

Machine Learning-Based Clinical Decision Support System for Predicting Fetal Hypoxemia during Non-Stress Test

Dissertation Submitted to



The IIHMR University, Jaipur

For the fulfilment for the award of the degree of
Master of Business Administration (MBA)

In

Hospital and Health Management

By

Dr. Sajal Baxi

Under the Supervision and Guidance of

Anoop Khanna

Machine Learning-Based Clinical Decision Support System for Predicting Fetal Hypoxemia during Non-Stress Test

Dissertation Submitted to



The IIHMR University, Jaipur

For the fulfilment for the award of the degree of
Master of Business Administration (MBA)

In

Hospital and Health Management

By

Dr. Sajal Baxi

Under the Supervision and Guidance of

Anoop Khanna

Declaration by the student

I hereby declare that this dissertation titled ‘Machine Learning-Based Clinical Decision Support System for Predicting Fetal Hypoxemia during Non-Stress Test’ is the bonafide record of my original research work. I declare that this has not been submitted to any other university or institution for the award of any degree or diploma. Information derived from the published or unpublished work of others has been duly acknowledged in the text.

(Signature)

Dr. Sajal Baxi

CERTIFICATE

This is to certify that dissertation titled ‘Machine Learning-Based Clinical Decision Support System for Predicting Fetal Hypoxemia during Non-Stress Test’ is a record of the research work undertaken by Dr. Sajal Baxi in partial fulfillment of the requirements for the award of the degree of MBA Hospital and Health Management from the IIHMR University, Jaipur under my guidance and supervision and his/her approach to the research study has been sincere, scientific, and analytical.

Faculty Guide
IIHMR University, Jaipur

Date

School Dean
IIHMR University, Jaipur

Date

Abstract

Background

Most under-five deaths occur within the first month after birth and intrapartum complications are a major contributor to the cause of death. These defects can be easily identified during the ante-natal check-up by use of a non-stress test. Due to the lack of availability of resources and medical experts in remote areas clinical decision support systems powered by machine learning models can provide information to the healthcare provider to make timely and better-informed decisions based on which course of treatment can be planned.

Aim

The study aims to develop an accurate and sensitive clinical decision support system model that can identify pathological fetuses based on the fetal heart rate recordings taken during the non-stress test.

Method

Foetal Heart rate recordings along with 10 other variables were collected from 1800 pregnant women in their third trimester. The data was put through a feature selection algorithm to identify important variables in the set. The data set was randomly divided into 2 independent random samples in the ratio of 70% for training and 30% for testing. After testing various machine learning algorithms based on specificity, sensitivity to accurately classify the fetus into normal, suspected, or pathological Random Forest algorithm was chosen, and using the “shiny” package of R the model was converted into a web-based platform for use by the health care provider.

Result

The fetal status determined by Obstetrician 77.85% observations from the normal category, 19.88% from the suspected category, and 8.28% from the pathological category. The Boruta algorithm revealed that all 11 independent variables in the data set were important to predict the outcome in the test set. In the training set the model had an accuracy of 99.04% and in the testing set accuracy was 94.7% (p-value= $< 2.2e-16$) with the precision of 97.56% to detect the pathological category.

Conclusion

With the ability of the model to accurately predict the pathological category the CDS can be used by healthcare providers in remote areas to identify high-risk pregnant women and take the decision on the medical care to be provided.

Key words- Clinical decision Support system, Machine Learning, Non-stress test, fetal hypoxemia

Acknowledgements

This study could not have been possible without the expertise of Dr. Anoop Khanna, IIHMR University, Jaipur, my thesis guide. I would also like to thank Nazaneen Saify, Esha Khadge of PharmaACE, Pune, who allowed me to work on the topic and provided me with their feedbacks for continuous improvements.

A debt of gratitude is also owed to Mr. Kirril Eremenko whose teaching modules on programming and data science in R, helped me learn the concepts of R and Machine Learning.

Last but not the least I would also like to thank my parents Col (Retd.) Salil Baxi and Mrs. Poonam Baxi, without you this project would be indeed impossible.

Table of Contents

Abstract.....	5
Background.....	5
Aim.....	5
Method.....	5
Result.....	5
Conclusion.....	5
1 Introduction.....	9
1.1 Non-Stress Test.....	9
1.2 Machine Learning.....	10
1.3 CDS classification.....	11
2 Review of literature.....	11
3 Data.....	14
3.1 Essential Attributes used.....	15
3.2 Distribution.....	16
3.3 Descriptive Statistics.....	16
3.4 Correlations.....	17
4 Methodology.....	17
4.1 Data Collection.....	17
4.2 Data Analysis and Model building.....	18
5 Algorithms.....	18
5.1 Feature Selection.....	18
5.2 Random Forest.....	18
6 Results.....	21
6.1 Feature selection.....	21
6.2 Important variables.....	21
6.3 Decision tree.....	22
6.4 Confusion matrix.....	22
6.5 Statistics by category.....	22
6.6 Overall Statistics.....	23
6.7 Error rates as a function of the number of trees.....	23
7 Comparison with reported results.....	24
8 Conclusions.....	24
9 R Code.....	24

10 References.....	24
--------------------	----

Table of Figures

Figure 4.1: Distribution of observations by class.....	16
Figure 6.1: Representation of correct and wrong classification.....	19
Figure 6.2: Random Forest for the training set.....	20
Figure 6.3: Random Forest for the Test set.....	20
Figure 7.1: Feature selection model.....	21
Figure 7.2: Decision tree model produced in R.....	22
Figure 7.3: Relation of prediction error with number of trees in the model.....	23

Table of Tables

Table 1.1: Readings categorization during non-stress test.....	10
Table 1.2 Classification of clinical decision support systems.....	11
Table 2.1 Review of literature.....	14
Table 3.1: Category of outcome.....	15
Table 3.2: List of variables and abbreviations.....	15
Table 3.3 Descriptive statistics.....	17
Table 3.4: Correlation matrix of the variables.....	17
Table 6.1: Ranking of variables by importance in prediction.....	21
Table 6.2: Confusion matrix of the model.....	22
Table 6.3: Prediction statistics by class.....	23
Table 7.1: Comparison with other studies.....	24

1 Introduction

1.1 Non-Stress Test

India in the year 2019-2020 recorded 824000 under-five deaths (1). Almost 24 out of 1000 neonatal deaths are due to fetal hypoxia associated with metabolic acidosis (45%) (2). Other leading causes of neonatal death are preterm birth complications (35%), intrapartum events (25%), and infections (10%) (3).

Fetal Hypoxia or Intra Uterine Hypoxemia is a condition in which the fetus does not receive adequate supply of oxygen (4). Such conditions can lead to irreversible damage to the Central Nervous System and even cause growth retardation (5). The readings are of two types Reactive and Non-Reactive (6).

A non-stress test is conducted in the 28th to 30th week of pregnancy to assess the fetal conditions. Non-stress test is a non-invasive procedure for screening pregnancies where there is a high risk of fetal hypoxia. Baseline fetal heart rate and fetal movements are the key features that are analysed during the test.

