National College of Ireland

Abnormal Foetuses Classification Based on Cardiotocography Recordings Using Machine Learning and Deep Learning Algorithms

MSc Research Project
MSc in Data Analytics

Jassem Alhaj Tamer
Student ID: X15021301

School of Computing
National College of Ireland

Supervisor: Dr Catherine Mulwa

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Jassem Alhaj Tamer
Student ID:	X15021301
Programme:	MSc in Data Analytics **Year:** 2020
Module:	Research Project
Supervisor:	Dr Catherine Mulwa
Submission Due Date:	17/08/2020
Project Title:	Abnormal Foetuses Classification Based on Cardiotocography Recordings Using Machine Learning and Deep Leaning Algorithms
Word Count:	**8611** **Page Count 25**

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 required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: 

Date: 17/08/2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

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):	

Abnormal Foetuses Classification Based on Cardiotocography Recordings Using Machine Learning and Deep Learning Algorithms

Jassem Alhaj Tamer
X15021301

Abstract

Cardiotocography (CTG), known as Electronic Foetal Monitoring, is typically performed at third trimester during pregnancy and in labour to primarily monitor the relationship between foetal heart rate (FHR) and contractions of the uterus (UC). CTG outcomes allow experts to determine the health of the foetus and to detect any foetal impairments. This research project examined the CTG recordings overlaid with various well-established machine learning and deep learning algorithms with the objective of identifying the best performing algorithm capable of classifying abnormal (suspect or pathologic) foetuses. Accuracy, recall and specificity are considered as significant performance metrics to identify the best performing model. Machine learning (support vector machine (SVM), C5.0 decision tree (C5.0), random forest (RF), generalised linear model (GLM), extreme gradient boosting (XGBoost), k-nearest neighbour (KNN) and naïve Bayes (NB)) and deep leaning models (multilayer perceptron neural networks (MLPNNs)) were applied. SVM model showed promising performance measures towards the classification of abnormal foetuses based on CTG recordings at accuracy of 90.65%, recall of 96.32% and specificity of 89.09%.

1 Introduction

CTG, which records (-graph) the foetal heart rate (cardio-) and the uterine contractions (-toco), is used as the ultrasound waves-based diagnostic monitor to evaluate the well-being of foetuses and to depict foetal distress such as foetal hypoxia or foetal tachycardia (Almström et al., 1992). CTG measures foetal heart rate (FHR) and the pressure inside the uterus antenatally and the intrapartum (Spencer, 1993). Foetal heart rates that are not within a baseline of 110-160 Beats Per Minute (BPM) can lead to invasive investigations or, in critical cases, to an emergency caesarean section or instrumental vaginal birth as babies are short of oxygen (Grivell et al., 2015). Foetuses with foetal heart rate variability greater than 6 BPM and accelerations with foetal movement (FM) developed a foetal well-being, whereas foetuses who do not meet this criterion died in the perinatal period (Parer and King, 2000). Clinicians' misinterpretations of CTG recordings led to an increase of obstetric litigations (Williams and Arulkumaran, 2004).

1.1 Motivation and Background

The well-being of a foetus has traditionally relied on tried and tested Obstetric techniques which focused principally on foetal heartbeat and uterine contractions. However, with advances in technology and the development of machine learning and deep learning it is interesting to examine how, and if, these scientific advances which rely heavily on data, can be used in

combination with the more traditional techniques to determine if greater efficiencies can be achieved in terms of cost, time and outcomes.

Clinical Data- Mining (CDM), which is an examination of clinical datasets to generate information for clinical decision-making, helps experts in diagnosis, prediction and treatment. Moreover, CDM is able to find hidden trends and patterns not readily observable by clinicians (Gracia Jacob and Geetha Ramani, 2012). Since 1970 researchers applied different data mining methods helping related-field staff to accurately interpret CTG trace patterns rather than relying on signal processing tools or computer-based techniques. The predictive capacity of traditional signal processing and computer-based systems to predict foetal well-being remained controversial (Ishtake and Sanap, 2013). With the development of computer science, researchers shifted towards the invention of computerised CTG which were complimentary and added synergy to the healthcare profession for assessing the well-being of babies. Computerised CTG led to a remarkable reduction in cerebral palsy perinatal mortality (Nidhal et al., 2011).

Neuro-developmental disability and cerebral palsy are often diagnosed several years after birth. Hence monitoring and appropriately interpreting foetal heart rates in the antenatal period and in labour are a priority. Since FHR signals are difficult to interpret, computer-aided systems and machine learning methods were merged with obstetrics techniques to help in interpreting FHR signals and predicting foetuses that would develop hypoxia or metabolic acidosis (Chung et al., 1995; Costa et al., 2009). This research follows the application of machine learning and deep learning methods to classify abnormal foetuses (suspect/pathologic) based on cardiotocography diagnostic features paving the way for experts to intervene at the right time and save babies' lives. This researcher's applied machine learning and deep learning methods were used to predict lung cancer (Spitz et al., 2007), breast cancer (Thongkam et al., 2009) and mortality prediction after cardiac surgery (Nilsson et al., 2006).

Artificial Neural Network (ANN), simple logistic and k-means clustering algorithms have been successfully used to classify the foetal state based on CTG recordings, thus reducing the diagnosis cost and time for hospitals, clinicians and obstetricians (Lee et al., 1999; Sahin and Subasi, 2010; Sundar et al., 2012). Discriminant Analysis (DA) and Decision tree (DT) are suitable for conducting medical classifications (Huang and Hsu, 2012). Support vector machine (SVM) with different kernels is considered an efficient classification algorithm to classify foetal distress (Krupa et al., 2011; Magenes et al., 2004; Yılmaz and Kılıkçıer, 2013). Adaptive network-based fuzzy inference system (ANIFS) and Weighted Fuzzy Scoring System (WFSS) were applied to classify foetal state and predict pathologic foetuses (Czabański et al., 2013; Ocak and Ertunc, 2013). Post-hoc power analysis, represented by chi-square goodness of fit and multivariate analysis of variance (MANOVA), was performed to determine the suitability of CTG data in the classification of abnormal foetuses (Gamage et al., 2004). Under-sampling majority class technique was used to resolve class imbalance problem avoiding biased results (Drummond and Holte, 2003). 10-fold cross-validation with 3 repeats approach was followed (Kim, 2009). As the target of this research is to classify abnormal cases, suspect and pathologic classes that have low frequencies compared to normal class were merged together converting the nature of the problem to that of binary classification. Applied algorithms are then evaluated through accuracy, recall and specificity for each model.

