Application of Machine Learning Techniques to classify Fetal Hypoxia

MSc Research Project
Data Analytics

Krishna Mohan Mishra
x15007901

School of Computing
National College of Ireland

Supervisor: Vikas Sahni

National College of Ireland
Project Submission Sheet – 2015/2016
School of Computing

Student Name:	Krishna Mohan Mishra
Student ID:	x15007901
Programme:	Data Analytics
Year:	2016
Module:	MSc Research Project
Lecturer:	Vikas Sahni
Submission Due Date:	21/12/2016
Project Title:	Application of Machine Learning Techniques to classify Fetal Hypoxia
Word Count:	6819

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature:	
Date:	21st December 2016

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
3. Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Contents

Application of Machine Learning Techniques to classify Fetal Hypoxia

Krishna Mohan Mishra
x15007901
MSc Research Project in Data Analytics

21st December 2016

Abstract

Aim of this research is to classify fetal hypoxia using machine learning approach based on Cardiotocography (CTG) data and patient's previous complications records. Classification method is very popular in analysing new born baby's health in critical cases. CTG data had been used by obstetricians for analysing fetal well-being during pregnancy complications which provides fetus information in detail. CTG data consist of fetal heart rate (FHR) signals from mothers abdomen in the form of continuous electronic time varying pattern. CTG data provides information about FHR which is used for prior prediction and diagnosis of long term embryo impairment. In this study high dimensional CTG data is used to classify fetal hypoxia having attributes like FHR and uterine contraction (UC) in normal and pathological terms. CTG data has been taken from UCI machine learning repository consist of 1832 instances and 21 attributes used in this study out of which 8 are continuous and 13 are discrete. Research follows stacked generalised ensemble approach consist of 6 machine learning classifiers (Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF), C4.5 Decision Tree Classier (J48), Simple Logistics (SL), Adaptive Boosting (AdaBoost)) for CTG data classification. Robustness of the model has been evaluated by 10-fold cross validation technique with accuracy, sensitivity and specificity as evaluation features for model. F-measure, ROC (receiver operation characteristics) and Kappa Statistics also used as tools to study test performance. We also compared Model Testing Time for each individual model with accuracy to find out which model performs best. Highest accuracy of 98.79 % provided by Stacking 2 (SL+C4.5(J48)+RF) model. If experienced obstetricians are not available this research will help physicians in prior detection and diagnosis of fetal hypoxia.

1 Introduction

Aim of this work is to classify fetal hypoxia using machine learning techniques based on high dimensional CTG data and patient's medical history. Research will provide solutions to so called third world or remote places where obstetricians are not accessible in terms of prior detection of fetal risks as well as its categorisation. Cesarelli et al. (2009) found that it is difficult to get direct information during pregnancy so obstetricians use FHR

as most important indirect information. This research follows machine learning classification of fetal hypoxia based on patient's medical history and present status through CTG data analysis. Feature identification is the main challenge those are responsible for fetal hypoxia in CTG data and will be separated from other common fetal risks during pregnancy. Gribbin and Thornton (2006) says that FHR must have baseline between 110-160 beats per minute with variance of 5 or more for healthy fetus.
Machine learning approach will save lot of diagnosis time for obstetricians as well as many lives of new born with early stage prevention. Diagnosis cost will also be reduced with this research approach. 'Electronics fetal monitoring' is second name of CTG however it contains other information like UC and movement of fetus inside mother's abdomen so its inappropriate. According to Grivell et al. (2012) analysis of FHR as an important biological indicator during pregnancy by antenatal CTG provides new born health assessment. Machine learning classifiers are used to classify low or high risk categories of patient in this study. Physicians at society level can provide optimisation by performing near perfect diagnosis with reduced source utilization. CTG monitoring is helpful only for the high risk cases practically. van Geijn et al. (1980) proposed that in false cases of fetal pain or fetal well-being diagnosis doesn't required.
Cross validation has been used in this research to analyse the robustness of the model and to avoid overfitting problem if tested at unknown dataset. Research is using machine learning approach which reduces cost of labour, laboratory experiments as well as save time. Predicting life sciences problem with machine learning approach is not easy task due to some vital cases. According to Sokolova et al. (2006) and Provost and Fawcett (1997) some specific features like F-measure and ROC can be added to test performance along with machine learning approach. Normal and pathological are the two main categories of CTG data as suggested by obstetricians, which have classified by machine learning approach in this research. As per Waikato (2010) accuracy, sensitivity and specificity are the other features along with F-measure and ROC to evaluate the performance of the model.
Research uses six machine learning algorithms to classify fetal hypoxia with Ensemble approach. Research provides comparative study of different machine learning algorithms and analyse the best predictor among all. Feature extraction plays significant role in this research. Cross Industry Standard Process for Data Mining (CRISP-DM) methodology has been used in this research. This research consist of the following sections: Section I provide introduction of the research. Section II reviews related works in this area and find what improvement is possible above the existing work. Section III elaborates the methodology being applied in this research. Section IV describes implementation of modelling work in details. Section V provides evaluation techniques with section VI which explores conclusion and future works.

2 Related Work

2.1 Fetal Hypoxia

Grivell et al. (2012) suggests acute and chronic are two main types of fetal hypoxia while CTG is applicable as monitoring test for fetus only after 7-month of pregnancy. Warrick et al. (2012) found that obstetricians choose caesarean section to avoid future negative outcomes after early detection of fetal hypoxia. According to Ma and Zhang (2015) maternal diseases, anaemia during pregnancy, placental insufficiency and high altitude pregnancy can be other causes of developing hypoxia. Ayres-de Campos et al. (2005) analysed that CTG information is very useful for obstetricians in identifying future complications and to avoid permanent damage of fetus it must be diagnose early. Ayres-de Campos et al. (2005) discovered that The SisPorto (Multicentre Study Group) chooses Apgar score using ROC curve to analyse CTG data consist of FHR signals by computer analysis provides results of prediction for Apgar score with less pH value required between 1 and 5 min scale which is better. Future work improves limitation of the work which is related to sensitivity in Apgar score prediction.

2.2 CTG feature extraction with Machine Learning

Huang and Hsu (2012) has quite impressive accuracy results of 82.1%, 86.36% and 97.78% respectively for machine learning techniques Artificial Neural Network (ANN), Decision Trees (DT) and Discriminant Analysis (DA) using CTG data. This study will help in efficient and effective manner to physicians for analysing status of fetal distress which is prior stage of our research aim i.e. fetal hypoxia. Ignorance of discrete attribute information and lower accuracy are main limitations of this study which can be improved by using logistic regression method in future work. In Krupa et al. (2011) SVM for FHR feature extraction with 5-fold cross validation and empirical mode decomposition (EMD) visual analysis provides 86% overall accuracy with excellent kappa values (performance measure) of 68.4 and 92.3 for test and train set respectively. This work is useful for machine learning classification of fetal status 'at risk' or 'normal' which is useful for our research. Model suitable for only large dataset which must be limitation of the study as well as EMD sifting process speed should be improved in future works.