Indications:

- Growth restrictions to the fetus
- Diabetes
- Chronic Hypertension
- Multiple pregnancies
- SLE
- Decreased fetal movements.
- Maternal heart disease, chronic renal insufficiency, chronic liver disease, Maternal drugs abuse (7)

Contraindications:

In cases of placental abruptions and cord prolapse non-stress test cannot be used; in such cases immediate delivery is indicated (8).

Equipment:

The non-stress test is conducted using an electronic fetal monitor. It records fetal heart rate, patterns of the heart rate, contractions, maternal heart rate, maternal blood pressure. To monitor uterine contractions and fetal movements it has a Doppler transducer built into it. Two transducers one at the level of fetal heart and other at the abdomen of the mother is placed using belts (9).

Readings: (10)

Uterine Contractions: Normal ≤ 5 ; High ≥ 5

	Reassuring Feature	Non-Reassuring	Abnormal Feature
FHR (bpm)	110-160	100-109	<100 or >180
Variability (bpm)	≥ 5	<5 (40 min, less than 90 mins)	<5 (for more than 90 mins)

Decelerations (15 per min)	No deceleration	Deceleration up to 3 mins	Deceleration for more than 3 mins
----------------------------	-----------------	---------------------------	-----------------------------------

Table 1.1: Readings categorization during non-stress test

Interpretation:

The interpretation of NST is:

1. Normal Trace: Base line fetal heart rate in range of 110-160 beats pre minute, variability of less than 5 beats per minute and no decelerations. i.e., all features fall in the reassuring category.
2. Suspicious Trace: Detection of three reassuring features and one non- reassuring feature.
3. Pathological trace: More than two non-reassuring features out of four detected.

Most term fetus have many of these accelerations in each 20-30 minutes period of active sleep (11) . In case of a non-reactive stress test, the procedure is generally extended to another 20 minutes to distinguish between an asphyxiated fetus and one in prolonged sleep phase (12). False positive NST occur at a rate of 4-5 per 1000 (13). False positive results were mostly found in cases with pre-existing metabolic problems that can be associated to fetal macrosomia.

1.2 Machine Learning

Machine learning is the discipline of science wherein computers learn from data (14) It is a combination of statistics, computer science, and data relationships. The key focus is on developing efficient computer algorithms (15). The aim is to build statistical models that can accurately predict the outcome of a certain event based on millions and billions of data points by recognizing patterns and make decisions based on minimal human intervention (16) The machines can also be made able to self-correct based on feedback. Machine learning based on the type of model is classified into supervised, non-supervised, semi-supervised, and reinforced learning.

The ability of the machines to learn from large volumes of data finds great use in the field of healthcare (17) (18) (19). They can be used to provide physicians with accurate and UpToDate information from journals, clinical data, and various other sources (20) (21). This information can be critically important for taking clinical decisions related to diagnosis and treatment. (22) It has also been shown that the time to make the diagnosis and the accuracy with which they are made by machine learning models is very high (23) (24) (25). These highly accurate algorithms can act as clinical decision support systems to aid in day-to-day health care operations.

A computer aided system or program that can help a healthcare provider in taking decisions in a clinical setup is called as a clinical decision support system (26), Sim et al. defined CDS as, “a software that aids in the clinical decision making in which the patient characteristics are matched or compared to the previously established patient-specific assessment or recommendations” (27). The patient-specific data is fed into the system by the healthcare provider or the doctor. This information is processed and

linked to the existing database and uploaded into the base algorithm. After the analysis, the results are communicated back to the clinician (28).

CDS have been developed for many medical processes and have proven to be successful (29) for example in the prevention of deep vein thrombosis (30), increased compliance to glucose level adherence in severely diabetic patients (31), implementation of vaccine programs, and other preventive health care measures (32), identify cases of thrombocytopenia in ICU induced by drugs (33).

1.3 CDS classification

Write et al. (34) classified the CDS based on the purpose the clinical decision support system will serve:

Sr. No:	CDS Type	Examples
1	Medication Dosing support	Suggest type and quantity of dosage
2	Order and facilitators	Templates for admission and discharge in hospital
3	Point of care alerts and reminders	Show drug to drug interactions
4	Expert systems	Diagnostic tests and treatment decision planning
5	Relevant information display	Monitoring system for blood gases and trace elements
6	Workflow support	Patient transfer, admission, and discharges

Table 1.2 Classification of clinical decision support systems.

2 Review of literature

Sr. No	Name of the study	Author	Year	Results
1	Prediction of the neonatal state by computer analysis of the heart rate tracing	D.S Ayres de Campos	2018	The use of modern computational and data analytical methods has proved effective in neonatal heart rate monitoring and predictions.
2	A random forest classifier for lymph diseases	Azar, Elshazly A, El- Kornany	2014	High accuracy rates of random forest classifier and the problem of overfitting.
3	Effect of clinical decision support systems, A Systematic review	T.J Bright	2012	Types of clinical decision support systems and their use cases.
4	What is the confusion matrix in	J.Brownlee	2016	Basics of ML analysis and statistical terma

	machine learning			
5	Random forest methods for diabetic retinopathy.	Casanova, Saladana, Y. Chew	2013	High accuracy rates of random forest classifier and the problem of overfitting.
6	Using AI to improve hospital inpatient care.	N. Daniel	2013	AI-enabled healthcare monitoring systems show a lot of promise and help clinicians detect diseases at a very early stage
7	A computerized reminder system to increase the use of preventive care for hospitalized patients	PR Dexter	2005	Regular patient monitoring and data collection can provide essential insights based on which physicians can make decisions regarding disease and treatment.
8	Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment	S.E Dilsizian	2014	AI and ML-based systems have shown better accuracy in identifying rare diseases and radiological diagnoses when put-put against practicing physicians.
9	Clinical Decision Support Systems: State of the Art.	Retrieved from healthit.ahr.gov: http://healthit.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf	2014	Basics of clinical decision support systems and their applications in the field of healthcare.
10	Every Child Alive: The Urgent Need to End Newborn Deaths	T.U Fund	2018	Features to select from the CTG.
11	Diagnostic error in internal medicine	M.L Graber	2015	Physicians sometimes overlook a few features in making a judgment for the diagnosis which has led to many medicolegal cases.
12	Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in retinal fundus photographs	V Gulshan	2016	AI-based systems have been more successful in identifying small variations in radiographs which were often missed by physicians to diagnose a disease.
13	Use of an abnormal	L.M	2012	ML and its uses to effectively

	laboratory value-drug combination alert to detect drug-induced thrombocytopenia in critically ill patients.	Hartenstein		predict anomalies and
14	The threshold of metabolic acidosis associated.	J.A Low	2012	Deaths due to various intrapartum conditions during 2012 and their future projections
15	Adapting to Artificial Intelligence: radiologists and pathologists as information specialists.	S Jha	2016	AI-based systems have been more successful in identifying small variations in radiographs which were often missed by physicians to diagnose a disease.
16	Classification of fetal and neonatal heart	H.W Jongsma	1986	Classification of fetal conditions based on the heart rate.
17	Electronic alerts to prevent venous thromboembolism among hospitalized patients	N Kucher	2015	Data monitoring and analysis have proven to be an effective tool in treating patients in ICU and enabled treating doctors to identify high-risk patients from the beginning and allot resources accordingly.
18	Empirical Evaluation of the Performance of Feature Selection Approaches on Random Forest	SS Kumar	2017	High accuracy rates of random forest classifier and the problem of overfitting.
19	Cognitive and system factors contributing to diagnostic errors in radiology	C.S Lee	2013	AI-based systems have been more successful in identifying small variations in radiographs which were often missed by physicians to diagnose a disease.
20	Trees Weighting Random Forest Method for Classifying High-Dimensional Noisy Data	H.W Ding, B Wang, J. Dong	2010	High accuracy rates of random forest classifier and the problem of overfitting.
21	Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach.	G.Y Lip	2014	Heart-based risk factors and their clinical features that can be identified in fetuses.