1.2 Research Question

The identification of abnormal foetuses based on CTG recordings can be considered as the central element for timetabled interventions by expert obstetricians to assess foetal distresses that could be a foetus is in a situation where oxygen supply is needed "hypoxia" resulting in life-long disability. The lack of a well- defined abnormal foetuses classification systems and non-clinically trained staff to correctly interpret cardiotocography recordings can mislead experts to intervene at a proper time saving foetuses' lives. To address this problem, the following question and sub-question are investigated.

RQ: "To what extent can machine learning (SVM, C5.0, RF, GLM, XGBoost, KNN and NB) and deep learning techniques (MLPNNs) help to identify the best performing algorithm to enhance the identification of abnormal foetuses, allowing practitioners to intervene early preventing foetal distress occurrence?".

Sub-RQ: "Can the current cardiotocography records (data size) and features sufficient to assist/enhance in the classification of abnormal foetuses?". To tackle the sub-RQ a post-hoc test analysis incorporating MANOVA and chi-square goodness of fit statistical techniques was used. To solve the research question and sub-RQ the following objectives were identified and selected.

1.3 Research Objectives and Contribution

The research objectives incorporate and asses a critical review of existing literature on the topics to ensure that previously applied methods and findings could be leveraged and integrated into this research and to attempt to identify significant gaps in the reviewed literature. Other objectives are shown in Table 1.

Table 1: Research Objectives

Objective	Description	Evaluation Method
1	Critical review of abnormal foetus was conducted and results presented	
2	Post-hoc test analysis	P-value, Observed Power Value
3	Feature Selection and data pre-processing	Correlation Matrix Testing
4	Exploratory analysis	
5	Implementation of Classification Models for Abnormal Foetuses	Accuracy, Recall and Specificity
5(a)	Implementation of Random Forest model	
5(b)	Implementation of C5.0 Decision Tree model	
5(c)	Implementation of Naïve Bayes model	
5(d)	Implementation of Support Vector Machine model	
5(e)	Implementation of K-Nearest Neighbour model	
5(f)	Implementation of Generalized Linear model	
5(g)	Implementation of XGBoost model	
5(h)	Implementation of Multilayer Perceptron Neural Network model	

Table1(b): Research Objectives

6	Evaluation and results of developed models (obj5(a)-5(h))	Accuracy, Recall and Specificity
7	Comparison of developed models (obj6)	
8	Comparison of developed models (obj7) verses existing models	

1.4 Contribution to the Body of Knowledge

The major contribution of this research project lies in the investigation of CTG data along with the application of artificial intelligence approach represented by developing machine learning and deep learning algorithms to find the best performing model capable of classifying abnormal foetuses. Performance-effective tuning parameters were applied for each model resulting in shorter execution times and better classification performance. This produced SVM algorithm capable of better classifying abnormal foetuses with an accuracy of 90.65%, a recall of 96.32% and a sensitivity of 89.09%, alerting clinicians and obstetricians to react accordingly preventing the occurrence of foetal complications. The investigation of existing literature on the topic, identifying gaps and achieving improvements to existing models are the minor contributions of this research.

The remainder of this report is designed as follows: Chapter 2 is an in-depth review of existing literature. Chapter 3 explains the research methodology, post-hoc analysis, data pre-processing, exploratory analysis and design specifications. Chapter 4 debates the implementation of developed models. Evaluating results are in Chapter 5. Chapter 6 discusses the research project in addition to deployment and skills learnt. To what extent applied algorithms have been successful in answering the research question and limitations of the research are shown in Chapter 7.

2 Related Work on Abnormal Foetuses

2.1 Introduction

The literature review investigates the relationship between CTGs and classification of abnormal foetuses and the application of machine learning and deep learning methods to classify abnormal foetuses. Reviewing related researches in similar abnormal foetuses' classification domain would be beneficial in the appropriate selection of machine learning and deep learning models based on medical data.

2.2 A Review of Relationship Between Cardiotocography and Abnormal Foetuses

The benefits of cardiotocography in the obstetrics field as an essential method used in the antenatal period and intrapartum for determining healthy and abnormal foetuses is explored in (Almström et al., 1992). The two primary components of CTG, namely FHR and UC which are conducted by placing two ultrasound waves-based transducers on the mother's abdomen, help obstetricians to decide on appropriate interventions preventing foetal problems to occur as stated by (Spencer, 1993). Traditional and computerised CTG are essential for assessing the

foetus's well-being. By correctly interpreting FHR by clinicians, which is between 110-160 BPM, expert obstetricians can identify foetuses with chronic hypoxia (Grivell et al., 2015).

Foetuses experienced foetal variability less than 6 BPM and deceleration with foetal movement (FM) could develop cerebral palsy or die in the perinatal period as discussed by (Parer and King, 2000). A huge number of litigations relating to delivery were related to misinterpretation of CTG recordings. Mandatory staff training courses with the help of "pulse oximetry" or "foetal electrocardiogram" to correctly interpret CTG recordings, which in turn reduces litigations, are a must (Williams and Arulkumaran, 2004). Data mining played a significant role in extracting, pre-processing, modelling and predicting clinical data providing a reliable ground for practitioners and obstetricians to diagnose abnormal cases from healthy ones (GraciaJacob and Geetha Ramani, 2012).

The Intelligent Heart Disease Prediction System (IHDPS) based on DT, NB and NN algorithms was developed to uncover hidden information from a historical heart disease database. NB achieved the highest accuracy of 95% (Ishtake and Sanap, 2013). This existing study is correlated and relevant to this research project in applying machine learning algorithms. An automated system for FHR baseline estimation based on computerised CTG to determine foetal status by (Nidhal et al., 2011), which helped in reducing cerebral palsy and perinatal mortality that had reached more than 60%.