Another study Micchelli (1986) has predicted fetal risks using CTG data with three machine learning algorithms fuzzy and multilayer perceptron with LSVM (Lagrangian SVM) based on specificity and sensitivity values. This study proves FHR as most relevant feature in fetal hypoxia prediction with UC which is important for our research. However better results than this study already published which limits the work but can be further improved. Georgoulas et al. (2006) used SVM approach for CTG data with Principal Component Analysis (PCA) for noise removal and dimension reduction provides accuracy of 75.61% which is less than previous work of same author using Hidden Markov Model (HMM) with accuracy of 83%. Proposed work has outstanding results in suspicious fetus prediction which will be hypoxic fetus in near future and quite related to our research. Another work of same author has better accuracy than this work which is lack of research.

Sundar et al. (2012b) proposed accuracy, recall and F-measure as the evaluation parameters with values 80%, 78% and 84% respectively in the research which uses ANN model for CTG data classification for pathological, suspicious and normal classes. Rand Index is also used as performance measure in this study. Proposed study found that urgent diagnosis required for pathological fetal classes which is useful for our research. Limitation of the model is in classification of suspicious data classes of CTG which can be improved by hybrid approach of both statistical and machine learning approaches in future works. Another paper Sundar et al. (2012a) used machine learning clustering method for CTG data classification with different performance measures like precision, recall and F-measure from previous study with values 80%, 35% and 45% respectively has used in this research which is not as impressive as previous study. For specific time window, just before hypoxia proposed model can analyse CTG data pattern to provide

useful information. Overfitting and lower accuracy of the model from previous study have been counted as research lack.
Georgoulas et al. (2006) proposed new method as Hidden Markov Model (HMM) for CTG classification while Georgoulas et al. (2004) provides overall accuracy of 83% for prediction of fetal well-being. In context of our research this study helps in prior detection and diagnosis of fetal hypoxia by obstetricians. Limitation of research found in pH value equal or less than normal of 7 which should be improved in future works. Ocak and Ertunc (2013) used new model Adaptive Neuro Fuzzy Inference System (ANFIS) for CTG classification in pathological and normal form with impressive accuracies of 96.6% and 97.2% respectively. Although study is based on machine learning classification with impressive results which is useful for our research but research gap found in misclassification of pathological classes by ANFIS model which mislead obstetricians towards caesarean section which is harmful for new born and mother health should be improved in future works.
Least Squares SVM (LS-SVM) method having binary decision tree algorithm is used for fetal state classification in Yilmaz and Klker (2013). Impressive accuracy of 91.62% has been achieved with Particle Swarm Optimization (PSO) as major optimization technique and 10-fold cross validation. This work got good results in machine learning classification of fetal state necessary for fetal hypoxia and beneficial to our research. However accuracy lower than other research as well as some misclassification in pathological classes those are acceptable but considered as limitation of work should be removed in future works. Radial Basis Function provides maximum accuracy of 99.8% in this particular area of study having machine learning classification of CTG data in Yarnguy and Kanarkard (2011). Study got best result till now among all researches based on machine learning classification of CTG data which is quite related to our work. Only gap in the research is found that data is unknown which reduces accuracy in suspicious classes should be removed in future works.
Salamalekis et al. (2002) provides Neural Network and k-means clustering algorithms as machine learning model for early detection of fetal hypoxia with results 83.3% and 97.9% as sensitivity and specificity respectively. Significance of the study is use of patient specific algorithms in quantitative way for fetal risks classification which may led to fetal hypoxia detection is helpful for our research. FHR interpretation reliability and efficacy of the system considered as limitation of the study and need improvement in future works. Discrete Wavelet Transformation in Georgoulas et al. (2005) used for feature extraction with SVM as machine learning classifier for real data with pH value 7 or less provides excellent accuracy of 90% in classifying normal or risk fetus. Study using real data for detection of hypoxic fetus closely related to fetal risks as well as our research. Lack in research found with value of pH which has always taken 7 or less should be improved in future works.
In Magenes et al. (2004) Intrauterine Growth Retarded (IUGR) detection in CTG data has been done by three step SVM classifier provides results in terms of sensitivity and specificity with values 78% and 79% respectively. Study using supervised machine learning techniques to detect fetal distress and sufferings those are prior stages of fetal hypoxia which is concern of our research. Research lack is that model is applicable to only small datasets which should be improved in future works. Lunghi et al. (2005) proposed Power Spectral Density (PSD) and Approximate Entropy (ApEn) with SVM base classifier as model input to Intrauterine Growth Restriction (IUGR) detection in abdomen by analysing CTG data provides accuracy of 84% with 65 attributes used for training. FHR

signals extracted features are used to detect fetal sufferance by proposed model which help obstetricians in detecting and diagnosing fetal hypoxia i.e. aim of our research. Study gap is lower accuracy which can be improved in future works by using SVM classifier as mono-pathology classifier.

FHR and Uterine Pressure (UP) signals in Warrick et al. (2006) have been used as input-output for model consist of SVM classifier with Pseudo inverse technique for noise analysis. Research classifies fetal hypoxia with ROC curve as best result provider. ROC curve analysis used to differentiate between normal and hypoxic fetus provides best results in this study related to our research. Overfitting of model and less variance values are counted as research lack and should be improved in future works. Chamidah and Wasito (2015) used hybrid K-SVM model which is ensemble of K-Means algorithm and SVM provides better accuracy of 90.64% than individual using 10-fold cross validation for robustness checking in fetal state classification with high dimensional CTG data. Proposed model provides better results in classifying fetal states which will help physicians in early detection and diagnosis of fetal hypoxia and related to our research. Limitations of study found in computational time and missing values those are common in bioinformatics research now days and should be resolved in future works.

In Comert et al. (2016) International Federation of Gynaecology and Obstetrics (FIGO) suggests some features like variability, number of acceleration and deaccleration patterns and baseline in CTG which have been classified by Extreme Learning Machine (ELM) and ANN provides impressive accuracy of 93.4% and 91.8% respectively. Proposed model based on features recommended by FIGO is used to classify fetal states for assessing fetal well-being related to fetal hypoxia and our research. ANN model hidden layers are responsible for greater time consumption which should be removed in future works. Information gain, ReliefF, Consistency and Correlation in Silwattananusarn et al. (2016) are used as SVM ensemble feature selection method which provides accuracy of 99.85% better than individual classifier for CTG classification of hypoxic fetus. Study used ensemble approach to classify fetal states and fetal sufferings those are initial stages of hypoxic fetus which is useful for our research and physicians to make better prediction in bioinformatics. Process speed after addition of more features is model limitation which should be improved in future works.