22	national institute of child health and human development workshop report on electronic fetal monitoring	G.A Macones	2008	Classification of fetal conditions based on the heart rate.
23	The coming of age of artificial intelligence in medicine	V.L Patel	2014	Positive effects of AI use in medicine and challenges in the field.
24	The use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation.	E Rood	2005	Data monitoring and analysis have proven to be an effective tool in treating patients in ICU for proper glucose intake and adherence to the regime.
25	Linear and Non-Linear parameters for the analysis of fetal heart rate signal from CTG recordings	MM Signorini	2013	Parameters to be obtained from CTG analysis

Table 2.1 Review of literature

3 Data

Primary data set is obtained from UCI repository available as open-source document. The data set comprises 2100 pregnant women in their third trimester of pregnancy. It contains data on 11 separate attributes used in the measurement of Foetal Heart Rate and Uterine contractions during the antenatal check-up. A Cardiotocography machine is used to record these variables for 20 minutes. The International Federation of Obstetrics and Gynaecology guidelines (35) and the National Institute of Child Health and Human Development (36) considers these below listed features to make a diagnosis about the fetal health. 8 features are continuous and were obtained by averaging the values over 10 minutes. 3 other variables are discrete and obtained by calculations. The features for Foetal Heart Rate are baseline heart rate, number of accelerations per second, number of light, severe and prolonged decelerations per second. The features for the Uterine contractions include Uterine tone, contraction frequency, duration, and strength (37). These pregnant women were classified by their respective obstetricians into 3 distinct categories as following (38):

Category	Interpretation
Category 1	Normal trace
Category 2	Suspicious trace
Category 3	Pathological

Table 3.1: Category of outcome

3.1 Essential Attributes used

- Base line Fetal heart rate –Average value of stable Fetal heart rates taken over a 10-minute duration, which excludes accelerations and decelerations. Unit of measurement is beats per minute (bpm).
- Variability - There can be minor fluctuations in the FHR termed as baseline variabilities. Variation of 6-25 bpm is considered normal, reduced variability is 3 -5 bpm, absent variability is <3 bpm and abnormal variability is > 25 bpm (39).
- Accelerations – A 15 beats per minute increase for 15 seconds or greater time period over the baseline heart rate is termed as acceleration. The amplitude and durations of the accelerations is less as compared to adults (40).
- Decelerations – Any transient decrease in the baseline heart rate by 15 bpm lasting for more than 15 seconds is a deceleration. Decelerations are always caused by some physiological factor (41) . They can be classified as:
 - Light deceleration: They occur before the peak of the uterine contractions when the baby's head gets compressed. They generally are not harmful.
 - Severe deceleration: They do not begin until the peak of uterine contraction or at the end of the contractions. They follow the same pattern of the contraction which caused them. They imply a lack of oxygen supply to the fetus, if the magnitude is high a caesarean section is indicated to save the fetus.
 - Prolonged deceleration: Decelerations in FHR by 15 bpm for more than 90 seconds but less than 5 minutes.
- Uterine Contractions – Also called as Braxton Hicks Contractions are a common occurrence that begin from 28th to 30th week of pregnancy. They are characterised by contraction of the muscles in the uterine lining. They are mild (0.5 in 10 minutes) in nature during this period but become severe during pregnancy 10 to 20 per minute.

Attribute	Interpretation
LB	Baseline Fetal heart rate (bpm)
ACC	Accelerations per second
FM	Fetal movements per second
UC	Uterine contractions per second
LD	Decelerations (Light) per second
SD	Decelerations (Severe) per second
PD	Decelerations (Prolonged) per second
LTV	Duration (%) of abnormal long-term variability
ASV	Duration (%) of short-term variability
MSV	Average short- term variability
MLV	Average long-term variability
Health_stat	Fetal category (1= Normal; 2= Suspected; 3= Pathological)

Table 3.2: List of variables and abbreviations

3.2 Distribution

The data has 77.85% observations from the normal category, 19.88% from the suspected category, and 8.28% from the pathological category.

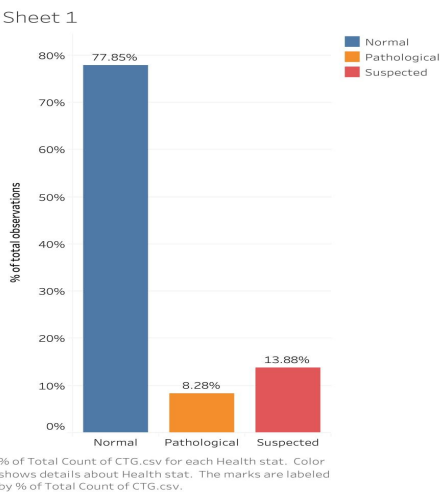


Figure 3.1: Distribution of observations by class

3.3 Descriptive Statistics

	LB	ACC	FM	UC	LD	SD	PD	LTV	ASV	MSV
nbr.val	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800

Table 3.3 Descriptive statistics

3.4 Correlations

Variable	1	2	3	4	5	6	7	8	9	10
LB										
ACC	-.08**									
	[-.12, -.04]									
FM	-.03	.05*								
	[-.08, .01]	[.01, .09]								
UC	-.15**	.09**	-.07**							
	[-.19, -.10]	[.05, .13]	[-.11, -.03]							
LD	-.16**	-.11**	.05*	.29**						
	[-.20, -.12]	[-.15, -.07]	[.01, .09]	[.25, .32]						
SD	-.05*	-.04*	-0.01	0.01	.11**					
	[-.10, -.01]	[-.09, -.00]	[-.05, .03]	[-.04, .05]	[.07, .15]					
PD	-.10**	-.13**	.27**	.08**	.23**	0.01				
	[-.15, -.06]	[-.17, -.09]	[.23, .30]	[.03, .12]	[.18, .27]	[-.03, .05]				
LTV	.31**	-.28**	-.10**	-.23**	-.12**	0.03	.05*			
	[.27, .34]	[-.32, -.24]	[-.15, -.06]	[-.27, -.19]	[-.16, -.08]	[-.01, .08]	[.00, .09]			
ASV	-.28**	.21**	.12**	.29**	.56**	0.03	.27**	-.43**		
	[-.32, -.24]	[.17, .25]	[.08, .16]	[.25, .33]	[.53, .59]	[-.01, .08]	[.23, .31]	[-.46, -.40]		
MSV	.29**	-.37**	-.07**	-.31**	-.27**	-0.03	-.14**	.46**	-.47**	
	[.25, .32]	[-.41, -.34]	[-.12, -.03]	[-.34, -.27]	[-.31, -.23]	[-.07, .01]	[-.18, -.10]	[.43, .49]	[-.50, -.44]	
MLV	-0.03	-.14**	0.01	-.07**	-.24**	-0.04	-.23**	-.32**	.07**	-.17**
	[-.07, .01]	[-.18, -.10]	[-.03, .05]	[-.11, -.02]	[-.28, -.20]	[-.08, .00]	[-.27, -.19]	[-.35, -.28]	[.03, .12]	[-.21, -.13]

