ORIGINAL RESEARCH PAPER

Health Science

KEYWORD: Clinical decision Support system, Machine Learning, Non-stress test, fetal hypoxemia

MACHINE LEARNING BASED CLINICAL DECISION SUPPORT SYSTEM TO PREDICT FETAL HYPOXIA IN WOMEN DURING ANTE-NATAL CHECK-UP.

Dr. Sajal Baxi

IIHMR University, 1, Prabhu Dayal Marg, near Sanganer Airport, Maruti Nagar, Jaipur, Rajasthan, 302029

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.

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.

1 INTRODUCTION

1.1 Non-Stress Test

India in the year 2019-2020 recorded 824000 under-five deaths (Bank, 2020). Almost 24 out of 1000 neonatal deaths are due to fetal hypoxia associated with metabolic acidosis (45%) (JA Low, 2019). Other leading causes of neonatal death are preterm birth complications (35%), intrapartum events (25%), and infections (10%) (Fund, 2018).

Fetal Hypoxia or Intra Uterine Hypoxemia is a condition in which the fetus does not receive adequate supply of oxygen (Maslova MV, 2003). Such conditions can lead to irreversible damage to the Central Nervous System and even cause growth retardation (Habek D, 2002). The readings are of two types Reactive and Non-Reactive (Preboth, 2000).

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 (F Keikha, 2016)

CONTRAINDICATIONS:

In cases of placental abruptions and cord prolapse non-stress test cannot be used; in such cases immediate delivery is indicated (A Brecher, 2002).

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 (M Campanile, 2020).

Readings: (Keegan KA, 1980)

Uterine Contractions: Normal =< 5; High >= 5

Table *1.1*: Readings Categorization During Non-stress Test

	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

Interpretation:

The Interpretation Of Nst Is:

1. Normal Trace: Base line fetal heart rate in range of 110-160 beats per 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 (KA Keegan, 1980) . 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 (J Patrick, 1984). False positive NST occur at a rate of 4-5 per 1000 (FA Manning, 1980). 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 (Lip, 2010) It is a combination of statistics, computer science, and data relationships. The key focus is on developing efficient computer algorithms (O'Mahony C, 2014). 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 (machine learning, 2015) 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 (Dilsizian SE, 2014) (Patel VL, 2009) (Jha S, 2016). They can be used to provide physicians with accurate and UpToDate information from journals, clinical data, and various other sources (Weingart SN, 2000) (Graber ML, 2005). This information can be critically important for taking clinical decisions related to diagnosis and treatment. (Winters B, 2012) 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 (Lee CS, 2013) (DB, 2013) (Gulshan V, 2016). 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 (EH, 1987), 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" (Sim I, 2001). 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 (ES, 2014).

CDS have been developed for many medical processes and have proven to be successful (Bright TJ, 2012) for example in the prevention of deep vein thrombosis (Kucher N, 2005), increased compliance to glucose level adherence in severely diabetic patients (Dexter PR, 2005), implementation of vaccine programs, and other preventive health care measures (Rood E, 2005), identify cases of thrombocytopenia in ICU induced by drugs (Harinstein LM, 2012).

2 CDS CLASSIFICATION

Write et al. (Wright A, 2011) classified the CDS based on the purpose the clinical decision support system will serve:

Table 3.1: Classification Of Clinical Decision Support Systems

Sr. No:	CDS Type.5	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

3 DATA

Primary data set is obtained from UCI repository available as open source document. The data set comprises 1800 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 (al., n.d.) and the National Institute of Child Health and Human Development (Macones GA, 2008) 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 (Signorini, 2003). These pregnant women were classified by their respective obstetricians into 3 distinct categories as following (Jongsma, 1986):

Table 4.1 : Category of outcome

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

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 (hospital, 2018).
- 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 (Baker L, 2016).
- 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" (Fetal Heart Monitoring: Whats Normal, Whats Not?, n.d.) .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 28^{th} to 30^{th} 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.

Table4.2: List OfVariables And Abbreviations

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)

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.