2.3 A Critique of Machine Learning and Deep Learning Techniques

A computer-aided system was used by (Chung et al., 1995) to predict foetal acidosis at birth. An accuracy of 77% and sensitivity of 88% were recorded. Extracting more correlated features with the response variable would likely improve the performance of the applied system. Omniview-SisPorto3.5 system was developed by (Costa et al., 2009) to predict online red alerts for foetuses developing neonatal umbilical artery acidaemia with $pH \leq 7.05$. A sensitivity of 57% was recorded. While features correlation were well explained, authors did not discuss in detail the stages of building this system.

Simple Logistic (SL) and ANN models were created to predict pathologic foetuses from CTG dataset (Sahin and Subasi, 2010). SL has beaten ANN with an accuracy of 98.74% and sensitivity of 99.52%. SL model's performance was affected by the domination of majority class as class imbalance problem was not resolved. Multilayer perceptron neural networks with 15 variables as an input layer, different hidden layers and an output layers, were applied by (Lee et al., 1999) to classify abnormal foetal heart rate patterns from CTG data. The resultant sensitivity ranged between 72% and 90%. The "class_weight" parameter was not applied to resolve class imbalance problem when building these networks. The methods and objective of (Lee et al., 1999) study is an integral part of the deep learning MLPNN methods used in this research.

The classification of foetal states into normal, suspect or pathologic based on CTG data was the intention of (Sundar et al., 2012) by applying ANN algorithm. Recall for normal, suspicious and pathologic class was 99.1%, 36.88% and 97.45% respectively. The ANN poorly performed to identify the "suspicious" group. An improvement can likely be achieved by merging "suspicious" and "pathologic" classes into one class, converting to a binary classification problem. The DT, ANN and Discriminant Analysis (DA) models were created by (Huang and Hsu, 2012) to classify foetal states from CTG dataset. ANN returned an

accuracy of 97.78%. Despite this high accuracy, but it cannot be reliable as a class-balanced training dataset has not been provided to the model during developing stage.

The SVM model with 5-fold cross-validation was built to classify normal or abnormal foetuses by (Krupa et al., 2011) where an accuracy of 81.5% was recorded. SVM model could possibly perform better by tuning "cost" and "kernel" values to the optimal. (Yılmaz and Kılıkçıer, 2013) adopted least squared support vector machine with gaussian radial kernel to classify foetal state normal, suspect or pathologic. An accuracy of 91.62% was recorded. The misclassification errors for suspect and pathologic classes are high compared to the normal class, which is not useful for clinicians to determine abnormal foetuses. SVM models with different kernels were used to predict distressed foetuses (Magenes et al., 2004). SVM model with polynomial kernel produced an accuracy of 78.26%, specificity of 79% and sensitivity of 78%. SVM can perhaps be improved by applying Independent Component Analysis (ICA) to select significantly correlated features.

An Adaptive Neuro-Fuzzy Inference Systems (ANFIS) of 13 fuzzy rules, 596 nodes and 832 parameters, and ANN model was used by (Ocak and Ertunc, 2013) to predict normal or pathologic foetuses based on CTG recordings. ANFIS outweighed ANN by an accuracy for normal and pathologic state of 97.2% and of 96.6% respectively. The parameters used for ANFIS were not well explained by authors. ANFIS performed better at normal class which is not desirable by clinicians. A Weighted Fuzzy Scoring System (WFSS) using MATLAB R2010a and JacketTM package was adopted by (Czabański et al., 2013) to classify foetal states. AUC reached to 79% which indicates a reasonable degree of certainty of WFSS. The model may probably be improved by excluding highly correlated features to avoid multicollinearity problem.

This research project also reviewed a significant body of work exists on the application of machine learning and deep techniques learning that provide useful and relevant insights. Logistic Regression algorithm was applied by (Spitz et al., 2007) to predict the probability of having lung cancer for never, former and current smokers. Area Under Curve (AUC) for never, former and current smokers was 57%, 63% and 58% respectively. The bias problem in the used Epidemiologic data was not tackled, which caused model overfitting that returned unreliable results.

AdaBoost, Bagging, C4.5 and SVM models were developed by (Thongkam et al., 2009) with a combination of outlier filtering and over-sampling (OOS) approach to predict the breast cancer survivability. SVM model, in OOS approach, showed an accuracy of 98.13%, sensitivity of 97.87% and specificity of 98.38%. SVM has performed better at predicting the "not survived" class, at a higher specificity, which contradicts with the main objective of the study. (Nilsson et al., 2006) adopted ANNs, with 4-fold and 6-fold cross-validation, to predict mortalities after cardiac surgery. The AUC of 81% and a sensitivity of 75% were recorded. This is considered an improvement compared to the logistic European System for Cardiac Operative Risk Evaluation model with an AUC of 79% and a sensitivity of 73%.

The MAOVA is a reliable technique to determine the power of a study regarding the effect of data size and the suitability of features throughout a desirable observed power value ≥ 40% (Gamage et al., 2004). Cost curves used in (Drummond and Holte, 2003) to show the interactions between over-sampling and under-sampling techniques applied on C.4.5 algorithm. Under-sampling technique produced a remarkable change in the model's predictive performance. In this research, under-sampling majority class effectively increased accuracy and sensitivity of the classifiers compared to Synthetic Minority Over-sampling Technique

(SMOTE). (Kim, 2009) performed a repeated comparison between .632+ bootstrap estimator and 10-fold cross-validation estimator showed the latter has a better performance when the classifier is built on the training sample. In addition, 10-fold cross-validation estimator can reduce the bias in a dataset which is efficient in regard to model's performance.

2.4 Comparison of Reviewed Classification Techniques and Datasets

Sourced datasets, targets, applied classifiers and results from reviewed literature are compared in Table 2. Obviously, (Sahin and Subasi, 2010) achieved the highest classification accuracy using Simple Logistic (SL) model. Consistent with the objectives of the research question, results achieved by (Lee et al., 1999), (Magenes et al., 2004) and (Krupa et al., 2011) to classify abnormal foetuses based on CTG dataset could be considered as a benchmark to be compared with results acquired from applied machine leaning and deep learning methods in this research project.