Table 3.4: Correlation matrix of the variables

The degree of correlation between the variables does not have a high magnitude but are statistically relevant in many instances so it can be inferred that there is no multi collinearity among the variables.

4 Methodology

4.1 Data Collection

An elastic belt containing two plates is placed around the abdomen of the mother. It is necessary that the plates are inContact with the skin. One plate measures the fetal heart rate and the other measures the intensity and number of uterine contractions. The belt is

connected to the cardiotocograph machine which interprets the signals from the plates and shows them in the screen attached. The fetal heart sounds can be heard as the bating and pulsating sounds made by the machine. The duration of the procedure is around 10 to 20 minutes so that all the cycles are properly recorded. The machine also gives out the readings in a printed format which can be analyzed to calculate different observations required to make the diagnosis about the fetal conditions.

4.2 Data Analysis and Model building

The dataset was first put into feature selection process to identify those variables that were important to predict the outcome using the ‘Boruta’ package in R. Subsequently the data set was divided into two independent random sets using random sampling, consisting of a training and test set with observations in a ratio of 70:30. The model was trained using the training set in which the outcome variable i.e., fetal health status was known to the machine. After sufficient tuning of the model by increasing the number of trees and the cut-off values, the model was applied on the test set with the outcome i.e., fetal health status hidden from the machine. The predicted and the actual outcome in the test set were then compared to calculate the accuracy, specificity, and sensitivity of the model.

5 Algorithms

5.1 Feature Selection

Machine learning-based classifiers have widely been used in healthcare decision making including screening of patients, diagnosis of a disease, and predicting the final prognosis and outcome of the treatment. Medical diagnosis is based on various factors and their interplay this may lead to multicollinearity and overfitting which lead to incorrect results. Therefore, it is necessary to select an appropriate number of variables for developing the model. Feature selection is employed in the development of a machine learning model to improve the performance for either classification or regression (42). The study uses the Boruta algorithm for conducting the feature selection as it has previously shown a high degree of accuracy in selecting the relevant variables based on their importance which is decided by their contribution to prediction (43). Feature selection is carried in two stages, stage 1 involves the formation of feature selection algorithms to reduce the number of dimensions and the multicollinearity in the dataset such as Principal Component Analysis (PCA), Sequential Forward Floating Search (SFFS), Boruta. Stage 2, construction of the model using the obtained subsets from stage 1 (44). In the analysis conducted in this study, all the 11 variables recorded which are base fetal heart rate, acceleration, fetal movements, number of uterine contractions, light decelerations, severe deceleration, prolonged decelerations, abnormal short-term variability, and percentage time of long-term variability were found to be important for prediction of the dependant variable that was the fetal health status.

5.2 Random Forest

Random Forest algorithm a form of the supervised machine learning algorithm, which utilizes multiple randomly grown decision trees based on the input sample and the split nodes of the given data set. The model has a very high classification and generalization ability and has thus been widely used in various domains. The trees are evaluated based

on the Out-of-bag error estimate of the training subset of the data (45) . In the below figure 6.1 (46) trees built from the same dataset is represented. The grey and the black dots represent the leaf nodes that give the output variable. Majority is calculated by averaging the regression results in each node. For any event, $A \subset \Omega$ of the sample space, the Indicator function, I_A is defined by, $I_A(x) = 1$, if $x \in A$ else 0,

Assuming there are n samples in the data with d features and classes C_1 and C_2 It can be represented by

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

Assuming there are S samples at the current node that need to be partitioned

$$P(S_j) = |S_j| / |S|$$

$$P(C_j/S_j) = |S_j \cap C_j| / |S_j|$$

$$g(S_j) = \sum P(C_j/S_j)(1 - P(C_j/S_j))$$

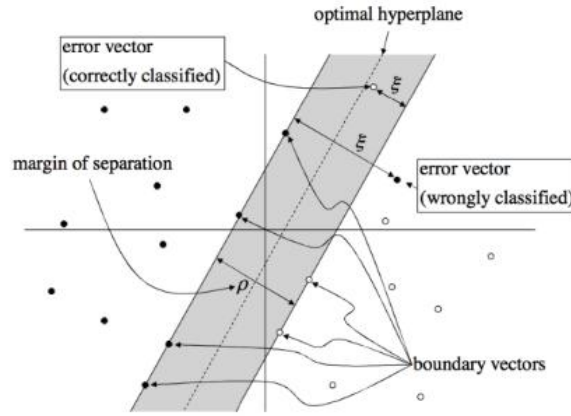


Figure 5.1: Representation of correct and wrong classification

Greatest variation is observed if S_j is divided equally amongst C_j and if S_j is just one of the C_j then the value of variation will be lowest. The entire sample is present in the top node and it branches into lower node also called as children until the lowest node present at the base contains only the category variable. Diversity at each child node is kept at a minimum by selecting a feature x_j and the associated threshold a . Gini criteria is used to measure diversity.

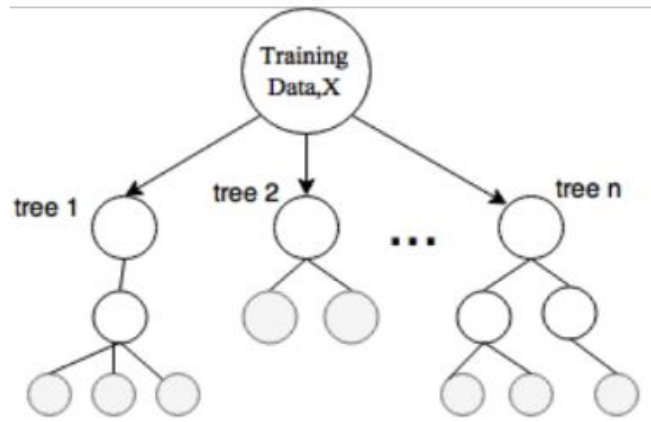


Figure 5.2: Random Forest for the training set

Gini Index is given as,

$$G = P(S1) g(S1) + P(S2) g(S2)$$

Where S1 and S2 are the splits at each subsequent child node.