Bagging approach in Shah et al. (2015) including decision tree family algorithms like RF, Reduced Error Pruning Tree (REPTree) and C 4.5 (J48) having correlation feature selection (cfs) with 10-fold cross validation method are used for fetal state classification using CTG data based on performance measures like F-measure, Recall and Precision provides accuracy of more than 90%. Proposed model will be used as better decision support tool for fetal risks classification like fetal hypoxia to obstetricians which is related to our research. Limitation of the study is use of public database so future works should verify model authentication on primary data. Hakan and Subasi (2012) used Simple logistic and ANN algorithms for fetal state classification using SisPorto 2.0 CTG data which is divided into pathological and normal states provides outstanding accuracy of 98.7 % and 98.5% respectively. This study provides better accuracy by using machine learning methods in fetal state classification which will led to fetal hypoxia aim of our research and can replace some expensive tests by obstetricians as decision support tool for diagnosis. Data type is the limitation of this work which should be resolved in future works.

Chamidah and Wasito (2015) verified that FHR and intra UP signals are main parts of CTG data signals used for evaluation and diagnosis fetal risks till pregnancy. Comert et al. (2016) proposed CTG data available in UCI machine learning repository contains

2126 records collected during pregnancy routine check-up for assessing fetal well-being. Shah et al. (2015) discovered that for at least 15 seconds with decrement due to reduced baseline pressure of 15 bpm or more with symptoms like UC can led to fetal hypoxia. CTG data analysis and classification in Hakan and Subasi (2012) is the most popular method for early detection of pathological states like fetal hypoxia or fetal distress and congenital heart defect.

2.3 CTG and Fetal Hypoxia Relationship

According to Ayres-de Campos et al. (2005) fetal hypoxia has some main risk factors like placenta previa, abnormalities in CTG, Rh immunization, fetal growth restriction, pre-eclampsia, health condition and maternal diabetes. Sundar et al. (2012b) found that FHR signals used for monitoring amount of oxygen taken by fetus during labour. To avoid fetal loss or irreversible damage Provost and Fawcett (1997) proposed early prediction of pathological states like fetal hypoxia required which can be done by CTG data analysis because of non-availability of direct observation from fetus. Huang and Hsu (2012) found fetal distress can be considered as early stage of metabolic acidosis or fetal hypoxia. Cesarelli et al. (2009) analysed that Central Nervous System (CNS) of fetus can be damaged by fetal hypoxia which is correlated with variable short term variability (STV) extracted from CTG and related to fetal heart.

Decrease in oxygen level of fetus can led to fetal risks like fetal hypoxia. Level of oxygen is reduced due to UC which may further affect FHR which is directly related to level of oxygen. Early prediction of hypoxic fetus will help physicians in saving patient's life with machine learning approach. Fetal movement is also related to fetal hypoxia. It causes lifelong disability or even death in worst cases. Long term neurological complications, brain development and increased risk in CNS can be affected by fetal hypoxia. Diagnosis of fetal hypoxia depends upon various factors like frequency and severity of hypoxia, medical history, person's age and overall health among large patient data. Accurate diagnosis or best treatment in case of fetal hypoxia is hard to find. This research will help obstetricians and patient both to avoid unnecessary treatments which do not lead to cure.

3 Methodology

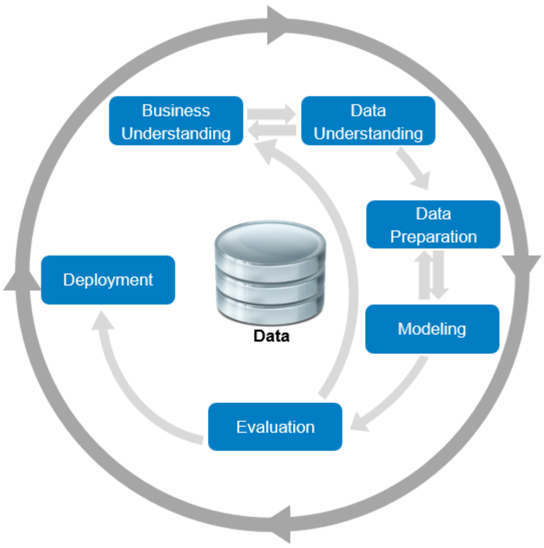

Figure 1: CRISP-DM Methodology

CRISP-DM framework is the base of methodology in this research. It provides blueprint for the complete project by dividing it into six phases which proved as better business process model and comprehensive data mining. Figure 1 shows according to CRISP-DM methodology there are six stages of process development. We translated the prediction task of business question into data mining process with methods like clustering, classification and prediction etc. Research problem is type of classification problem because based on 21 extracted features of CTG data like FHR, UC, pathological and normal states we have to predict fetal hypoxia. Next step will be the assessment of tools and techniques to be used in this research. UCI machine learning repository is the source of data. After getting data initial data preparation task was carried out. Acquiring target attribute, finding relationships and statistical analysis are the main parts of initial data pre-processing. Machine learning approach will be successful if feature selection process is better which can lead to better predictive model and improved performance on test data which is unknown.
Implementation part consist of feature engineering process in details. interesting insights from the data can be explored with appropriate bar charts, histograms and correlation between attributes. Next step of exploration consists of improving data quality with the help of imputing missing values and rectifying errors in the data. Data pre-processing provides prepared data for designing the machine learning model for major analysis. Process also includes transformation of values for existing variables. WEKA machine learning tool is already selected during business understanding phase for this research. We are using different machine learning algorithms like ANN, SVM, Simple Logistic, C 4.5 (J48), Adaboost and RF with stacked generalised ensemble approach having adaboost meta classifier.
Hakan and Subasi (2012) said that robustness of the model can be tested by 10-fold cross validation. Model Testing time is also taken into consideration. In pre-processing stage, we identified the attribute which has significant importance in the model. Mic-

chelli (1986) found various performance measures are considered for model evaluation which have made the model more accurate. Accuracy, Sensitivity (Recall), Specificity, F-measure, ROC Curve area, Kappa Score and Testing time are taken into consideration for prediction performance matrix in context of model evaluation. Final report consists of all data mining results with future scope of improvement at the end in the form of comprehensive presentation. All models have evaluated and compared to find the best performing model for aim of understanding. Different business applications favours different models depends upon which aspect of performance matrix have been focused in each business scenario. Accuracy of the model is more focused in this study for the scope of this project.