Table 2: Comparison of Reviewed Classification Techniques

Datasets	Target	Techniques	Results	Authors
CTG dataset	Predict pathologic foetuses	SL	Accuracy of 98.74%	(Sahin and Subasi, 2010)
CTG dataset of 53 patients	Classify Abnormal foetuses	ANNs	Sensitivities of 70% - 90%	(Lee et al., 1999)
CTG dataset	Foetal distress prediction	SVM	Accuracy of 78.26% Sensitivity of 78%	(Magenes et al., 2004)
CTG dataset	Classify Abnormal foetuses	SVM	Accuracy of 81.5%	(Krupa et al., 2011)

2.5 Identified Gaps

According to the reviewed literature findings, there has been no post-hoc test analysis to determine the power of the existing studies regarding the correlation between features and foetal health. The class imbalance problem in CTG dataset was not adequately/incorporated that could have mitigated the risk relative to criticisms that the domination of the high-frequency class on the model's overall performance could mislead and fail to correctly classify abnormal foetuses. In addition, handling outlier problem was not tackled and the reason for selecting applied algorithms has not been justified by reviewed literature.

2.6 Conclusion

Based on the literature discussed above, machine learning and deep learning algorithms were applied on medical data to classify a binary or a multi-class targeted response variable to identify non-healthy foetuses. The relationship between CTG data and machine learning and deep learning applications was well presented by the reviewed literature. To address the identified gaps and to answer the research question, research methodology and design specification have been tailored as outlined in the next chapter.

3 Research Methodology, Data Pre-processing and Design Specification

3.1 Introduction

The research methodology for abnormal foetuses classification followed the approach of the widely popular data mining Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology to answer the research question (Wirth and Hipp, 2000). Additionally, post-hoc test analysis, data pre-processing, exploratory analysis and design specification are presented in this chapter.

3.2 Abnormal Foetuses Methodology

The adopted methodology (refer Figure 1) tracks the following stages: (i) understanding the critical components relative to the research question and sub-question; (ii) extracting raw data from the repository in Excel ".xls" format; (iii) post-hoc analysis is performed to ensure the data suitability and features to be selected; (iv) data is pre-processed and transformed; (v) machine leaning and deep learning classifiers are developed; (vi) models are evaluated based on accuracy, recall and specificity; (vii) final results are visualised using a data mining visualisation tool to get information at a glance.

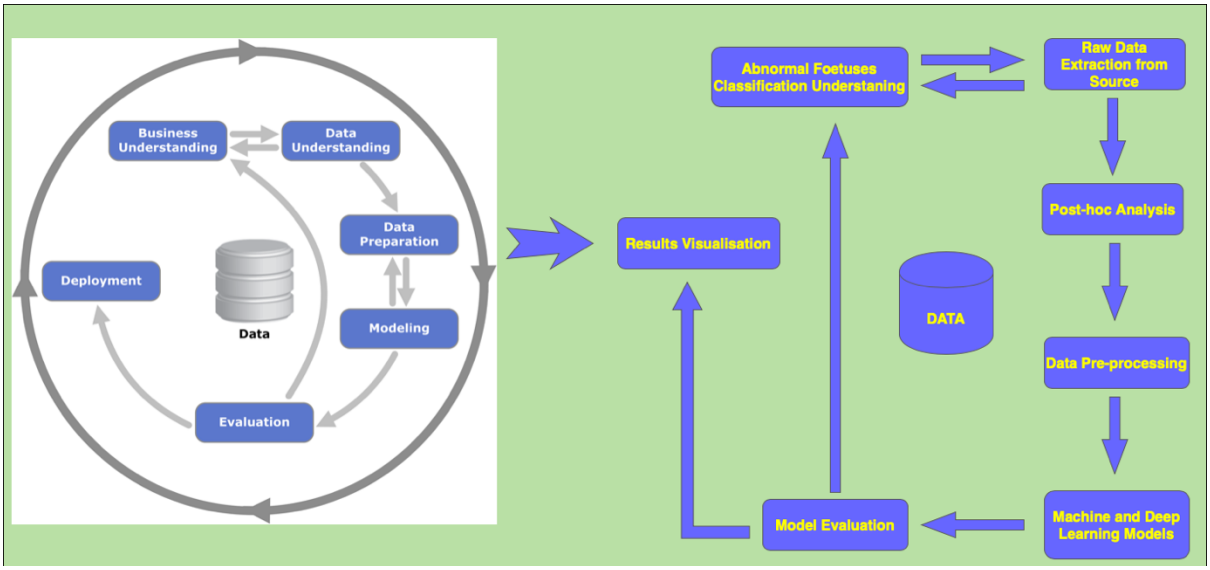

Figure 1: Abnormal Foetuses Methodology

3.3 Post-hoc Test Analysis

Raw data was extracted from UCI Machine Learning Repository in Excel extension ".xls" format. The data consists of measurements of FHR, UC features and classification labels recorded by three expert obstetricians from 2126 foetal cardiotocographs that were automatically extracted and analysed using SisPorto 2.0 (Ayres-de-campos et al., 2000). Experts classified 1655 cases as normal, 295 cases as suspicious and 176 cases as pathologic. This research focused on (NSP = Normal, Suspect or Pathologic) classification label as the predicted response variable to classify abnormal "Suspect or Pathologic" foetuses based on 21

diagnostic features. Data was converted to Comma Separated Values ".csv" format for speed and to consume less memory when importing data into a data mining tool.

From the sourced raw data, the classification label NSP and all other measurement diagnostic features in CTG dataset were selected to be examined by the post-hoc test analysis for suitability. The first part of post-hoc analysis represented by a chi-square goodness of fit test (refer Figure 2), was performed to scientifically examine the significant relationship between the different groups relating to the response variable NSP. A p-value less than 0.05 indicated a substantial evidence to reject the null hypothesis of this test, indicating no existing relationship between the different groups. This confirmed the suitability of the response variable.

Test Statistics

	NSP
Chi-Square	64.864[a]
df	2
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 212.6.

Figure 2: Chi-Square Goodness of Fit

The second part of post-hoc analysis was performed by applying MANOVA where the four multivariate tests (refer Figure 3) were significant to the response variable NSP with p-value > 0.05.