In a classification problem such as this the final output from majority of the trees is taken as the predicted value y . To prevent overfitting in the model splitting of trees should only be done till the point where the next split does not add to the overall accuracy in prediction.

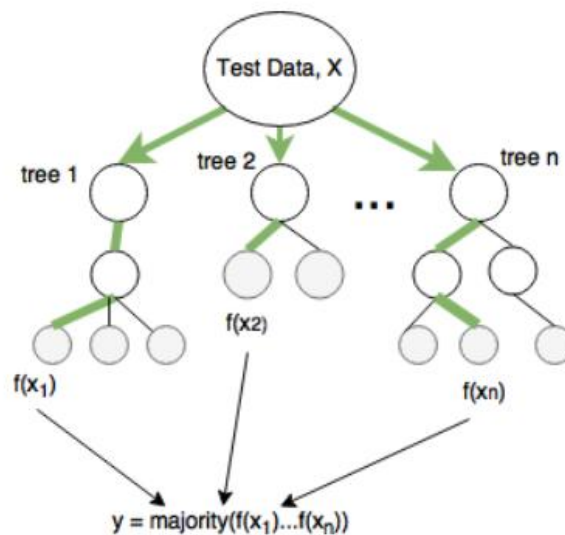


Figure 5.3: Random Forest for the Test set

Random Forest classifier also can rank variables in the data set as per their importance in which they affect the model performance (47). Features ranked as per their importance for prediction by the model.

6 Results

The performance of a classifier is evaluated based on predicted classification values in the test set in comparison to the actual values.

6.1 Feature selection

All 11 independent variables were found to be important to predict the fetal health category. Features coloured in green are categorized as important when compared with the shadow variables which include minimum, maximum, and mean of the data frame.

Based on this result all 11 variables can be used to develop the prediction model.

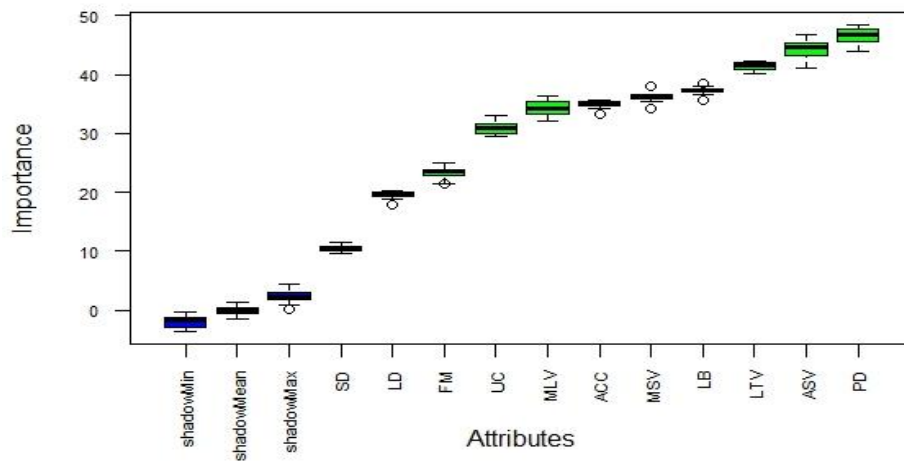


Figure 6.1: Feature selection model

6.2 Important variables

The algorithm rates the dependent variables used in the data based on their impact on predicting the final value.

Rank	Name of variable	Mean Decrease Gini
1	Decelerations (Prolonged) per second	107.92
2	Duration (%) short-term variability	90.11
3	% time with abnormal long-term variability	80.93
4	Baseline Fetal heart rate	58.75
5	Average short-term variability	56.08
6	Accelerations per second	53.04
7	Average long-term variability	33.99
8	Uterine Contractions per second	33.93
9	Fetal movements per second	21.11
10	Decelerations (Light) per second	16.97
11	Decelerations (Severe) per second	1.58

Table 6.1: Ranking of variables by importance in prediction.

6.3 Decision tree

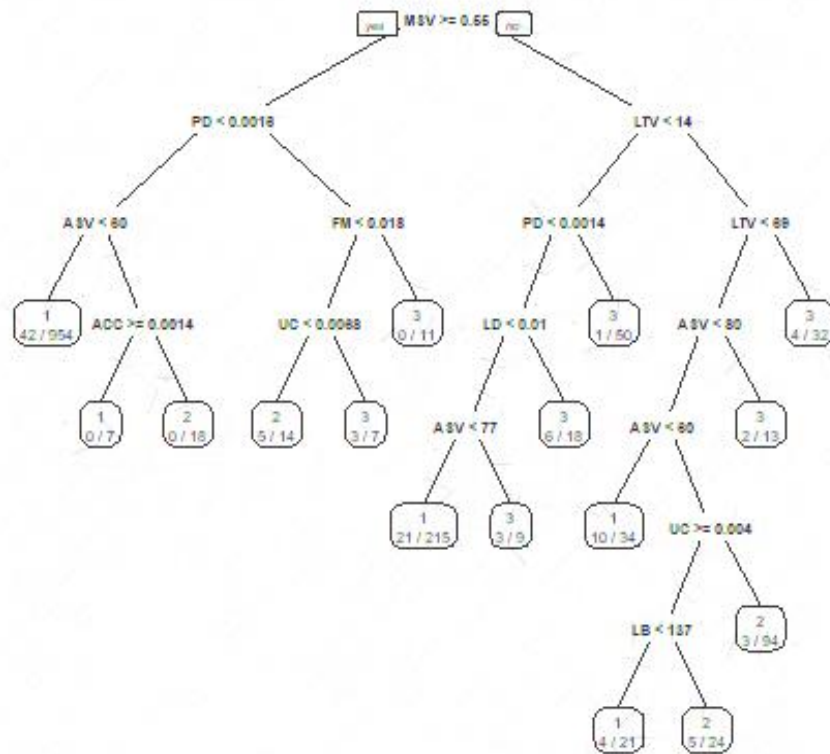


Figure 6.2: Decision tree model produced in R

6.4 Confusion matrix

It is a table that shows how the classification model is confused when predicting various classes (48). A confusion matrix can be used in a multilevel classifier by dividing the sum of diagonals by the sum of the table.

The model produced the following confusion matrix on the test set containing 605 (30% of total) observations.

	Predicted class 1 (Normal)	Predicted class 2 (Suspected)	Predicted class 3 (Pathological)
Actual class1	468	21	2
Actual class 2	5	65	3
Actual class 3	1	0	40

Table 6.2: Confusion matrix of the model

6.5 Statistics by category

	(Normal)	(Suspected)	(Pathological)
Sensitivity	0.9873	0.7558	0.8889
Specificity	0.8244	0.9846	0.9982
Positive prediction value	0.9532	0.8904	0.9756

Negative prediction value	0.9474	0.9605	0.9913
Prevalence	0.7835	0.1421	0.0743
Detection rate	0.7736	0.1074	0.0661
Detection prevalence	0.8116	0.1207	0.0677
Balanced accuracy	0.9059	0.8702	0.9435

Table 6.3: Prediction statistics by class

6.6 Overall Statistics

1. Accuracy: 0.9471
2. 95% Confidence Interval: (0.9261, 0.9635)
3. P-Value < 2.2e-16
4. Kappa: 0.8453
5. McNamara's Test P-Value: 0.004264

6.7 Error rates as a function of the number of trees

The out-of-bag error rate stabilizes after 400 trees in the forest model and any further increase does not significantly reduce the predictability error.

