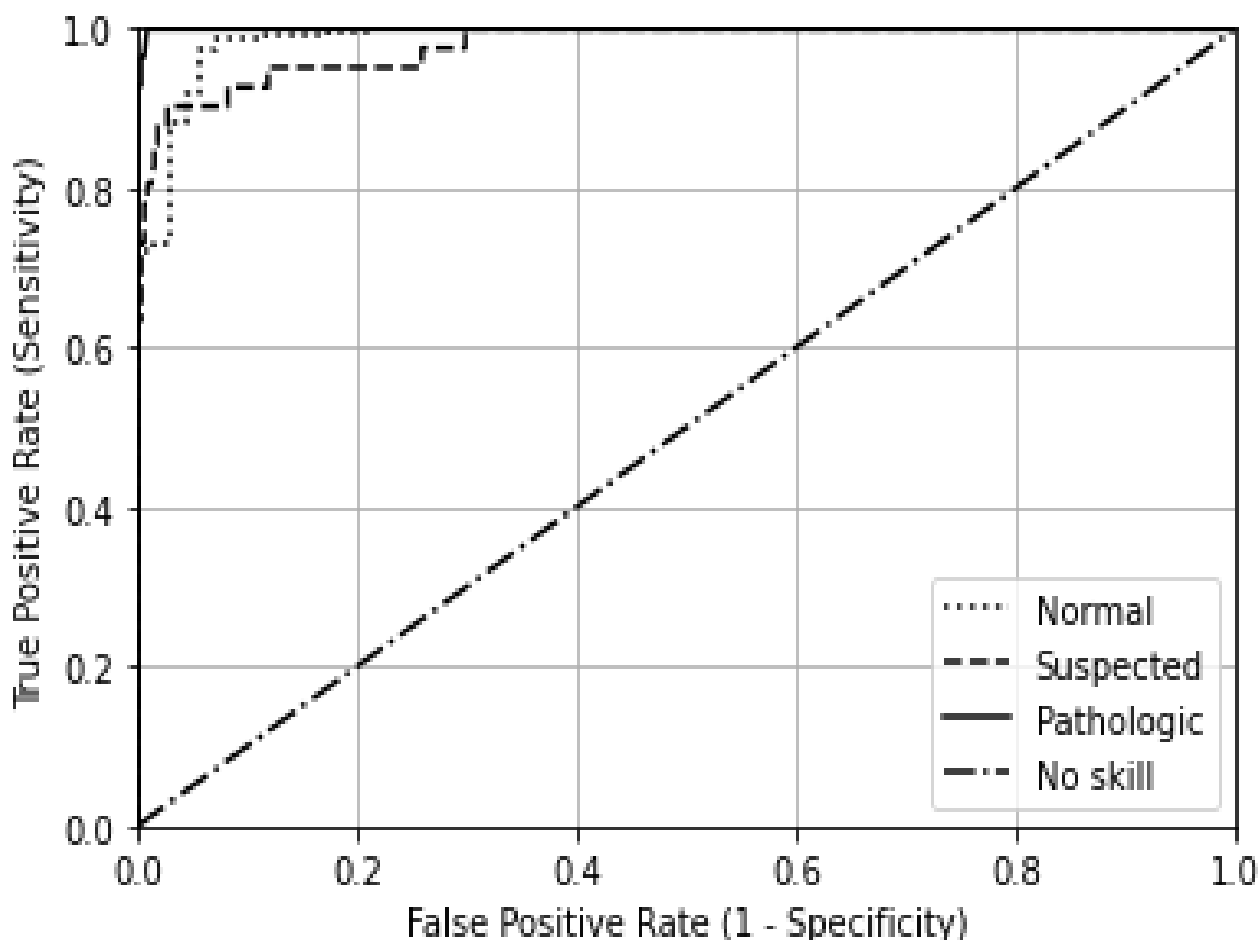


Fig. 8. ROC curves for XGB with all features model for normal, suspected and pathologic class.