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[d]
Intercept	Pillai's Trace	.992	13167.832[b]	20.000	2104.000	.000	.992	263356.645	1.000
	Wilks' Lambda	.008	13167.832[b]	20.000	2104.000	.000	.992	263356.645	1.000
	Hotelling's Trace	125.170	13167.832[b]	20.000	2104.000	.000	.992	263356.645	1.000
	Roy's Largest Root	125.170	13167.832[b]	20.000	2104.000	.000	.992	263356.645	1.000
NSP	Pillai's Trace	.896	85.338	40.000	4210.000	.000	.448	3413.539	1.000
	Wilks' Lambda	.277	94.745[b]	40.000	4208.000	.000	.474	3789.790	1.000
	Hotelling's Trace	1.990	104.610	40.000	4206.000	.000	.499	4184.411	1.000
	Roy's Largest Root	1.601	168.484[c]	20.000	2105.000	.000	.616	3369.686	1.000

a. Design: Intercept + NSP
b. Exact statistic
c. The statistic is an upper bound on F that yields a lower bound on the significance level.
d. Computed using alpha =

Figure 3: Multivariate Tests

Testing the homogeneity of variance (refer Figure 4), Leven's test was significant with p-value less than 0.05 for majority of predictor variables indicating a rejection of null hypothesis, where the variances across the different groups of the dependent variable are equal.

Levene's Test of Equality of Error Variances[a]

	F	df1	df2	Sig.
LB	8.453	2	2123	.000
AC	367.699	2	2123	.000
FM	40.583	2	2123	.000
UC	23.247	2	2123	.000
DL	146.646	2	2123	.000
DS	126.383	2	2123	.000
DP	1003.532	2	2123	.000
ASTV	54.549	2	2123	.000
MSTV	75.446	2	2123	.000
ALTV	506.172	2	2123	.000
MLTV	31.742	2	2123	.000
Width	30.045	2	2123	.000
Min	25.427	2	2123	.000
Max	12.053	2	2123	.000
Nmax	4.788	2	2123	.008
Nzeros	4.429	2	2123	.012
Mode	170.252	2	2123	.000
Mean	139.261	2	2123	.000
Median	64.015	2	2123	.000
Variance	261.187	2	2123	.000
Tendency	.894	2	2123	.409

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + NSP

Figure 4: Levene's Test

Tests of between-subject effects (refer Table 3), that was taken from SPSS software, reverted observed power values greater than 40% for all measurable independent variables used for diagnosis. Additionally, the p-value is less than 0.05 for majority of the independent variables. As a such, post-hoc test analysis showed that CTG data size and features are sufficiently efficient to be used for developing machine learning and deep learning models.

Table 3: Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[v]
NSP	LB	24072.758	2	12036.379	140.621	.000	.117	281.242	1.000
	AC	.005	2	.002	196.321	.000	.156	392.643	1.000
	FM	.050	2	.025	11.662	.000	.011	23.323	.994
	UC	.002	2	.001	94.760	.000	.082	189.521	1.000
	DL	.001	2	.001	68.709	.000	.061	137.419	1.000
	DS	2.306E-7	2	1.153E-7	29.907	.000	.027	59.813	1.000
	DP	.000	2	.000	524.375	.000	.331	1048.750	1.000
	ASTV	153677.101	2	76838.550	343.820	.000	.245	687.641	1.000
	MSTV	168.221	2	84.111	119.882	.000	.101	239.764	1.000
	ALTV	176471.737	2	88235.868	345.156	.000	.245	690.313	1.000
	MLTV	4174.078	2	2087.039	70.174	.000	.062	140.348	1.000
	Width	159098.710	2	79549.355	55.088	.000	.049	110.176	1.000
	Min	141165.958	2	70582.979	87.341	.000	.076	174.681	1.000
	Max	1585.194	2	792.597	2.465	.085	.002	4.930	.497
	Nmax	208.419	2	104.209	12.105	.000	.011	24.210	.996
	Nzeros	2.187	2	1.094	2.196	.111	.002	4.393	.450
	Mode	117372.516	2	58686.258	275.118	.000	.206	550.235	1.000
	Mean	113151.960	2	56575.980	297.625	.000	.219	595.251	1.000
	Median	84436.765	2	42218.382	248.772	.000	.190	497.544	1.000
	Variance	221956.634	2	110978.317	150.797	.000	.124	301.594	1.000
	Tendency	31.930	2	15.965	44.542	.000	.040	89.085	1.000

3.4 Feature Selection and Data Pre-processing

Multicollinearity problem, where there is an excessive correlation greater than 95% between independent variables was not identified in CTG dataset as shown in the correlation matrix (refer Figure 5). Multicollinearity, if exists, will result in the instability of the model's estimates and, accordingly, the classification model may not be accurate due to model overfitting A positive correlation between heartbeat acceleration (AC) and UC for normal foetuses was identified. Additionally, a negative correlation was observed between the abnormal long-term availability (ALTV) and heartbeat decelerations (DP) for suspect or pathologic foetuses. This significant correlation and the absence of multicollinearity confirmed the suitability of the measurement features to be selected for machine learning and deep learning applications.

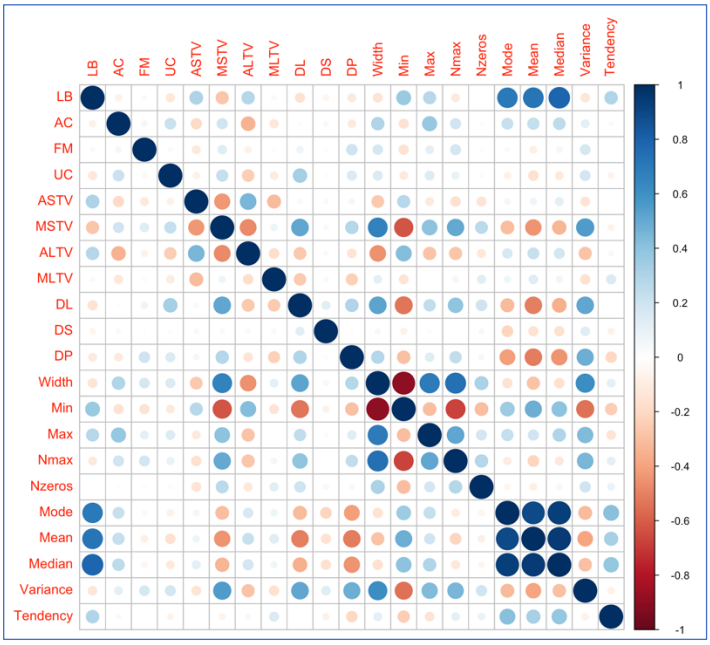

Figure 5: Correlation Matrix for Independent Variables

No missing values were detected in CTG dataset. The three classes of response variable have been transformed to a binary class where normal fetuses were coded to "0" and suspect or pathologic fetuses were coded to "1". Data points distribution of the binary class is noticeably imbalanced (refer Figure 6) with a domination of the Normal class, which will inevitably affect the predictive models' accuracy, recall and specificity. Under-sampling majority class techniques were applied to resolve the class imbalance problem.