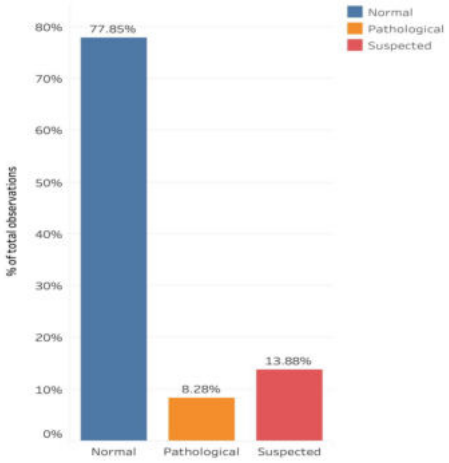

Figure 4.1: Distribution Of Observations By Class

3.3 Descriptive Statistics

Table 4.3: Descriptive Statistics

	LB	ACC	FM	UC	LD	SD	PD	LTV	ASV	MSV	MLV
nbr.val	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800	1800
nbr.na	0	0	0	0	0	0	0	0	0	0	0
min	106	0	0	0	0	0	0	12	0.2	0	0
max	160	0.019	0.481	0.015	0.015	0.001	0.005	87	7	91	50.7
range	54	0.019	0.481	0.015	0.015	0.001	0.005	75	6.8	91	50.7
sum	283404	6.757	20.156	9.283	4.017	0.007	0.337	99901	2833.5	20934	17406.9
median	133	0.002	0	0.004	0	0	0	49	1.2	0	7.4
mean	133.3039	0.003178	0.009481	0.004366	0.001889	3.29E-06	0.000159	46.99012	1.332785	9.84666	8.187629
SE.mean	0.213428	8.38E-05	0.001012	6.39E-05	6.42E-05	1.24E-06	1.28E-05	0.372877	0.019156	0.39899	0.122065
CI.mean.0.95	0.418549	0.000164	0.001985	0.000125	0.000126	2.44E-06	2.51E-05	0.731242	0.037566	0.782453	0.23938
var	96.84222	1.49E-05	0.002178	8.68E-06	8.76E-06	3.28E-09	3.48E-07	295.5928	0.780115	338.4452	31.67716
std.dev	9.840844	0.003866	0.046666	0.002946	0.00296	5.73E-05	0.00059	17.19281	0.883241	18.39688	5.628247
coef.var	0.073823	1.216257	4.922186	0.674711	1.566692	17.40278	3.721746	0.365881	0.662704	1.868337	0.687409

3.4 Correlations

Table 4.4: Correlation Matrix Of The Variables

Variable	1	2	3	4	5	6	7	8	9	10
LB										
ACC	-.08**									
	[-.12, -.04]									
FM	-0.03	.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]

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 (Kumar & Shaikh, 2017). 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 (Rudnicki, Wrzesie□, & Paja, 2015). 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 (Azar, Elshazly, Hassanien, & El-Korany, 2014). 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 (Li, Wang, Ding, & Dong, 2010) . In the below figure 6.1 (Ayres-de-campos, 2005) 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 of the sample space, the Indicator function, I am defined by, IA (x) = 1, if x ε 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

$D = \{(x1, y1), (x2, y2)....,(xn, yn)\}$

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

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

$P(Cj/Sj) = |Sj \cap Cj| / |Sj|$

$g(Sj) = \Sigma P(Cj/Sj)(1 - P(Cj/Sj))$

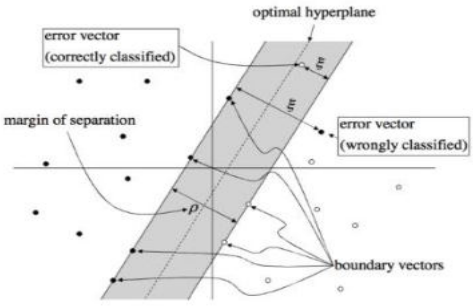

FIGURE 6.1: REPRESENTATION OF CORRECT AND WRONG CLASSIFICATION

Greatest variation is observed if Sj is divided equally amongst Cj and if Sj is just one of the Cj 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 xj and the associated threshold a. Gini criteria is used to measure diversity.