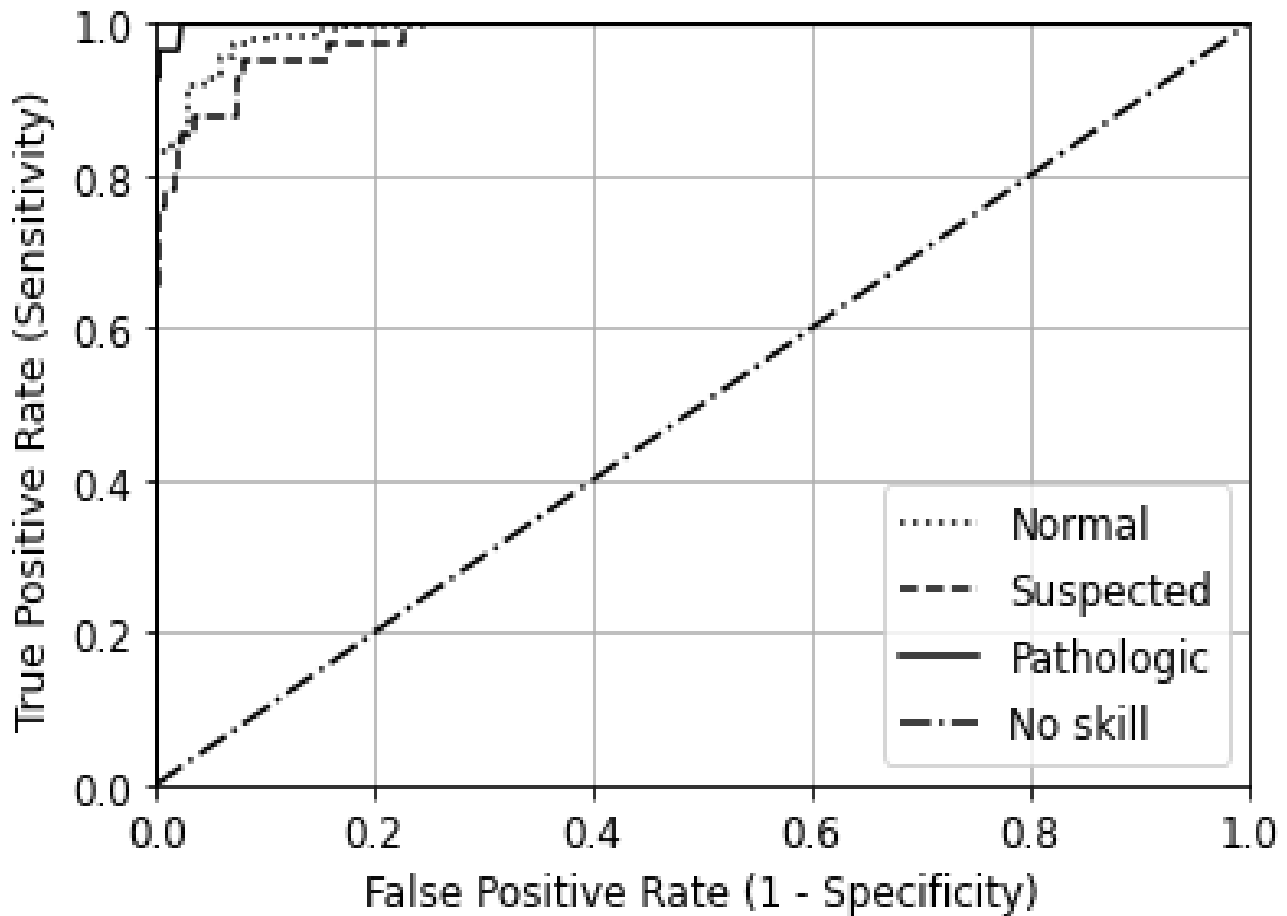


Fig. 9. ROC curves for RF with selected features model for normal, suspected and pathologic class.