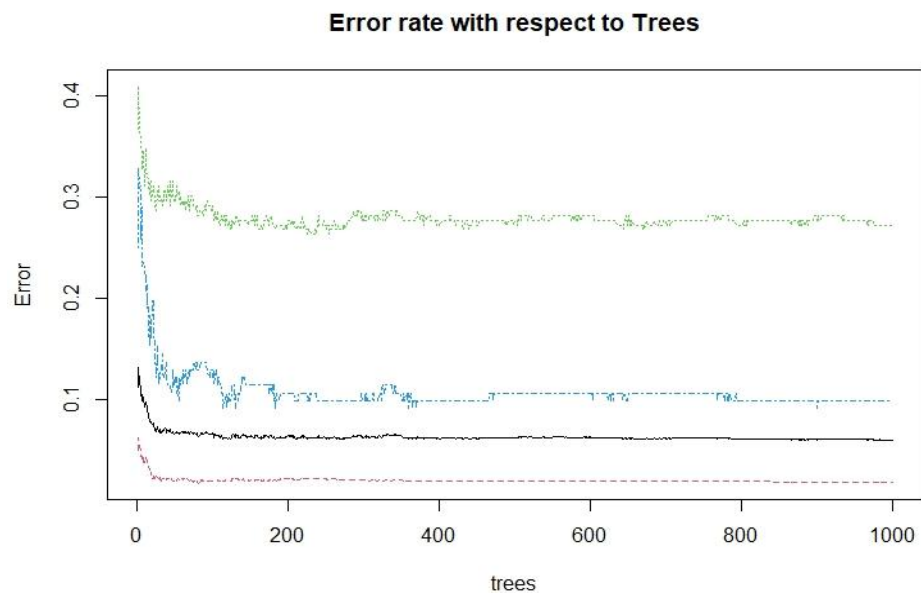


Figure 6.3: Relation of prediction error with number of trees in the model.

7 Comparison with reported results

Sr. No	Reference	Method	Result (Category 3)	
			Accuracy	Specificity
1	Sundar et al 2013 (49)	Neural network-based classifier	0.91	0.90
2	Menai Mohder et al 2013 (50)	Relief F-15	0.939	0.958
3	Jezewski NSKI et al. (51)	LSVM classifier	0.90	0.92
4	Chen et al 2012 (52)	FG- K means	0.76	0.81
5	Zhou and Sun 2014 (53)	Active learning of Gaussian process	0.89	0.79
6	Cruz et al (54)	Meta- DES Ensemble Classifier	0.846	-
7	This paper	Random forest with feature selection	0.94	0.99

Table 7.1: Comparison with other studies

8 Conclusions

The data collected during antenatal check-up is extremely important to the obstetrician and the Gynaecologist to diagnose the condition of the growing fetus. Only visual diagnosis of the data could not always be objective and correct. The chances of making an error in diagnosis increase if the healthcare provider is not well trained such as in remote and rural areas that lack efficient healthcare providers and infrastructure. Therefore, use of clinical decision support tools can be beneficial. The supervised machine learning model Random forest (total number of trees in the forest = 1000) used in this study has shown an overall accuracy of 94.71% and p-value $\leq 2.2e-16$ which indicates the statistically significant results of the model to predict the health category. After 400 trees the model does not show any significant reduction in the error rate. The model has a sensitivity and specificity of 0.88 and 0.99 for predicting category 3 that is the pathological condition of the fetus. Hence the model can easily distinguish between the three categories of fetal health, and the result can be used to take a course of action that can save the life of the fetus and the mother during parturition. Thereby reducing the infant mortality rate and the maternal mortality rate.

9 R Code

```
data <- read.csv(file.choose())  
data$Health_stat <- ifelse(test = data$Health_stat == 1)
```

```
data$Health_stat <- as.factor(data$Health_stat)
```

```
#Descriptive Statistics
```

```
str(data)
```

```
install.packages("pastecs")
```

```
library(pastecs)
```

```
df <- stat.desc(data)
```

```
df
```

```
write.csv(df,"stats_Health.csv")
```

```
install.packages("descr")
```

```
library(descr)
```

```
descry (data,  
        headings = FALSE,
```

```
        stats = "common"
```

```
)
```

```
help(descr)
```

```
install.packages("apaTables")
```

```
library(apaTables)
```

```
apa.cor.table(M, "apa Correlation.doc")
```

```
#partition the data
```

```
set.seed(111)
```

```
ind <- sample(2,nrow(data), replace = TRUE, prob= c(0.7,0.3))
```

```
training <- data[ind == 1,]
```

```
testing <- data[ind == 2,]
```

```
#####  
#####
```

```
#FEATURE SELECTION
```

```
library(Boruta)
```

```
library(mlbench)
```

```
library(caret)
```

```

library(randomForest)

set.seed(111)

boruta <- Boruta(fetal_health ~ ., data = data, doTrace = 2, maxRuns = 500)
print(boruta)
plot(boruta, las = 2, cex.axis = 0.7)

#blue boxes are shadow attributes, important attributes should have higher performance
than these

#Red boxes are not important, yellow are yet to be decided, Green are important
plotImpHistory(boruta)

# Tentative Fix
bor <- TentativeRoughFix(boruta)
print(bor)
attStats(boruta)

#scatterplot and correlations
library(psych)
pairs.panels(training[,-12],
              gap = 0,
              bg = c("red", "yellow", "blue")[training$fetal_health],
              pch=21)
pairs.panels(training[,-12])

#Form principle components
pc <- prcomp(training[,-12],
              center = TRUE,
              scale. = TRUE)

print(pc)
summary(pc)

```

```

#orthogonality of PC
pairs.panels(pc$x, gap = 0, bg = c('red', 'yellow', 'blue')
            [training$fetal_health], pch=21)

#bi-plots
library(devtools)
install_github("vqv/ggbiplot")
library(ggbiplot)

g <- ggbiplot(pc, obs.scale = 1, var.scale = 1, groups = training$fetal_health,
             ellipse = TRUE, circle = TRUE, ellipse = 0.68)
g <- g + scale_color_discrete(name="")
g <- g + theme(legend.direction = 'horizontal', legend.position = 'top')
plot(g)

#predictions with PC
trg <- predict(pc, training)
trg <- data.frame(trg, training[12])
trg<- trg[-13]
tst <- predict(pc, testing)
tst <- data.frame(tst, testing[12])

#####
#####

#Applying the models
#1. Multinomial Logistic Regression
library(nnet)
data$fetal_health <- relevel(data$fetal_health , ref = '1')
mod_multinom <- multinom(fetal_health ~ . , data = training)
summary(mod_multinom)

#check for accuracy