4 Implementation

A. Data Pre-processing and Splitting

Original data set has 2127 instances with 40 attributes like file name and date. Model performance can be decreased by nominal categorical variables which is common challenge. Normal state represented by 1 while pathological represents 0. Original data collected from UCI machine learning repository presented below in Excel format as in Bache and Lichman (2013):

1	FileName	Date	SegFile	b	e	LBE	LB	AC	FM	UC	ASTV	MSTV	ALTV	MLTV	DL	DS	DP	DR	Width	Min	Max	Nmax	Nzeros	Mode	Mean	Median	Variance
2																											
3	Variab10.txt	01-12-1996	CTG0001.txt	240	357	120	120	0	0	0	73	0.5	43	2.4	0	0	0	0	64	62	126	2	0	120	137	121	73
4	Fmcs_1.txt	03-05-1996	CTG0002.txt	5	632	132	132	4	0	4	17	2.1	0	10.4	2	0	0	0	130	68	198	6	1	141	136	140	12
5	Fmcs_1.txt	03-05-1996	CTG0003.txt	177	779	133	133	2	0	5	16	2.1	0	13.4	2	0	0	0	130	68	198	5	1	141	135	138	13
6	Fmcs_1.txt	03-05-1996	CTG0004.txt	411	1192	134	134	2	0	6	16	2.4	0	23	2	0	0	0	117	53	170	11	0	137	134	137	13
7	Fmcs_1.txt	03-05-1996	CTG0005.txt	533	1147	132	132	4	0	5	16	2.4	0	19.9	0	0	0	0	117	53	170	9	0	137	136	138	11
8	Fmcs_2.txt	03-05-1996	CTG0006.txt	0	953	134	134	1	0	10	26	5.9	0	0	9	0	2	0	150	50	200	5	3	76	107	107	170
9	Fmcs_2.txt	03-05-1996	CTG0007.txt	240	953	134	134	1	0	9	29	6.3	0	0	6	0	2	0	150	50	200	6	3	71	107	106	215
10	Hasc_1.txt	22-02-1995	CTG0008.txt	62	679	122	122	0	0	0	83	0.5	6	15.6	0	0	0	0	68	62	130	0	0	122	122	123	3

Figure 2: UCI Cardiotocography Data

Raw data has been collected and domain knowledge has gathered by literature review as well as identification of relevant features for research. Feature engineering is required to reduce complexity of data as well as to improve the model performance. All categorical variables are converted into numerical variables because it only accepts 0 or 1. Three experienced obstetricians generated UCI machine learning repository CTG data to classify the embryo into normal or pathological states. UCI CTG dataset has been generated by SISPORTO 2.0 software. Normal, suspicious and pathological are the three main classes of CTG data. We are using only 21 features out of 40 from UCI CTG data out of which 13 are discrete and 8 are continuous.

We are taking only two states of fetus normal or pathological into consideration under this research as suggested by literatures. We have proposed stacked generalised ensemble approach of ANN, SVM and RF. Individual performance of algorithms and model robustness can be analysed by 10-fold cross validation. Model performance and accuracy can be improved by training several times the model. Feature selection has done before because research require only selected features. Features like LB (FHR baseline beats per minute), FM (fetal movements per second), UC (uterine contraction per second), STV (short-term variability) and NSP (fetal state class) are suitable for our research. Correlation between attributes and time dependent changes analysed by features selection method within CTG data.

Symbol	Attribute information
LB	FHR baseline (beats per minute)
AC	# of accelerations per second
FM	# of fetal movements per second
UC	# of uterine contractions per second
DL	# of light decelerations per second
DS	# of severe decelerations per second
DP	# of prolonged decelerations per second
ASTV	Percentage of time with abnormal short-term variability
MSTV	Mean value of short-term variability
ALTV	Percentage of time with abnormal long-term variability
MLTV	Mean value of long-term variability
Width	Width of FHR histogram
Min	Minimum of FHR histogram
Max	Maximum of FHR histogram
Nmax	# of histogram peaks
Nzeros	# of histogram zeros
Mode	Histogram mode
Mean	Histogram mean
Median	Histogram median
Variance	Histogram variance
Tendency	Histogram tendency
NSP	Fetal state class (code (N = normal; P = pathological))

Figure 3: UCI Cardiotocography Data Attribute Description

In this experiment, we have 2127 instances in UCI CTG dataset out of which 295 belong to suspicious class. Since aim of the research is to help obstetricians during early prediction and diagnosis of fetal hypoxia these 295 instances are not providing any useful information in this study so we have excluded them by "RemoveWithValue" filter in WEKA. But before that we have converted UCI CTG dataset from Excel to CSV and then to Attribute-Relation File Format (ARFF) because WEKA only performs best on ARFF format of data. After that we resolved the issue of missing values in data with "ReplaceMissingValue" filter. "NumericToNominal" filter has been used because WEKA only predict nominal attribute and we converted predicted attribute to binary i.e. 1 for normal and 0 for pathological with "NominalToBinary" filter. After all the feature engineering, we have found our dataset in ARFF format as below:-

No.	LB Numeric	AC Numeric	FM Numeric	UC Numeric	ASTV Numeric	MSTV Numeric	ALTV Numeric	MLTV Numeric	DL Numeric	DS Numeric	DP Numeric	Width Numeric	Min Numeric	Max Numeric	Nmax Numeric	Nzeros Numeric	Mode Numeric	Mean Numeric	Median Numeric	Variance Numeric	Tendency Numeric	**NSP** Nominal
1	131.9...	3.125	7.262...	3.912...	44.60...	1.44405	6.775...	8.21048	1.757...	0.003...	0.135...	73.87...	90.38...	164.2...	4.188...	0.336...	135.9...	132.9...	136.6...	20.704...	0.302402	1
2	132.0	4.0	0.0	4.0	17.0	2.1	0.0	10.4	2.0	0.0	0.0	130.0	68.0	198.0	6.0	1.0	141.0	136.0	140.0	12.0	0.0	1
3	133.0	2.0	0.0	5.0	16.0	2.1	0.0	13.4	2.0	0.0	0.0	130.0	68.0	198.0	5.0	1.0	141.0	135.0	138.0	13.0	0.0	1
4	134.0	2.0	0.0	6.0	16.0	2.4	0.0	23.0	2.0	0.0	0.0	117.0	53.0	170.0	11.0	0.0	137.0	134.0	137.0	13.0	1.0	1
5	132.0	4.0	0.0	5.0	16.0	2.4	0.0	19.9	0.0	0.0	0.0	117.0	53.0	170.0	9.0	0.0	137.0	136.0	138.0	11.0	1.0	1
6	134.0	1.0	0.0	10.0	26.0	5.9	0.0	0.0	9.0	0.0	2.0	150.0	50.0	200.0	5.0	3.0	76.0	107.0	107.0	170.0	0.0	0
7	134.0	1.0	0.0	9.0	29.0	6.3	0.0	0.0	6.0	0.0	2.0	150.0	50.0	200.0	6.0	3.0	71.0	107.0	106.0	215.0	0.0	0
8	122.0	0.0	0.0	0.0	83.0	0.5	6.0	15.6	0.0	0.0	0.0	68.0	62.0	130.0	0.0	0.0	122.0	122.0	123.0	3.0	1.0	0
9	122.0	0.0	0.0	1.0	84.0	0.5	5.0	13.6	0.0	0.0	0.0	68.0	62.0	130.0	0.0	0.0	122.0	122.0	123.0	3.0	1.0	0
10	122.0	0.0	0.0	3.0	86.0	0.3	6.0	10.6	0.0	0.0	0.0	68.0	62.0	130.0	1.0	0.0	122.0	122.0	123.0	1.0	1.0	0

Figure 4: Feature Engineering of Attributes

K-fold cross validation is used in this study that divides data into K subsets which repeat the method K times. Each time K-1 parts of the data have been used to build the model as training data and rest one of the k subset used for testing purpose. In WEKA, every time 90% of the data is used for training and rest 10% has been used for testing the model. Overall accuracy is calculated by averaging the all K trials. How data is divided it matters least that is advantage of this method. Each data point is present once in test set and K-1 times in train set. Computation time is less due to small dataset. Model overfitting problem can be overcome with k-fold cross validation. UCI CTG dataset has been broken into train and test set for performance calculation of each method. This process is called K-fold cross validation which avoids picking a part of data for training and testing.