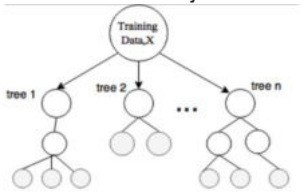

FIGURE 6.2: RANDOM FOREST FOR THE TRAINING SET

Gini Index is given as, $G = P(S1)\, g(S1) + P(S2)\, g(S2)$

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.

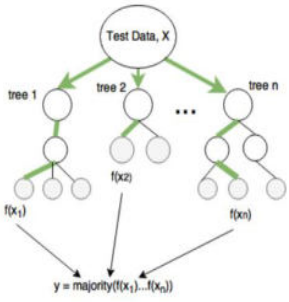

Figure 6.3: Random Forest For The Test Set

Random Forest classifier also can rank variables in the data set as per their importance in which they affect the model performance (Casanova, et al., 2014). 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.

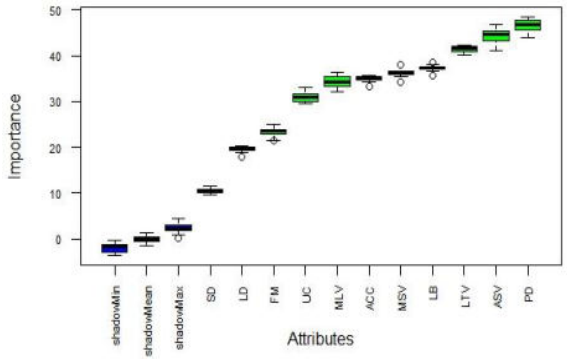

FIGURE 7.1: FEATURE SELECTION MODEL

6.2 Important Variables

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

Table 7.1: Ranking Of Variables By Importance In Prediction.

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

6.3 DECISION TREE

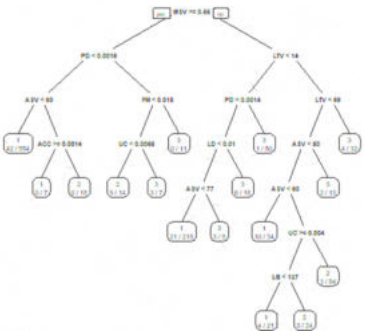

Figure 7.2: Decision Tree Model Produced In R

6.4 Confusion Matrix

It is a table that shows how the classification model is confused when predicting various classes (Brownlee, 2016). 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.

Table 7.2: Confusion Matrix Of The Model

	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

6.5 Statistics by category

Table 7.3: Prediction Statistics By Class

	(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

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.

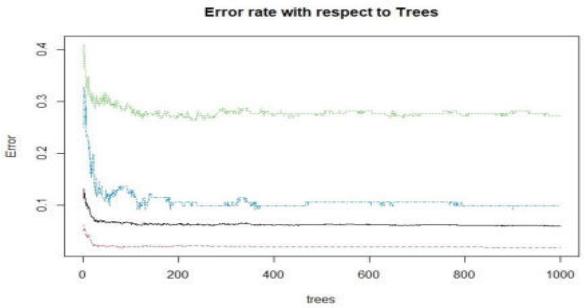

Figure 7.3: Relation Of Prediction Error With Number Of Trees Is The Model

7 Comparison With Reported Results

Table 7.1: Comparison with other studies

Sr. No	Reference	Method	Result (Category 3)	
			Accuracy	Specificity
1	Sundar et al 2013 (C. Sundar)	Neural network-based classifier	0.91	0.90
2	Menai Mohder et al 2013 (MEB Menai, 2013)	Relief F-15	0.939	0.958

3	Jezewski NSKI et al. (M. Jezewski, 2007)	LSVM classifier	0.90	0.92
4	Chen et al 2012 (H.Y. Chen, 2012)	FG- K means	0.76	0.81
5	Zhou and Sun 2014 (Sun, 2014)	Active learning of Gaussian process	0.89	0.79
6	Cruz et al (RM. Cruz, 2015)	Meta- DES Ensemble Classifier	0.846	-
7	This paper	Random forest with feature selection	0.94	0.99