```

```

fitted.proBABilities <- predict(mod_multinom, testing, type = "class")
misClassError <- mean(fitted.proBABilities != testing$fetal_health)
misClassError
accuracy_multinom <- 100 - (misClassError*100)

#Confusion Matrix
pred <- predict(mod_multinom , training)
table <- table(pred , training$fetal_health)
acc.train <- sum(diag(table))/sum(table)

pred.test <- predict(mod_multinom , testing)
tab.test <- table(pred.test , testing$fetal_health)
acc.test_multinom <- sum(diag(tab.test))/sum(tab.test)
# Multinomial logistic regression accuracy(test) = 85.9%

#Performance Evaluation
install.packages("ROCR")
library(ROCR)
pred <- predict(mod_multinom, training, type = "class")
head(pred)
pred <- as.factor(pred)
hist(pred)
P <- prediction(pred , training$fetal_health)
eval_multinom <- performance(pred, "acc")
plot(eval)

#ROC curve
roc <- performance(pred, "tpr", "fpr")
plot(roc,
      colorize = TRUE,
      main = "ROC Curve",
      ylab = "Sensitivity",

```

```

      xlab = "1-Specificity")
abline(a=0 , b=1)
#AUC
auc <- performance(pred, "auc")
auc <- unlist(slot(auc, "y.values"))
auc
#####
#####

```

#2.Decision tree and Random forest:

```

library(rpart)
tree <- rpart(Health_stat ~ . , data = training,
              ntree = 500,
              mtry = 8,
              importance = TRUE,
              proximity = TRUE)
print(tree)
plot(tree)
library(rpart.plot)
rpart.plot(tree)
prp(tree, extra = 3)
predict(tree , testing, type = 'prob')

library(randomForest)
rf.model <- randomForest(Health_stat ~ . ,ntree =1000, proximity = TRUE,
                        data = training)
plot(rf.model, main = 'Error rate with respect to Trees', legend( ))
summary(rf.model)
rf.model

rf.model$importance

```

```
varImpPlot(rf.model)
```

```
#Confusion Matrix
```

```
pred_rf.model <- predict(rf.model , training)
```

```
table <- table(pred_rf.model , training$fetal_health)
```

```
acc.train <- sum(diag(table))/sum(table)
```

```
pred.test_rf <- predict(rf.model , testing)
```

```
tab.test <- table(pred.test_rf , testing$fetal_health)
```

```
acc.test_rf <- sum(diag(tab.test))/sum(tab.test)
```

```
#Random forest accuracy(test) = 94.7%
```

```
#Accuracy of prediction of pathological cases = 40/41 = 97.5
```

```
summary(tab.test)
```

10 References

1. **Bank, The World.** Mortality rate, under 5. *data.worldbank.org*. [Online] 2020. <https://data.worldbank.org/indicator/SH.DYN.MORT>.
2. *Threshold of metabolic acidosis associated.* **JA Low, BG Lindsay, EJ Derrick.** 2019, *Am J Obstet Gynecol*, p. 1391.
3. *Every Child Alive: The Urgent Need to End Newborn Deaths.* **Fund, The United Nations Children's.** 2018, The United Nations Children's Fund.
4. *The effects of ante- and postnatal hypoxia on the central nervous system and their correction with peptide hormones.* **Maslova MV, Maklakova AS, Sokolova NA, Ashmarin IP, Goncharenko EN, Krushinskaya YV.** July 2003, *Neuroscience and Behavioral Physiology*, Vol. 6, pp. 607-11.
5. *[Intrauterine hypoxia and sudden infant death syndrome.* **Habek D, Habek JC, Jugović D, Salihagić A.** 2002, *Acta Medica Croatica*, Vol. 56(3), pp. 109-18.
6. *ACOG guidelines on antepartum fetal surveillance.* **Preboth, M.** s.l. : American College of Obstetrician and Gynecologists, September 1, 2000, *Am Fam Physicians*, Vol. 5, pp. 62(50):1184, 1184-8.
7. *The effects of maternal opium abuse on fetal heart rate.* **F Keikha, FG Vandani, S Latifi.** s.l. : Iran Journal of Medical sciences, Nov 2016, Vol. 6, pp. 41(6):479-485.
8. *Perinatal mortality in diabetic patients undergoing antepartum fetal evaluation.* **A Brecher, T Williams, L Baxi.** Dec 2002, *Journal of Maternal, Fetal and Neonatal Medicine*, pp. 12(6): 423-427.
9. *Intrapartum cardiotocography with and without computer analysis.* **M Campanile, P D'Allesandro, L Della Corte.** Jul 7, 2020, *Journal of maternal fetal and neonatal Medicine* , pp. 33(13):2284-2290.
10. *Antepartum fetal heart rate testing: Non stress test as a promary approach.* **Keegan KA, RH Paul.** Jan 1980, *American Journal of Obstetrician and Gynecologists* , pp. 01;136(1):75-80.
11. *Antepartum fetal heart rate testing: An Out patient approach.* **KA Keegan, RH Paul, D McCart.** January 1980, *American Journal of Obstetrician and Gynecologists* , pp. 01;136(1):81-3.
12. *Accelerations of the normal fetal heart rate at 38 to 40 weeks gestational age.* **J Patrick, L Carmichael, L Chess, C Staples.** Jan 1984, *American Journal of Obstetrician and Gynecologists* , pp. 01;148(1):35-41.
13. *Antepartum fetal evaluation: development of a fetal biophysical profile.* **FA Manning, LD Platt, LSipos.** March 15, 1980, *American Journal of Obstetrician and Gynecologists* , pp. 136(6):787-95.
14. *Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation.* **Lip, G. Y., Nieuwlaat, R., Pisters, R., Lane, D. A., & Crijns, H. J.** 2010, *Chest*, pp. 137(2), 263–272.