In this research CTG dataset splits into 10 parts due to K which is equal to 10. 9 data elements use to train the model while rest for testing. Process has repeated 10 times until each fraction of data being used as test dataset. In this research, like Bakshi (1998) features are same for all the classifiers while classifier parameters chosen according to K-fold cross validation during training phase.

B. Machine Learning Models

WEKA workbench allows us to change and find appropriate parameters for classification. We are using default values of parameters for all the algorithms and with the help of confusion matrices, testing has been done to achieve best accuracy by respective classifiers.

1) Artificial Neural Networks (ANNs): Biological neural networks are inspiration for computing structures like ANN. Learning capability of ANN is due to weight adjustment of interconnection through input data. Neurons construct ANN which have one or more weighted input and output when connecting other neurons individually. Feed forward multilayer networks with non-linear node functions can overcome perceptron limitations. Between transportation of single layer to multiple layer of perceptrons conversion of simple perceptron learning algorithm cannot possible. "Multilayer perceptrons (MLPs)" are these types of feed forward networks which has been used in this research through WEKA. When back propagation learning method is training feed forward neural network then expression called as "Back propagation network". ANN has seven parameters but we are testing only three like hidden layers equal to "a" with learning rate 0.21 and momentum 0.20 default settings.

2) Support Vector Machines (SVM): Binomial classification is main application of SVMs which is now one of the popular machine learning algorithm. SVMs are most widely used machine learning algorithm in biomedical sciences like protein compounds classifications as well as computer science problems like hypertext, text, images and handwritings classification. Multidimensional data can be splited by SVM into two classes with

hyperplane. Real life problems usually generate nonlinear data which is not easy to separate into two groups with a line. SVM has concept of "kernel-induced feature space" which coverts data into separable by transferring it into higher dimensional space for accomplishing nonlinearity problem. It is interesting to know how time-consuming complexity of transformation and model overfitting problems have been handled by SVM. According to Boswell (2002) SVM uses dot product during computation in higher dimensional space which provides immunity towards these problems of computation and overfitting. SVM can be used for regression problems besides classification. In this research SVM has been used with two parameters i.e. kernel as PolyKernel and cost value as 8500.

3) Random Forest (RF): It is one of most popular machine learning algorithm in terms of classification accuracy. RF is a reliable predictor due to its way of creating group and arbitrary instantiation besides better generalizations. In Yang et al. (2010) bagging sketch is used to derive generalization property, which diminish variance to improve generalisation which can also be performed using boosting dependent process by bias diminishing. In this research, we tested two out of four parameters for RF such as number of trees as 10 and seed as 1. RF has three features those are primary:
a) Different field of techniques can be reliably classified with RF.
b) In model training significance of each attribute can be evaluated by RF.
c) Among samples pairwise proximity can be measured by trained model.

4) C4.5 Decision Tree Classifier (J48): In each dataset type of classes must be known to experimental learning system. Correspondence between classes and attribute values is responsibility of mapping function and vector of attribute value which exists in each system. Continuous and discrete attributes having numeric and nominal values respectively can be assembled into attributes applied for defining examples. Quinlan (1996) said that decision tree classifier can be learn by a system called C 4.5 (J48). Number of folds as 3, seed as 1 and confidence factor as 0.25 are three parameters have tested out of four for C 4.5 (J48) in this research.

5) Simple Logistics (SL): Binary responses have been provided by medical datasets frequently not numerical. Binary logistic regression model method suits this type of response to find correlations between risk factors and disease prediction. Accoding to Hakan and Subasi (2012) this regression model suits to dichotomous variable response with any kind of risk factor of illness. logistic regression model needs either normally distributed variables or linearly correlated risk factors and variables response. In WEKA, there are four parameters available for simple logistics but we testing only two in this research like heuristics stop as 50 and max boosting iterations as 500.

6) Adaptive Boosting (AdaBoost): In this research, we are using Ensemble machine learning approach with AdaBoost like Freund and Schapire (1996) which is most popular variant for boosting now days. In the training sample, each candidate has been assigned a weight by AdaBoost differently. Greater weight has been assigned to training sample candidate which is incorrectly classified by previous base classifier. Next base classifier has been assigned a training set which consists of selected candidates according to their weights. Therefore, samples whose correct classification is difficult are matter of concern for AdaBoost. Until error ratio has not reached lower value process of adding base classifier continues. Weighted votes are factor for AdaBoost to decide strategy unlike majority vote for boosting algorithms. Classifier's training accuracy decides weights of votes.

7) Ensemble Approach (Stacking): AdaBoost has been used as meta classifier for Stacking ensemble approach in this research to improve the accuracy of classification of

CTG data for early and better prediction of fetal hypoxia. Karabulut and Ibrikci (2014) says that ensemble approach consists of more than one machine learning classifiers those are trained simultaneously to contribute in combined manner to final decision of system. These classifiers contributing to final decision are called as base classifiers produced one after another by boosting process. Each base classifier is using incorrectly classified instances in training set by previous base classifier due to its dependency. Therefore, previous errors can be fixed by new base classifier for strengthening ensemble.
Data normalization required before neural network training because sometimes before number of maximum iterations allowed algorithms have not converged which depends on dataset. According to Hall et al. (2009) WEKA data mining tool is used to compare classifier's performances with and without AdaBoost ensemble which has machine learning algorithms collection in JAVA programming language. Each classification algorithms are trained with default parameters provided by WEKA. Classifier performance validation has done by 10-fold cross validation which divides dataset into 10 subsets and 1 subset is used as test while rest as train set in a hold out operation which is repeated 10 times for each classifier. Therefore, averaging 10 accumulated accuracies provides eventual accuracy of classification for a particular model. Stability can be provided by 10-fold cross validation method.

5 Evaluation

5.1 Performance Evaluation

It is difficult to predict performance for inadequate data using machine learning approach. Therefore, small amount of data can be analysed by cross-validation which is favourite of researchers. How to divide data into train and test must be made before applying machine learning approach. Entire CTG is splitted into train and test set for performance evaluation of machine learning method and afterwards famous method of evaluation 10-fold cross validation is applied. Model is formed using train set while verified using test set in classification process. Efficiency of the classifier can be calculated by number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Checkout tests having statistical measurements like sensitivity and specificity. Sensitivity provides positive test results.
Negative test results ratio indicates specificity while overall measure shows by accuracy. Class distribution or error costs have no effect on classifier performance represented by ROC. Salzberg (1997) says Curve between all specific values and correspondent sensitivity values called as ROC curve . Amount of threshold tested is responsible for approximation excellence. Classification of cases cannot use ROC plot as standard method regardless of its good sides. ROC drawing can provide decision formulas by applying some different methods. F-measure as another statistical measurement is used to evaluate performance characterization.