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 =< 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)

#orthognality 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,
         elipse = 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. A Brecher, T. W. (2002, Dec). Perinatal mortality in diabetic patients undergoing antepartum fetal evaluation. Journal of Maternal, Fetal and Neonatal Medicine, 12(6):423-427.
2. al., K. M. (n.d.). FIGO global voice for women's health. Retrieved from FIGO: https://obgyn.onlinelibrary.wiley.com/doi/full/10.1002/ijgo.12235
3. Ayres-de-campos, D. S. (2005). Predicton of neonatal state by computer analysis of the heart rate tracing. Eur. J. Obstet, 52-60.
4. Azar, A. T., Elshazly, H. I., Hassanien, A. E., & El-Korany, A. (2014). A random forest classifier for lymph diseases. Computer Methods and Programs in Biomedicine, 113(2), 465-473. Retrieved 1 27, 2021, from https://sciencedirect.com/science/article/pii/s0169260713003751
5. Baker L, B. M. (2016). Assessing Fetal Wellbeing A Practical Guide. Southern Health and RANZCOG, 3.
6. Bank, T. W. (2020). Mortality rate, under 5. Retrieved from data.worldbank.org: https://data.worldbank.org/indicator/SH.DYN.MORT
7. Bright TJ, W. A. (2012). Effect of Clinical Decision-Support SystemsA Systematic Review. Ann Intern Med, 157(1):29–43.
8. Brownlee, J. (2016, November 16). What is a Confusion Matrix in Machine Learning. Retrieved from Machine learning mastry: https://machine learningmastery.=com/confusion-matrix-machine-learning/
9. C. Sundar, M. c. (n.d.). An overview of research challenges for classification of cardiotogram data. 9(2), 198.
10. Casanova, R., Saldana, S., Chew, E. Y., Danis, R. P., Greven, C. M., & Ambrosius, W. T. (2014). Application of Random Forests Methods to Diabetic Retinopathy Classification Analyses. PLOS ONE, 9(6). Retrieved 1 27, 2021, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098587
11. DB, N. (2013). Using arti□cial intelligence to improve hospital inpatient care. IEEE Intell syst, 28:92–5.
12. Dexter PR, P. S. (2005). A computerized reminder system to increase the use of preventive care for hospitalized patients. N Engl J Med, 12(2):172–80.
13. Dilsizian SE, S. E. (2014). Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep, 16:441.
14. EH, S. (1987). Computer programs to support clinical decision making. JAMA, 258(1):61–6.
15. ES, B. (2014, May 4). Clinical Decision Support Systems: State of the Art. Retrieved from healthit.ahrg.gov: http://healthit.ahrq.gov/sites/default/files/docs/page/09–0069–EF_1.pdf
16. F Keikha, F. V. (2016, Nov). The effects of maternal opium abuse on fetal heart rate. 6, 41(6):479-485.
17. FA Manning, L. P. (1980, March 15). Antepartum fetal evaluation: development of a fetal biophysical profile. American Journal of Obstetrician and Gynecologists , 136(6):787-95.
18. Fetal Heart Monitoring: What's Normal, What's Not? (n.d.). (Johns Hopkins Medicine Health Library) Retrieved from healthline Parenthood: https://www.healthline.com/health/pregnancy/abnormal-fetal-heart-tracings
19. Fund, T. U. (2018). Every Child Alive: The Urgent Need to End Newborn Deaths. The United Nations Children's Fund.
20. Graber ML, F. N. (2005). Diagnostic error in internal medicine. Arch Intern Med, 165:1493–9.
21. Gulshan V, P. L. (2016). Development and Validation of a Deep Learning Algorithm for detection of Diabetic Retinopathy in retinal fundus photographs. JAMA, 316:2402–10.
22. H. Y. Chen, S. C. (2012). Electric fetal heart rate monitoring and its relationship with neonatal and infant mortality. American journal of Obstetrics and Gynecology, 204(6), 491-e1.
23. Habek D, H. J. (2002). [Intrauterine hypoxia and sudden infant death syndrome. Acta Medica Croatica, 56(3), 109-18.