Discussion

Automating the fetal health status detection process by analyzing CTG data is critical for preventing fetal and neonatal death. Automatic detection can lead to a correct decision taken by the physician in less amount of time. Also, it can reduce the pressure on doctors as multiple patients can be monitored by the system under the supervision of a single doctor. The doctors will be able to pay attention only to the critical patients. This study found two effective algorithms, XGB and RF, to perform this automation with 96.7% and 95.6% accuracy, respectively. Besides, other parameters such as F1-score, precision, recall, ROC and AUC were also taken into account. This was done because the dataset was highly skewed where the number of data points for Normal, Suspected, and Pathologic class is 1655, 295, and 176, respectively.

Since support for pathologic class is very low compared to normal class, special techniques such as balanced class weight were applied to get better metrics. The XGB model correctly classified 26 out of 28 data points despite the skewness of the dataset. This balanced performance across all classes is more important in this case rather than a high accuracy model that performs well only in the normal class.

Several studies have been conducted on classifying the CTG dataset from UCI (Ayres-de-campos *et al.*, 2000; Dua and Graff, 2017). A recent study using a bootstrap aggregating ensemble of random forest classifiers achieved 99.02% accuracy in the binary classification of the dataset into normal and abnormal data (Subasi *et al.*, 2020). In contrast, this study retained the suspected class in the dataset, solving a multiclass classification problem. This study also

used ensemble models. The best performing model in this study was the XGB classifier which achieved 96.70% overall accuracy for the three-class classification problem dividing the dataset into normal, suspected and pathologic classes. A similar study by Huang and Hsu achieved a maximum of 97.78% accuracy using artificial neural networks on all three classes (Huang and Hsu, 2012). Huang and Hsu also applied discriminant analysis and decision tree, which yielded 82.10% and 86.36% overall accuracy, respectively. Although the current work has a lower maximum accuracy than the study of Huang and Hsu, the performances of developed models using XGB and RF algorithms were comparable to neural networks. Additionally, XGB and RF algorithms have better model explainability and as better computational efficiency than artificial neural networks.

Sensitivity in the pathologic class is an important metric in this classification problem. Low sensitivity in the pathologic class means more pathologic fetuses will be classified as a normal fetuses and will not receive the desired medical attention. Yet, sensitivity in the pathologic class is hard to achieve in this dataset because of fewer data samples for pathologic fetuses. Due to special measures taken in this study, sensitivity was as high as 97.80% achieved in the pathologic class, which is a considerable improvement over 94.10% found in another study by Sahin and Subasi on the same CTG dataset (Sahin and Subasi, 2015). High sensitivity in pathologic class implies that the pathologic fetus will be identified and thus help the doctors make life-saving decisions for the mother and fetus.

In summary, this study used computation friendly machine learning algorithms to determine the optimal outcomes by using data preprocessing techniques, such as outlier detection, feature selection, and hyperparameter tuning, etc., while training the models. The use of balanced class weights on all the models also improved the overall performance rather than improving the overall accuracy only.

Conclusion

This study was motivated by the sustainable development goal of attaining zero stillbirth rate,

which is expected to be gained by 2030 in Bangladesh. Currently, we have biomedical instrumentation such as Cardiotocographs to monitor the fetal heart rate patterns. This tool is used to prevent stillbirths before and during the labor period of pregnancy. The goal here was to enhance the diagnostic capability of medical practitioners by automating the process of CTG signal interpretation. Five machine learning algorithms were used to build automatic prediction models. Extreme Gradient Boosting with all features and Random Forest with selected features produced useful models among those five algorithms. XGB with all features model had an accuracy of 96.7% and an F1 score of 0.963. Whereas RF with selected features model had an accuracy of 95.6% and an F1 score of 0.963. The F1 scores in the pathologic class for both models were 0.963, which was very close to 1, i.e., very high and realistic. Hence, we can conclude that incorporating the power of machine learning with Cardiotocography can lead to a more efficient and effective diagnosis of fetal heart conditions. In contrast to other similar studies, the performances of this study's predictive models were improved in terms of balanced performances in all three classes. Furthermore, the computational complexity of the machine learning models used in this study was much lower than neural networks, implying that these models can be run on embedded devices with low memory and processing power. These findings in this study thus have the potentials to be incorporated with the commercially available Cardiotocographs to interpret the outcomes better and thus help the doctors make better and quick decisions without continuous supervision. Last but not least, further work can be done to improve the performance of these models. Primary data can be collected from hospitals in Bangladesh and the models can be trained using the extended dataset, which is the next logical step of this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this article.