5.2 Accuracy, Specificity and Sensitivity

Salzberg (1997) proposes division of CTG data into train and test set for calculation of performance of each machine learning method using K-fold cross validation. Figure 5 shows the comparison of the performance testing of algorithms. As it displayed in figure 5, Stacking 2 (SL+C4.5(J48)+RF) provides maximum accuracy, specificity and sensitivity as compared to others. This result shows that stacked generalised ensemble approach always performs better than individual classifiers. For each algorithm, respected F-measure, ROC, kappa value and model testing time also computed at the end of the process. K-fold cross validation is being applied with value K=10 for each classifier to assure the performance. Table 1 provides comparison chart for performance of each machine learning algorithm in terms of percentage value for accuracy, sensitivity and specificity. Highlighted Stacking 2 provides best values among all and proven as best predictor in classifying CTG data with values 98.79% as accuracy, 94.33% as specificity and 98.78% as sensitivity.

Table 1: Machine Learning Classifiers Performance Comparison (%)

	ANN	SVM	RF	Stacking 1	AdaBoost	SL	C4.5 (J48)	Stacking 2
Accuracy	98.51	97.92	98.58	98.13	97.72	98.75	98.64	98.79
Specificity	89.78	83.02	93.24	91.45	82.42	91.45	91.53	94.33
Sensitivity	98.52	97.94	98.63	98.11	97.67	98.74	98.56	98.78

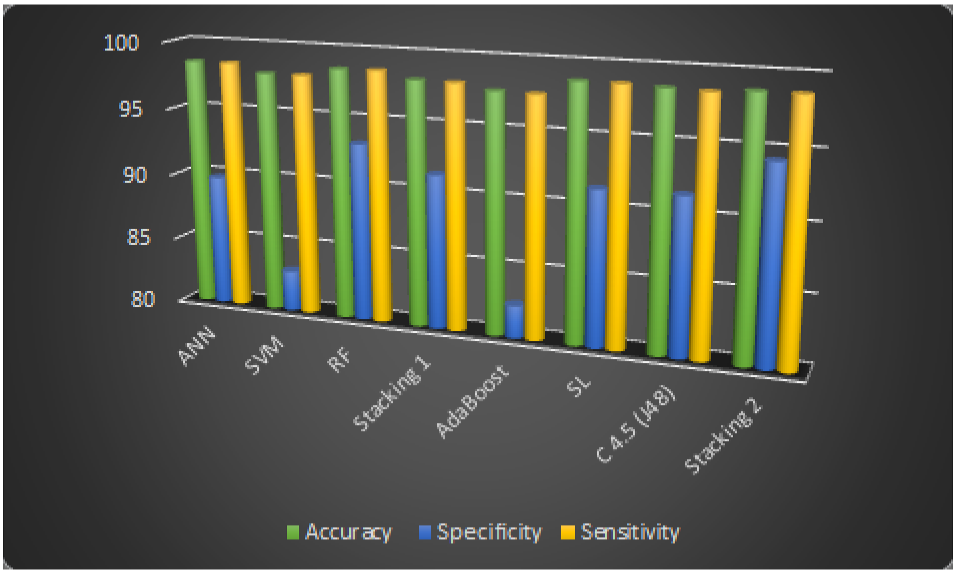

Figure 5: Performances of the respective machine learning methods

5.3 F-Measure, ROC Area and Kappa Statistics

Figure 6 shows comparison among respective machine learning algorithms with their F-Measure, ROC Area and Kappa Statistics values out of which we can see that Stacking 2 provides best values in all three parameters which proves it as best classifier among all for CTG data in this research. Table 2 provides performance comparison for each machine learning algorithm in terms of F- Measure, ROC Area and Kappa Statistics values. Stacking 2 method provides maximum values of 0.988, 0997 and 0.931 for all F-Measure, ROC Area and Kappa Statistics respectively.

Table 2: Performance Comparison with F- Measure, ROC Area and Kappa Statistics

	ANN	SVM	RF	Stacking 1	AdaBoost	SL	C4.5 (J48)	Stacking 2
F- Measure	0.985	0.978	0.986	0.982	0.976	0.987	0.986	0.988
ROC Area	0.995	0.912	0.995	0.989	0.986	0.996	0.928	0.997
Kappa Statistics	0.911	0.871	0.919	0.894	0.861	0.926	0.921	0.931

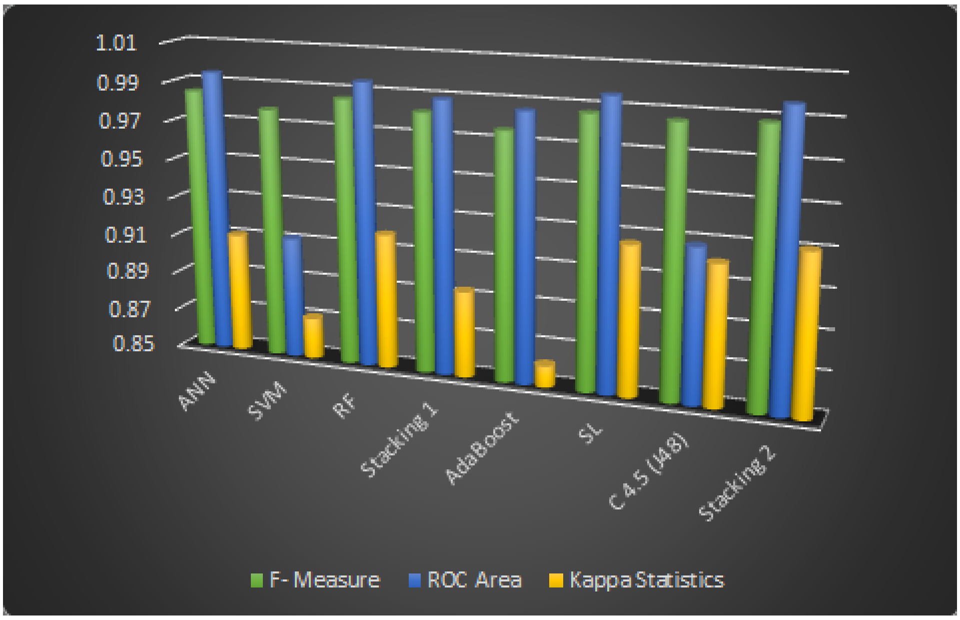

Figure 6: F- Measure, ROC Area and Kappa Statistics of the respective machine learning methods

5.4 Model testing time

Model testing time for 30% of the data which is test set has been compared below in table 3 for each machine learning model we have used in this research. SVM is fastest as we can see, only 0.02 seconds required to verify the model while if we want moderate model with high accuracy and low testing time then we can select C 4.5 (J48) which takes 0.05 seconds to verify the model with an accuracy of 98.64% which is better.