24. Harinstein LM, K.-G. S. (2012). Use of an abnormal laboratory value-drug combination alert to detect drug-induced thrombocytopenia in critically Ill patients. J Crit Care, 27(3):242–9.
25. hospital, T. W. (2018). Cardiotocographic interpretations and response. Retrieved from thewomes.r.worldssl.net: https://thewomens.r.worldssl.net/images/uploads/downloadable-records/clinical-guidelines/ctg-interpretation-and-response_280720.pdf
26. J Patrick, L. C. (1984, Jan). Accelerations of the normal fetal heart rate at 38 to 40 weeks gestational age. American Journal of Obstetrician and Gynecologists , 01;148(1):35-41.
27. JA Low, B. L. (2019). Threshold of metabolic acidosis associated. Am J Obstet Gyneco, 1391.
28. Jha S, T. E. (2016). Adapting to Artificial Intelligence: radiologists and pathologists as information specialists. JAMA, 316:2353–4.
29. Jongsma, H. W. (1986). Classification of fetal and neonatal heart. Eur. Obstet. Gynecol. Reprod., 21:293–299.
30. KA Keegan, R. P. (1980, January). Antepartum fetal heart rate testing: An Out patient approach. American Journal of Obstetrician and Gynecologists , 01;136(1):81-3.
31. Keegan KA, R. P. (1980, Jan). Antepartum fetal heart rate testing: Non stress test as a promary approach. American Journal of Obstetrician and Gynecologists , 01;136(1):75-80.
32. Kucher N, K. S. (2005). Electronic alerts to prevent venous thromboembolism among hospitalized patients. N Engl J Med, 352(10):969–77.
33. Kumar, S. S., & Shaikh, T. (2017). Empirical Evaluation of the Performance of Feature Selection Approaches on Random Forest. Retrieved 1 27, 2021, from https://researchportal.hw.ac.uk/en/publications/empirical-evaluation-of-the-performance-of-feature-selection-appr
34. Lee CS, N. P. (2013). Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol, 201:611–7.
35. Li, H. B., Wang, W., Ding, H. W., & Dong, J. (2010). Trees Weighting Random Forest Method for Classifying High-Dimensional Noisy Data. Retrieved 1 27, 2021, from http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ieee-000005704290
36. Lip, G. Y. (2010). 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. Chest, 137(2), 263–272.

```
#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. A Brecher, T. W. (2002, Dec). Perinatal mortality in diabetic patients undergoing antepartum fetal evaluation. Journal of Maternal, Fetal and Neonatal Medicine, 12(6):423-427.
2. al., K. M. (n.d.). FIGO global voice for women's health. Retrieved from FIGO: https://obgyn.onlinelibrary.wiley.com/doi/full/10.1002/ijgo.12235
3. Ayres-de-campos, D. S. (2005). Predicton of neonatal state by computer analysis of the heart rate tracing. Eur. J. Obstet, 52-60.
4. Azar, A. T., Elshazly, H. I., Hassanien, A. E., & El-Korany, A. (2014). A random forest classifier for lymph diseases. Computer Methods and Programs in Biomedicine, 113(2), 465-473. Retrieved 1 27, 2021, from https://sciencedirect.com/science/article/pii/s0169260713003751
5. Baker L, B. M. (2016). Assessing Fetal Wellbeing A Practical Guide. Southern Health and RANZCOG, 3.
6. Bank, T. W. (2020). Mortality rate, under 5. Retrieved from data.worldbank.org: https://data.worldbank.org/indicator/SH.DYN.MORT
7. Bright TJ, W. A. (2012). Effect of Clinical Decision-Support SystemsA Systematic Review. Ann Intern Med, 157(1):29–43.
8. Brownlee, J. (2016, November 16). What is a Confusion Matrix in Machine Learning. Retrieved from Machine learning mastry: https://machine learningmastery.=com/confusion-matrix-machine-learning/
9. C. Sundar, M. c. (n.d.). An overview of research challenges for classification of cardiotogram data. 9(2), 198.