References

- Abir T, Agho KE, Ogbo FA, Stevens GJ, Page A, Hasnat MA, Dibley MJ and Raynes-Greenow C. Predictors of stillbirths in Bangladesh: evidence from the 2004–2014 nation-wide household surveys. *Glob. Health Action*, 2017; 10(1): 1410048.
- Ayres-de-campos D, Bernardes J, Garrido A, Marques-de-sá J and Pereira-leite L. SisPorto 2.0: A program for automated analysis of cardiotocograms. *J. Matern. Fetal Med.*, 2000; 9(5): 311-318.
- Chen T and Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016; p.785-794.
- Cortes C and Vapnik V. Support-vector networks. *Mach. Learn.*, 1995; 20(3): 273-297.
- CTG Case 19. Neoventa Medical. *Stancases*, 2012.
- CTG Case 20. Neoventa Medical. *Stancases*, 2012.
- Dua D and Graff C. *{UCI} Machine Learning Repository*, 2017.
- Fauveau V, Wojtyniak B, Mostafa G, Sarder AM and Chakraborty J. Perinatal mortality in Matlab, Bangladesh: A community-based study. *Int. J. Epidemiol.*, 1990; 19(3): 606-612.
- Friedman JH. Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 2001; 29(5).
- Frøen JF, Friberg IK, Lawn JE, Bhutta ZA, Pattinson RC, Allanson ER, Flenady V, McClure EM, Franco L, Goldenberg R L, Kinney MV, Leisher SH, Pitt C, Islam M, Khera A, Dhaliwal L, Aggarwal N, Raina N, Temmerman M and Leisher SH. Stillbirths: progress and unfinished business. *The Lancet*, 2016; 387(10018): 574-586.
- Heazell AEP, Siassakos D, Blencowe H, Burden C, Bhutta ZA, Cacciatore J, Dang N, Das J, Flenady V, Gold KJ, Mensah OK, Millum J, Nuzum D, O'Donoghue K, Redshaw M, Rizvi A, Roberts T, Toyin Saraki HE, Storey C and Budd J. Stillbirths: economic and psychosocial consequences. *The Lancet*, 2016; 387(10018): 604-616.
- Ho TK. Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1995; 1(1): 278-282.
- Huang ML and Hsu YY. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. *J. Biomed. Sci. Eng.* 2012; 5: 526-533
- Hug L, Mishra A, Lee S, You D, Moran A, Strong KL and Cao B. A neglected tragedy The global burden of stillbirths: report of the UN inter-agency group for child mortality estimation. In: *United Nations Children's Fund*, 2020.
- Liu FT, Ting KM and Zhou ZH. Isolation Forest. 2008 *Eighth IEEE International Conference on Data Mining*, 2008; 413-422.
- Mistry H and Heazell AEP, Vincent O and Roberts T. A structured review and exploration of the healthcare costs associated with stillbirth and a subsequent pregnancy in England and Wales. *BMC Pregnancy Childbirth*, 2013; 13: 236.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay É. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 2011; 12(85): 2825-2830.
- Peng H, Long F and Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005; 27(8): 1226-1238.
- Pettker CM and Campbell KH. 14- Antepartum Fetal Assessment. In C. A. Gleason & S. E. Juul (Eds.), *Avery's Diseases of the Newborn (Tenth Edition)* Elsevier. 2018; p.145-157.
- Sahin H and Subasi A. Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Appl. Soft Comput.*, 2015; 33: 231-238.
- Subasi A, Kadasa B and Kremic E. Classification of the Cardiotocogram Data for Anticipation of Fetal Risks using Bagging Ensemble Classifier. *Procedia Comput. Sci.*, 2020; 168: 34-39.
- Tarricone R. Cost-of-illness analysis: What room in health economics? *Health Policy*, 2006; 77(1): 51-63.
- Walker SH and Duncan DB. Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 1967; 54(1-2): 167-179.