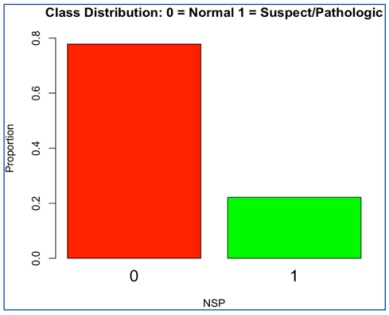

Figure 6: Class Imbalanced Distribution

The data was inspected for upper and lower outliers to avoid negative impacts on the developed classifiers' performance. Some of the independent variables showed outliers, which are the points outside the whiskers in Figure 7 that could be a result of SisPorto 2.0 errors or extreme-tails of the population distribution. A "winsorising" technique was followed to handle outliers where upper outliers were replaced by $Q3 + 1.5 * IQR$, (Q3 is 3rd quantile and IQR is the interquartile range). The lower outliers were replaced by $Q2 - 1.5 * IQR$, (Q2 is 2nd quartile) (Hoo et al., 2002).

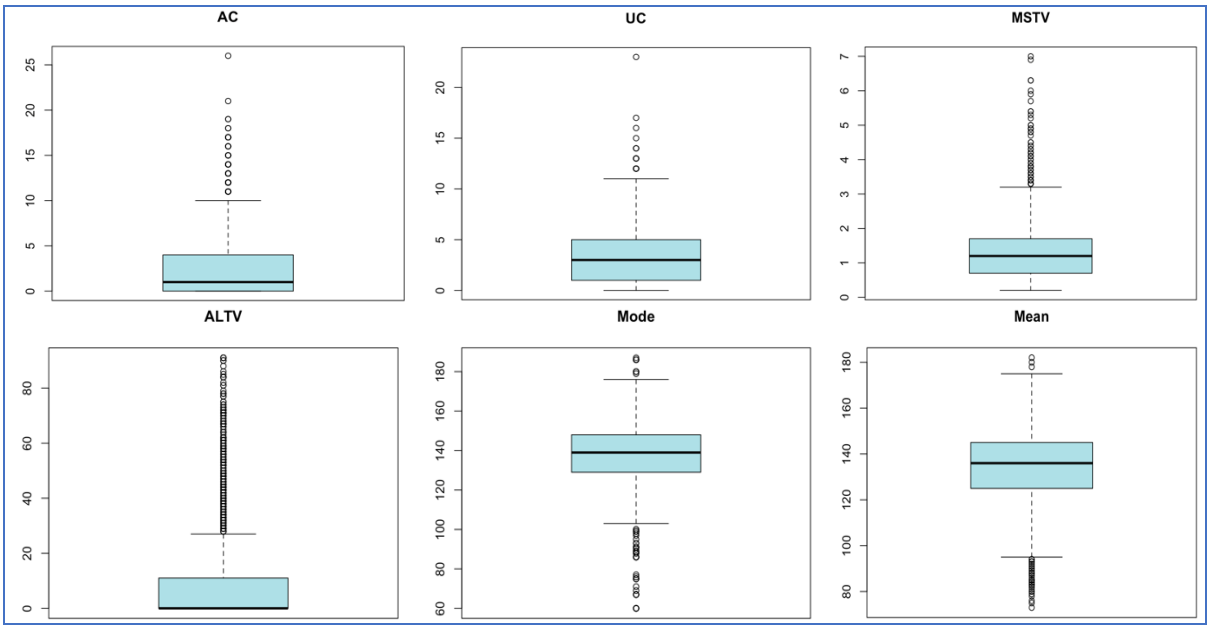

Figure 7: Boxplots for Some Features Suffering Outliers

The CTG data was divided into two subsets, a training set of 70% for building the models and a testing set of 30% for evaluating the developed models. A 10-fold cross-validation technique with 3 repeats was applied, where in each cross-validation training set is split into 10 folds, 9 of them are used for creating the model and the remaining one is used for error estimation.

3.5 Exploratory Data Analysis

Inspecting a histogram for data points distribution for some independent variables (refer Figure 8) showed that features tend to follow symmetric distribution through the examination of the mode, median and variance.