10. Casanova, R., Saldana, S., Chew, E. Y., Danis, R. P., Greven, C. M., & Ambrosius, W. T. (2014). Application of Random Forests Methods to Diabetic Retinopathy Classification Analyses. PLOS ONE, 9(6). Retrieved 1 27, 2021, from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098587
11. DB, N. (2013). Using arti□cial intelligence to improve hospital inpatient care. IEEE Intell syst, 28:92–5.
12. Dexter PR, P. S. (2005). A computerized reminder system to increase the use of preventive care for hospitalized patients. N Engl J Med, 12(2):172–80.
13. Dilsizian SE, S. E. (2014). Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep, 16:441.
14. EH, S. (1987). Computer programs to support clinical decision making. JAMA, 258(1):61–6.
15. ES, B. (2014, May 4). Clinical Decision Support Systems: State of the Art. Retrieved from healthit.ahrg.gov: http://healthit.ahrq.gov/sites/default/files/docs/page/09–0069–EF_1.pdf
16. F Keikha, F. V. (2016, Nov). The effects of maternal opium abuse on fetal heart rate. 6, 41(6):479-485.
17. FA Manning, L. P. (1980, March 15). Antepartum fetal evaluation: development of a fetal biophysical profile. American Journal of Obstetrician and Gynecologists , 136(6):787-95.
18. Fetal Heart Monitoring: What's Normal, What's Not? (n.d.). (Johns Hopkins Medicine Health Library) Retrieved from healthline Parenthood: https://www.healthline.com/health/pregnancy/abnormal-fetal-heart-tracings
19. Fund, T. U. (2018). Every Child Alive: The Urgent Need to End Newborn Deaths. The United Nations Children's Fund.
20. Graber ML, F. N. (2005). Diagnostic error in internal medicine. Arch Intern Med, 165:1493–9.
21. Gulshan V, P. L. (2016). Development and Validation of a Deep Learning Algorithm for detection of Diabetic Retinopathy in retinal fundus photographs. JAMA, 316:2402–10.
22. H. Y. Chen, S. C. (2012). Electric fetal heart rate monitoring and its relationship with neonatal and infant mortality. American journal of Obstetrics and Gynecology, 204(6), 491-e1.
23. Habek D, H. J. (2002). [Intrauterine hypoxia and sudden infant death syndrome. Acta Medica Croatica, 56(3), 109-18.
24. Harinstein LM, K.-G. S. (2012). Use of an abnormal laboratory value-drug combination alert to detect drug-induced thrombocytopenia in critically Ill patients. J Crit Care, 27(3):242–9.
25. hospital, T. W. (2018). Cardiotocographic interpretations and response. Retrieved from thewomes.r.worldssl.net: https://thewomens.r.worldssl.net/images/uploads/downloadable-records/clinical-guidelines/ctg-interpretation-and-response_280720.pdf
26. J Patrick, L. C. (1984, Jan). Accelerations of the normal fetal heart rate at 38 to 40 weeks gestational age. American Journal of Obstetrician and Gynecologists , 01;148(1):35-41.
27. JA Low, B. L. (2019). Threshold of metabolic acidosis associated. Am J Obstet Gyneco, 1391.
28. Jha S, T. E. (2016). Adapting to Artificial Intelligence: radiologists and pathologists as information specialists. JAMA, 316:2353–4.
29. Jongsma, H. W. (1986). Classification of fetal and neonatal heart. Eur. Obstet. Gynecol. Reprod., 21:293–299.
30. KA Keegan, R. P. (1980, January). Antepartum fetal heart rate testing: An Out patient approach. American Journal of Obstetrician and Gynecologists , 01;136(1):81-3.
31. Keegan KA, R. P. (1980, Jan). Antepartum fetal heart rate testing: Non stress test as a promary approach. American Journal of Obstetrician and Gynecologists , 01;136(1):75-80.
32. Kucher N, K. S. (2005). Electronic alerts to prevent venous thromboembolism among hospitalized patients. N Engl J Med, 352(10):969–77.
33. Kumar, S. S., & Shaikh, T. (2017). Empirical Evaluation of the Performance of Feature Selection Approaches on Random Forest. Retrieved 1 27, 2021, from https://researchportal.hw.ac.uk/en/publications/empirical-evaluation-of-the-performance-of-feature-selection-appr