Table 3: Model Testing Time Performance Comparison

	ANN	SVM	RF	Stacking 1	AdaBoost	SL	C4.5 (J48)	Stacking 2
Testing Time	4.21 s	0.02 s	0.05 s	47.02 s	0.07 s	0.94 s	0.05 s	8.13 s
Accuracy	98.51	97.92	98.58	98.13	97.72	98.75	98.64	98.79

5.5 Other Algorithms and Approaches

We have tried other algorithms and approach with different combinations to improve the accuracy of the classification of UCI CTG dataset but we have found above listed results as better than others while as we have tried so its better to include them in our research results. Results we summarized in Table 2 and 3 are better than other classifiers used in this research. Accuracies of classifiers are in the range (97.72-98.79) which includes ANN, SVM, RF, Stacking 1 (ST) (ANN+SVM+RF), AdaBoost (AB), SL, C 4.5 and Stacking 2 with accuracies 98.51, 97.92, 98.58, 98.13,97.72, 98.75, 98.64 and 98.79 respectively. Area under curves have values 0.995,0.912,0.995, 0.989,0.986,0.996, 0.928 and 0.997 for algorithms ANN, SVM, RF, Stacking 1, AdaBoost, SL, C 4.5 and Stacking 2 respectively. F-Measure has following results 0.985 for ANN,0.978 for SVM,0.986 for RF, 0.982 for Stacking 1, 0.976 for AdaBoost, 0.987 for SL, 0.986 for C 4.5 and 0.988 for Stacking 2. This research will help obstetricians to detect and diagnose fetal hypoxia with the help of CTG data classification as well as proves the supremacy of stacked generalised ensemble approach in handling CTG data over conventional machine learning methods.

Table 4: Performance Comparison of Other Algorithms

	AB-RF	AB-ANN	AB-SVM	ST-AB-SVM+RF	ST-AB-RF
Accuracy	98.53	98.47	98.37	98.54	98.53
Specificity	91.50	89.78	89.78	93.23	94.34
Sensitivity	98.45	98.45	98.43	98.51	98.53
F- Measure	0.985	0.985	0.983	0.985	0.985
ROC Area	0.995	0.974	0.994	0.983	0.984
Kappa Statistics	0.915	0.910	0.904	0.913	0.914
Testing Time	0.07 s	12.37 s	0.22 s	1 s	0.55 s

5.6 ROC Curve Analysis

Stacking 2 has the highest value of 0.997 for ROC curve which is close to 1. It means classification of CTG data can be done with high precision by Stacking 2 model. Figure 6 shows the ROC curve for Stacking 2 model. In biomedical datasets specificity is much important than sensitivity in terms of analysis. For obtaining high specificity algorithm should show its power in classifying pathological states. Classifiers are showing high specificity it means all models are showing same performance in classifying CTG dataset. Different face of data can be seen by analysis of algorithm sensitivities. In our research sensitivities have higher values for each classifier it means they are better in classifying positively labelled data. F- measure is another important evaluation parameter which shows similarity between two classification results.

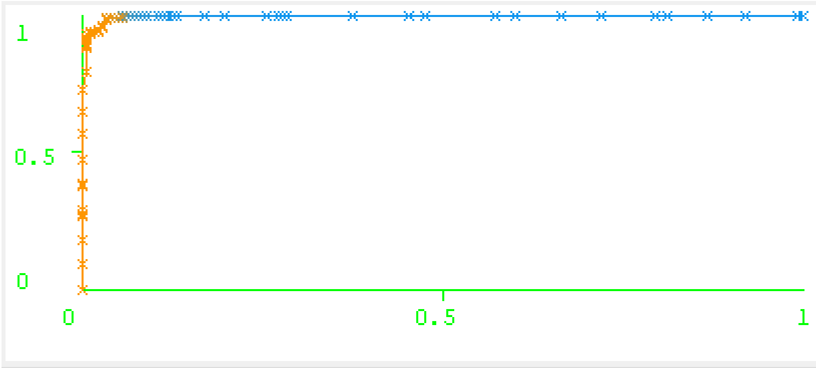

Figure 7: ROC Curve for Stacking 2 Algorithm

5.7 Recommendation of Viva Panel

Viva panel suggested me to analyse suspicious class of CTG data with machine learning model presented in this research. After evaluation, I found maximum accuracy of 93.98 % for Random Forest in CTG data classification with suspicious class which is less than result of my research with Stacking 2 ensemble approach.

6 Conclusion and Future Work

UCI CTG dataset has been used in this research to evaluate the performance of machine learning classifiers mainly ANN, SVM, RF, SL, C 4.5 (J48) and stacked generalised ensemble approach with AdaBoost as meta classifier. 10-fold cross validation has been used to verify the robustness of the model. Stacking 2 (SL+ C 4.5(J48) +RF) performs best among all machine learning classifiers used in this research for classification of UCI CTG dataset into normal and pathological states. Future work consists of testing of model over larger dataset. However biomedical dataset which is reliable and well maintained is hard to find due to confidentiality of patient information unless major hospital or health board itself initiated research for its own accord. Fetal abnormalities can be diagnosed by obstetricians with CTG data and before persistent damage of fetus medical intervention can be applied.
Model performance on large dataset is also be our concern in future work. Future study also includes improvement of model accuracy further. Discussion provides two main tasks for our future research. First, Other machine learning models we are also going to apply those are not included in this study as level-0 models as well as different combination of algorithms in ensemble approach are also our concern to improve accuracy of classification. Second, for level -1 model better prediction can be achieved by more complex algorithms to solve the non-linearity problem in level -0 prediction. Last but not the least our Stacking 2 model is going to be tested with different types of biomedical datasets like cancer, breast cancer etc. to verify the model performance in our future works.

Acknowledgements

I am very grateful to my thesis supervisor Mr. Vikas Sahni for his insights and able direction to this research thesis which is individual piece of work. For his encouragement, advice, knowledge and continuous support throughout the entire process, i wish to thank him. I also wish to acknowledge UCI machine learning repository for publicly available CTG data and SisPorto project for generating data for research use.

References

Ayres-de Campos, D., Costa-Santos, C., Bernardes, J. and Group, S. M. V. S. (2005). Prediction of neonatal state by computer analysis of fetal heart rate tracings: the antepartum arm of the sisporto multicentre validation study, *European Journal of Obstetrics Gynecology and Reproductive Biology* **118**(1): 52–60.

Bache, K. and Lichman, M. (2013). Uci machine learning repository.

Bakshi, B. (1998). Multiscale pca with application to multivariate statistical process monitoring, *AIChE journal* **44**(7): 1596–1610.