15. *A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD).* O'Mahony C, Jichi F, Pavlou M, Monserrat L, Anastasakis A, Rapezzi C, Biagini E, Gimeno JR, Limongelli G, McKenna WJ, Omar RZ, Elliott PM. 2014, Eur Heart J, pp. 35:2010–2020.
16. machine learning. SAS. [Online] 2015.
https://www.sas.com/en_in/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention..
17. *Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment.* Dilsizian SE, Siegel EL. 2014, Curr Cardiol Rep, p. 16:441.
18. *The coming of age of artificial intelligence in medicine.* Patel VL, Shortliffe EH, Stefanelli M, et al. 2009, Artif Intell Med, pp. 46:5–17.
19. *Adapting to Artificial Intelligence: radiologists and pathologists as information specialists.* Jha S, Topol EJ. 2016, JAMA, pp. 316:2353–4.
20. *Epidemiology of medical error.* Weingart SN, Wilson RM, Gibberd RW, et al. 2000, BMJ, pp. 320:774–7.
21. *Diagnostic error in internal medicine.* Graber ML, Franklin N, Gordon R. 2005, Arch Intern Med, pp. 165:1493–9.
22. *Diagnostic errors in the intensive care unit: a systematic review of autopsy studies.* Winters B, Custer J, Galvagno SM, et al. 2012, BMJ Qual Saf, pp. 21:894–902.
23. *Cognitive and system factors contributing to diagnostic errors in radiology.* Lee CS, Nagy PG, Weaver SJ, et al. 2013, AJR Am J Roentgenol, pp. 201:611–7.
24. *Using artificial intelligence to improve hospital inpatient care.* DB, Neill. 2013, IEEE Intell syst, pp. 28:92–5.
25. *Development and Validation of a Deep Learning Algorithm for detection of Diabetic Retinopathy in retinal fundus photographs.* Gulshan V, Peng L, Coram M, et al. 2016, JAMA, pp. 316:2402–10.
26. *Computer programs to support clinical decision making.* EH, Shortliffe. 1987, JAMA, pp. 258(1):61–6.
27. *Clinical decision support systems for the practice of evidence-based medicine.* Sim I, Gorman P, Greenes RA, et al. 2001, Journal of the American Medical Informatics Association: JAMIA, pp. 8(6):527–34.
28. ES, Berner. Clinical Decision Support Systems: State of the Art. *healthit.ahrq.gov*. [Online] May 4, 2014.
http://healthit.ahrq.gov/sites/default/files/docs/page/09-0069-EF_1.pdf.
29. *Effect of Clinical Decision-Support SystemsA Systematic Review.* Bright TJ, Wong A, Dhurjati R, et al. 2012, Ann Intern Med, pp. 157(1):29–43.
30. *Electronic alerts to prevent venous thromboembolism among hospitalized patients.* Kucher N, Koo S, Quiroz R, et al. 2005, N Engl J Med, pp. 352(10):969–77.

31. *A computerized reminder system to increase the use of preventive care for hospitalized patients.* Dexter PR, Perkins S, Overhage JM, Maharry K, Kohler RB, McDonald CJ. 2005, N Engl J Med, pp. 12(2):172–80.
32. *Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation.* Rood E, Bosman RJ, van der Spoel JI, Taylor P, Zandstra DF. 2005, Journal of the American Medical Informatics Association: JAMIA, pp. 12(2):172–80.
33. *Use of an abnormal laboratory value-drug combination alert to detect drug-induced thrombocytopenia in critically ill patients.* Harinstein LM, Kane-Gill SL, Smithburger PL, Culley CM, Reddy VK, Seybert AL. 2012, J Crit Care, pp. 27(3):242–9.
34. *Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems.* Wright A, Sittig DF, Ash JS, et al. 2011, Journal of the American Medical Informatics Association: JAMIA, pp. 18(3):232–42.
35. al., Kristen M. Shellenberg et. FIGO global voice for women's health. *FIGO*. [Online] <https://obgyn.onlinelibrary.wiley.com/doi/full/10.1002/ijgo.12235>.
36. *The 2008 national institute of child health and human development workshop report on electronic fetal monitoring: Update on definitions, interpretation, and research guidelines.* Macones GA, Hankins GD, Spong CY, Hauth J, Moore T. 2008, pp. 37:510–5.
37. *Linear and Non Linear paramenters for the analysis of foetal heart rate signal form CTG recordings.* Signorini, M.G. Magenes, G. Cerutti and Arduini D. 3, s.l. : IEEE Trans., 2003, Vol. 50.
38. *Classification of fetal and neonatal heart.* Jongsma, H. W., and Nijhuis, J. G. 1986, Eur. Obstet. Gynecol. Reprod., pp. 21:293–299.
39. hospital, The Women's The royal womens. Cardiotocographic interpretations and response. *thewomes.r.worldssl.net*. [Online] 2018. https://thewomens.r.worldssl.net/images/uploads/downloadable-records/clinical-guidelines/ctg-interpretation-and-response_280720.pdf.
40. *Assessing Fetal Wellbeing A Practical Guide.* Baker L, Beaves M, and Wallace E. 2016, Southern Health and RANZCOG, Vol. 3.
41. *Fetal Heart Monitoring: What's Normal, What's Not?* *healthline Parenthood*. [Online] Johns Hopkins Medicine Health Library. <https://www.healthline.com/health/pregnancy/abnormal-fetal-heart-tracings>.
42. Kumar, Smitha S and Shaikh, Talal. Empirical Evaluation of the Performance of Feature Selection Approaches on Random Forest. [Online] 2017. [Cited: 1 27, 2021.] <https://researchportal.hw.ac.uk/en/publications/empirical-evaluation-of-the-performance-of-feature-selection-appr>.
43. Rudnicki, Witold R., Wrzesień, Mariusz and Paja, Wieslaw. All Relevant Feature Selection Methods and Applications. [Online] 2015. [Cited: 1 27, 2021.] https://link.springer.com/chapter/10.1007/978-3-662-45620-0_2.
44. *A random forest classifier for lymph diseases.* Azar, Ahmad Taher, et al. 2, 2014, Computer Methods and Programs in Biomedicine, Vol. 113, pp. 465–473.

45. Li, Hong Bo, et al. Trees Weighting Random Forest Method for Classifying High-Dimensional Noisy Data. [Online] 2010. [Cited: 1 27, 2021.] <http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ieee-000005704290>.
46. *Prediction of neonatal state by computer analysis of the heart rate tracing.* Ayres-de-campos, D'costa- Santos. 2005, Eur. J. Obstet, pp. 52-60.
47. *Application of Random Forests Methods to Diabetic Retinopathy Classification Analyses.* Casanova, Ramon, et al. 6, 2014, PLOS ONE, Vol. 9.
48. Brownlee, Jason. What is a Confusion Matrix in Machine Learning. *Machine learning mastery*. [Online] November 16, 2016. <https://machinelearningmastery.com/confusion-matrix-machine-learning/>.
49. *An overview of research challenges for classification of cardiogram data.* C. Sundar, M chitradevi and G Geetharamani. 2, s.l. : Journal of computer science, Vol. 9, p. 198.
50. *Influence of feature selection on naive Bayes Classifier for recognizing patterns in coardiotocograms.* MEB Menai, FJ Mohder and F Al-muttairi. 37, s.l. : Journal of Medical sytems, 2013, Vol. 2, pp. 1-9.
51. *Some practical remarks on neural network approach to fetal cardiogram classification.* M. Jezewski, Wrobel M, P. Leski, N. Henzel, J. Jezewski. s.l. : Engineering in medicine and biology socciety, EBMS, 2007. 29th Annual International Conference of the IEEE.
52. *Electric fetal heart rate monitoring and its relationship with neonatal and infant mortality.* H. Y. Chen, SP Chauhan, CV Ananth. 2012, American journal of Obstetrics and Gynecology, pp. 204(6), 491- e1.
53. *Active learning of the Gaussian process with manifold-preserving graph reduction.* Sun, J Zhou and S. 7-8, s.l. : Neural Computer Application, 2014, Vol. 25, pp. 1615-1625.
54. *META-DES: A dynamic ensemble selection for meta learning.* RM. Cruz, R. Sabourin, GD Cavalcanti. 2015, Patt. Recognition, pp. 48(5), 1925-1935.