34. Lee CS, N. P. (2013). Cognitive and system factors contributing to diagnostic errors in radiology. AJR Am J Roentgenol, 201:611–7.
35. Li, H. B., Wang, W., Ding, H. W., & Dong, J. (2010). Trees Weighting Random Forest Method for Classifying High-Dimensional Noisy Data. Retrieved 1 27, 2021, from http://yadda.icm.edu.pl/yadda/element/bwmeta1.element.ieee-000005704290
36. Lip, G. Y. (2010). 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. Chest, 137(2), 263–272.

37. M Campanile, P. D. (2020, Jul 7). Intrapartum cardiotocography with and without computer analysis. Journal of maternal fetal and neonatal Medicine , 33(13):2284-2290.
38. M. Jezewski, W. M. (2007). Some practical remarks on neural network approach to fetal cardiotogram classification. 29th Annual International Conference of the IEEE. Engineering in medicine and biology socciety, EBMS.
39. machine learning. (2015). Retrieved from SAS: https://www.sas.com/en_in/insights/analytics/machine-learning.html#:~:text=Machine%20learning%20is%20a%20method,decisions%20with%20minimal%20human%20intervention.
40. Macones GA, H. G. (2008). The 2008 national institute of child health and human development workshop report on electronic fetal monitoring: Update on definitions, interpretation, and research guidelines. 37:510–5.
41. Maslova MV, M. A. (2003, July). The effects of ante- and postnatal hypoxia on the central nervous system and their correction with peptide hormones. Neuroscience and Behavioral Physiology, 6, 607-11. doi:10.1023A:1023938905744
42. MEB Menai, F. M.-m. (2013). Influence of feature selection on naive Bayes Classifier for recognizing patterns in coardiotocograms. 2(37), 1-9.
43. O'Mahony C, J. F. (2014). A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM risk-SCD). Eur Heart J, 35:2010–2020.
44. Patel VL, S. E. (2009). The coming of age of artificial intelligence in medicine. Artif Intell Med, 46:5–17.
45. Performance metrics in Classification and Regression. (n.d.). Retrieved from open genious: https://iq.opengenus.org/performance-metrics-in-classification-regression/
46. Preboth, M. (2000, September 1). ACOG guidelines on antepartum fetal surveillance. Am Fam Physicians, 5, 62(50):1184, 1184-8.
47. RM. Cruz, R. S. (2015). META-DES: A dynamic ensemble selection for meta learning. Patt. Recognition, 48(5), 1925-1935.
48. Rood E, B. R. (2005). Use of a computerized guideline for glucose regulation in the intensive care unit improved both guideline adherence and glucose regulation. Journal of the American Medical Informatics Association: JAMIA, 12(2):172–80.
49. Rudnicki, W. R., Wrzesie□, M., & Paja, W. (2015). All Relevant Feature Selection Methods and Applications. Retrieved 1 27, 2021, from https://link.springer.com/chapter/10.1007/978-3-662-45620-0_2
50. Signorini, M. M. (2003). Linear and Non Linear paramenters for the analysis of foetal heart rate signal form CTG recordings. 50(3).
51. Sim I, G. P. (2001). Clinical decision support systems for the practice of evidence-based medicine. Journal of the American Medical Informatics Association: JAMIA, 8(6):527–34.
52. Sun, J. Z. (2014). Active learning of the Gaussian process with manifold-preserving graph reduction. 25(7-8), 1615-1625.
53. Weingart SN, W. R. (2000). Epidemiology of medical error. BMJ, 320:774–7.
54. Winters B, C. J. (2012). Diagnostic errors in the intensive care unit: a systematic review of autopsy studies. BMJ Qual Saf, 21:894–902.
55. Wright A, S. D. (2011). Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. Journal of the American Medical Informatics Association: JAMIA, 18(3):232–42.
56. You D, H. L. (2015). 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, 386:2275–86.