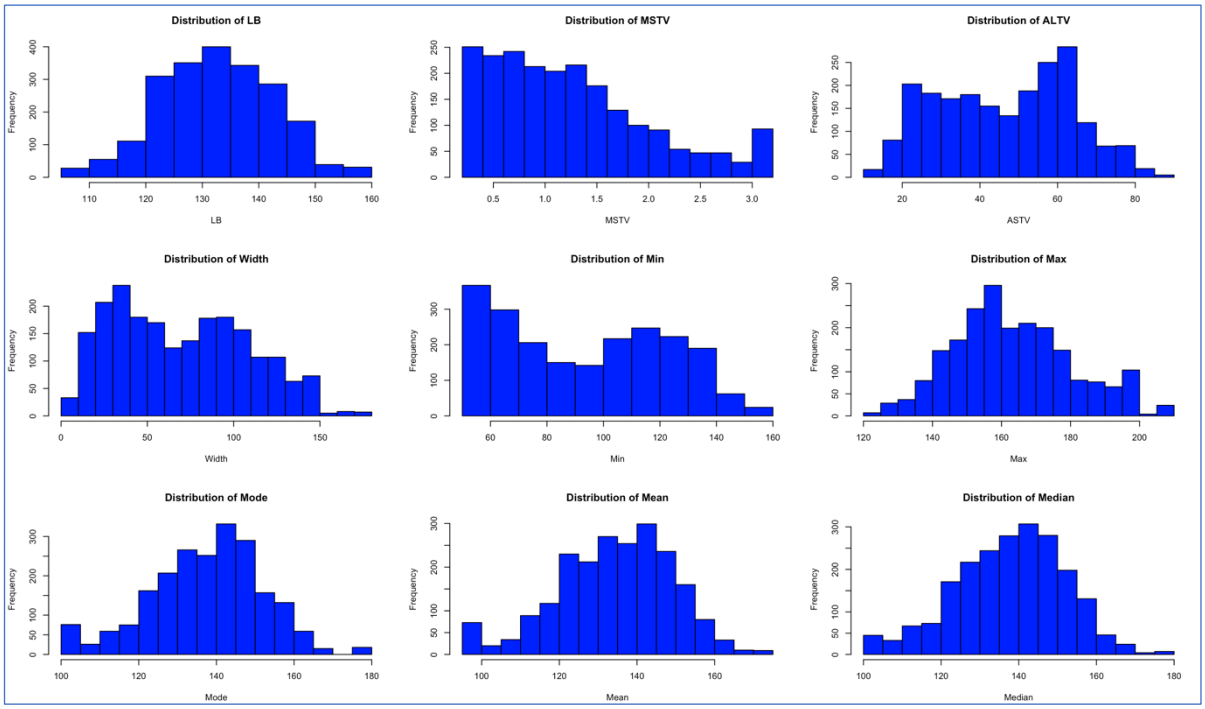

Figure 8: Histogram for Data Points Distribution of Independent Variables

Based on the RF algorithm built of 500 trees (refer Figure 9), variables that played important roles to classify abnormal foetuses were shown. The first graph showed variables having significant importance based on accuracy, whereas the second graph showed significantly important variables in terms of the "mean decrease Gini". The variable "DS", according to both graphs, is not important for classification as it has an almost-zero contribution.

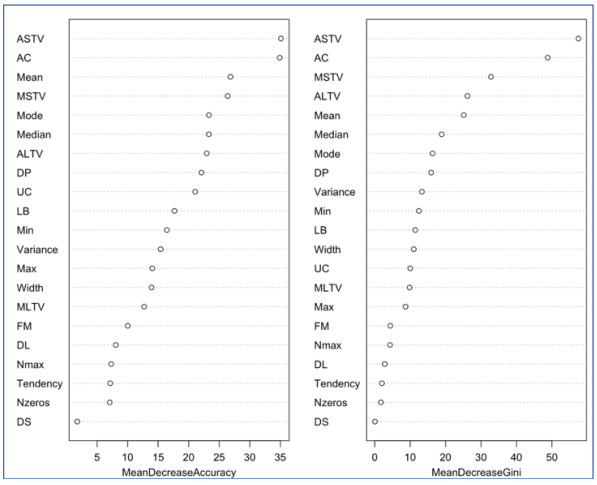

Figure 9:Variable Importance Based on 500 Tree Random Forest

Exploring the relationship between the NSP variable and other independents (refer Figure 10), the average of uterine contraction UC is higher for normal class coded as "0", whereas it is lower for abnormal (suspect or pathologic) class coded as "1", and vice versa for abnormal long terms variability (ALTV) variable.

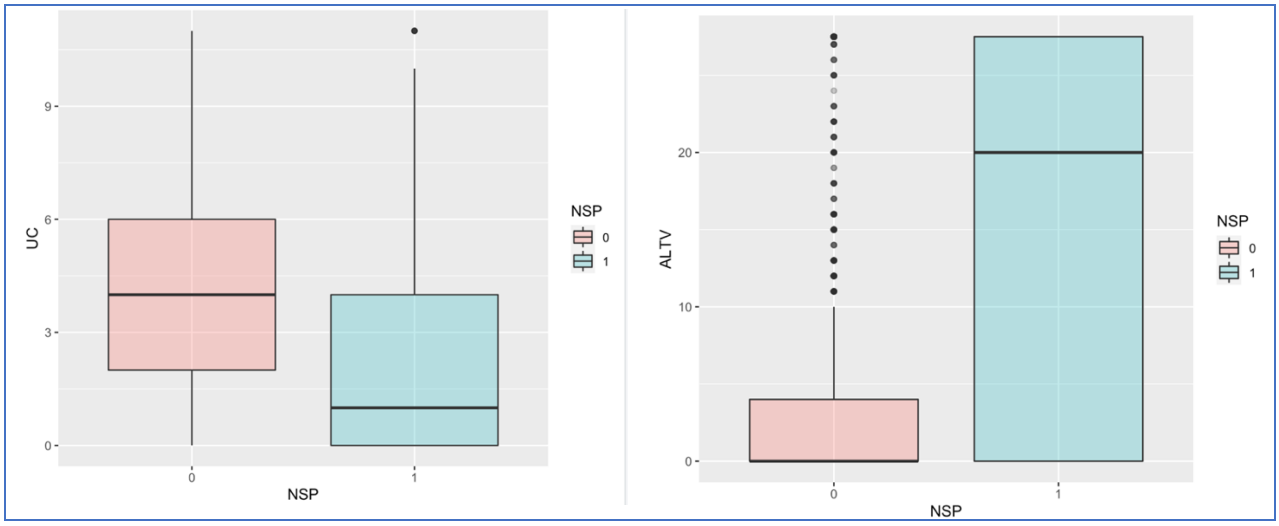

Figure 10: Relationship Between Predicted and Predictor variables (UC, ALTV)

3.6 Research Technical Design

The research process flow (refer Figure 11) of classifying abnormal foetuses consists of (i) a client tier where raw data is extracted from the source and (ii) computational tier where post-hoc test analysis, data pre-processing and machine learning and deep learning models were implemented, (iii) the presentation tier that displays the visualisation of results.