Boswell, D. (2002). Introduction to support vector machines:, *University of carlifornia. San Diego.* .

Cesarelli, M., Romano, M. and Bifulco, P. (2009). Comparison of short term variability indexes in cardiotocographic foetal monitoring, *Journal of Computers in biology and medicine* **39**(2): 106–118.

Chamidah, N. and Wasito, I. (2015). Fetal state classification from cardiotocography based on feature extraction using hybrid k-means and support vector machine, *Proceedings of 2015 IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS)* **33**(2): 37–41.

Comert, Z., Kocamaz, A. and Gngr, S. (2016). Cardiotocography signals with artificial neural network and extreme learning machine, *Proceedings of 24th IEEE Signal Processing and Communication Application Conference (SIU)* **1542**(9): 1493–1496.

Freund, Y. and Schapire, R. (1996). Experiments with a new boosting algorithm, *In Icml* **96**: 148–156.

Georgoulas, G., Stylios, C. and Groumpos, P. (2005). Classification of fetal heart rate using scale dependent features and support vector machines, *Proceedings of 16th IFAC World Cong.* **38**(1): 313–318.

Georgoulas, G., Stylios, C. and Groumpos, P. (2006). Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines, *IEEE Transactions on Biomedical Engineering BME* **5**(53): 875–884.

Georgoulas, G., Stylios, C., Nokas, G. and Groumpos, P. (2004). Classification of fetal heart rate during labour using hidden markov models. in neural networks, *Proceedings of IEEE International Joint Conference* **3**(2473): 2471–2475.

Gribbin, C. and Thornton, J. (2006). Critical evaluation of fetal assessment methods. high-risk pregnancy management options, **3**(2): 229–239.

Grivell, R., Alfirevic, Z., Gyte, G. and Devane, D. (2012). Antenatal cardiotocography for fetal assessment, *Journal of Cochrane Database Syst Rev* **1**(2): 12.

Hakan, S. and Subasi, A. (2012). Classification of fetal state from the cardiotocogram recordings using ann and simple logistic, *Proceedings of CEUR Workshop* **1542**(9): 33–36.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. (2009). The weka data mining software: an update, *Journal of ACM SIGKDD explorations newsletter* **11**(1): 10–18.

Huang, M. and Hsu, Y. (2012). Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network, *Journal of Biomedical Science and Engineering* **5**(9): 526–533.

Karabulut, E. and Ibrikci, T. (2014). Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach, *Journal of Computer and Communications* **09**(2): 32.

Krupa, N., Mohd, A., Zahedi, E., Ahmed, S. and Hassan, F. (2011). Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine, *Journal of Biomedical Engineering Online* **10**(1): 6.

Lunghi, F., Magenes, G., Pedrinazzi, L. and Signorini, M. (2005). Detection of fetal distress though a support vector machine based on fetal heart rate parameters, *IEEE Transactions In Computers in Cardiology* **32**(1): 247–250.

Ma, Q. and Zhang, L. (2015). Epigenetic programming of hypoxicischemic encephalopathy in response to fetal hypoxia, *Journal of Progress in neurobiology* **124**: 28–48.

Magenes, G., Pedrinazzi, L. and Signorini, M. (2004). Identification of fetal sufferance antepartum through a multiparametric analysis and a support vector machine, *Proceedings of 26th Annual International Conference of the IEEE, In Engineering in Medicine and Biology Society, 2004. IEMBS'04* **1**(1): 462–465.

Micchelli, C. (1986). Interpolation of scattered data: distance matrices and conditionally positive definite functions, *Journal of Constr. Approxim* pp. 11–22.

Ocak, H. and Ertunc, H. (2013). Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems, *Journal of Neural Computing and Applications* **23**(6): 1583–1589.

Provost, F. and Fawcett, T. (1997). Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions, *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining* **2049**: 43–48.

Quinlan, J. (1996). Improved use of continuous attributes in c4. 5, *Journal of artificial intelligence research* **4**(4): 77–90.

Salamalekis, E., Thomopoulos, P., Giannaris, D., Salloum, I., Vasios, G., Prentza, A. and Koutsouris, D. (2002). Computerised intrapartum diagnosis of fetal hypoxia based on fetal heart rate monitoring and fetal pulse oximetry recordings utilising wavelet analysis and neural networks, *BJOG: An International Journal of Obstetrics Gynaecology* **109**(10): 1137–1142.

Salzberg, S. (1997). On comparing classifiers: Pitfalls to avoid and a recommended approach, *Journal of Data mining and knowledge discovery* **3**(1): 317–328.

Shah, S., Aziz, W., Arif, M. and Nadeem, M. (2015). Decision trees based classification of cardiotocograms using baggingapproach, *Proceedings of 13th IEEE International Conference on Frontiers of Information Technology (FIT)* **1542**(9): 12–17.

Silwattananusarn, T., Kanarkard, W. and Tuamsuk, K. (2016). Enhanced classification accuracy for cardiotocogram data with ensemble feature selection and classifier ensemble, *Journal of Computer and Communications* **4**(04): 20.

Sokolova, M., Japkowicz, N. and Szpakowicz, S. (2006). Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, *Australasian Joint Conference on Artificial Intelligence,Springer Berlin Heidelberg* **1**: 1015–1021.

Sundar, C., Chitradevi, M. and Geetharamani, G. (2012a). An analysis on the performance of k-means clustering algorithm for cardiotocogram data clustering, *International Journal of Computer Applications* **47**(2): 11–20.

Sundar, C., Chitradevi, M. and Geetharamani, G. (2012b). Classification of cardiotocogram data using neural network based machine learning technique, *International Journal of Computer Applications* **47**(14): 19–25.

van Geijn, H., Jongsma, H., de Haan, J. and Eskes, T. (1980). Analysis of heart rate and beat-to-beat variability: Interval difference index, *American journal of obstetrics and gynecology* **138**(3): 246–252.

Waikato, M. (2010). The weka classification algorithms.

Warrick, P., Hamilton, E., Kearney, R. and Precup, D. (2012). A machine learning approach to the detection of fetal hypoxia during labor and delivery, *AI Magazine* **33**(2): 79.

Warrick, P., Kearney, R., Precup, D. and Hamilton, E. (2006). System-identification noise suppression for intra-partum cardiotocography to discriminate normal and hypoxic fetuses, *IEEE Transactions In Computers in Cardiology* **33**(2): 937–940.

Yang, P., Y., H. Y., B Zhou, B. and Y Zomaya, A. (2010). A review of ensemble methods in bioinformatics, *Current Bioinformatics* **5**(4): 296–308.

Yarnguy, T. and Kanarkard, W. (2011). A radial basis function committee machine for cardiotocography classification, *Proceedings of 12th Graduate Research Conference* **10**(6): 262–267.

Yilmaz, E. and Klker, . (2013). Determination of fetal state from cardiotocogram using ls-svm with particle swarm optimization and binary decision tree, *Journal of Computational and mathematical methods in medicine* **15**(8): 1583–1589.