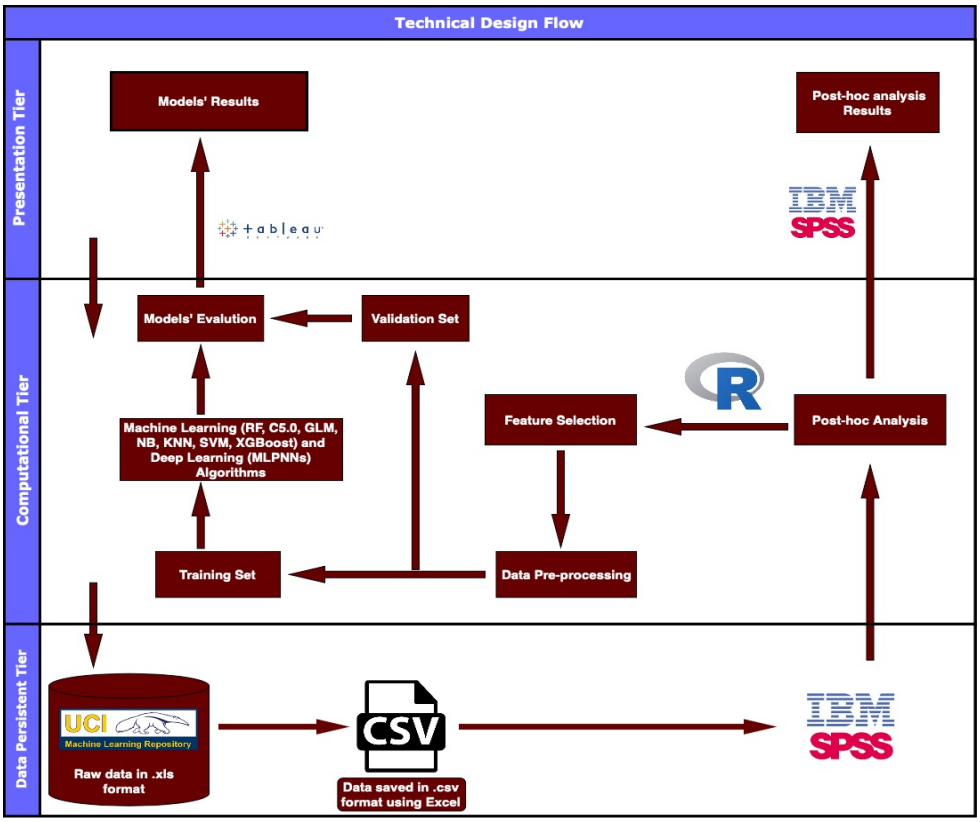

Figure 11: Abnormal Foetuses Classification Design Flow

3.7 Conclusion

The research selected and deployed adopted abnormal foetuses' methodology with tailored procedures to improve the different applied classification models' performance which, in turn,

lead to reliable results. Post-hoc test analysis and data pre-processing made CTG data ready for the implementation of machine learning and deep learning algorithms as will be shown in the following chapter.

4 Implementation of Abnormal Fetuses Classification Models

4.1 Introduction

Different machine learning and deep learning classifiers were developed to address the two-class classification problem in this research. Each model was trained using class-balanced training dataset, 10-fold cross-validation technique with 3 repeats, and tuning parameters to identify the best performing classifier based on accuracy, recall and specificity. In order to get consistent results for each time running machine learning models, R-base "set.seed" function has set to a specific number for each model.

4.2 Implementation of Random Forest Model

The RF model was developed using "randomForest" package (Liaw and Wiener, 2002) with 500 trees and a number of variables that are randomly sampled at each split (mtry = 4). This was obtained by the square root of the number of independent variables in classification problems. Additional tuning parameters such as "importance", "proximity" and "keep.forest" were applied when building RF model by setting their value to "TRUE". As the number of trees grows, out of bag error estimate "OOB" initially drops down then it becomes more or less constant after around 400 trees (refer Figure 12).

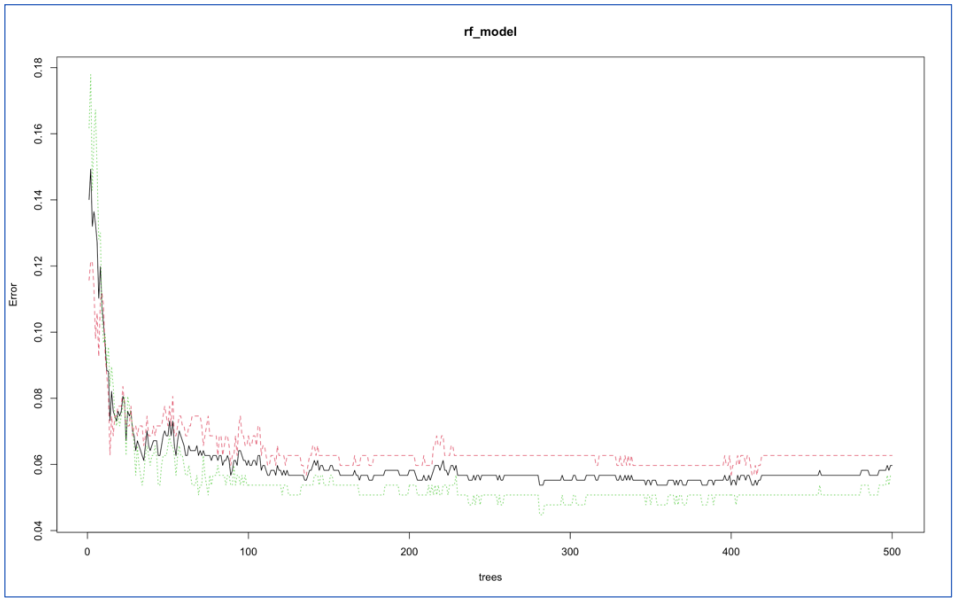

Figure 12: RF Out of Bag Error Estimate

4.3 Implementation of C5.0 Decision Tree Model

A well-known implementation of decision trees is C5.0 decision tree algorithms with its developed mechanism concentrating on post-prune the tree preventing overfitting the model.

The model was created using "C50" (Kuhn et al., 2020) package, where the most critical variable "AC" at which the tree starts to split was identified. Additionally, the model returned 21 trees with a high accuracy and a substantial Kappa value.

4.4 Implementation of Naïve Bayes Model

The simple, fast and linearly scalable NB model was developed using "naivebayes" package (Majka, 2020). To improve the performance, the model was tuned by setting kernel-based density to "true" as some numerical variables are not normally distributed. NB was trained to obtain reliable recall value indicating the probability of classifying abnormal foetuses.

4.5 Implementation of Support Vector Machine Model

SVM algorithm was developed using "e1071" package (Meyer et al., 2019) along with effective tuning parameters to obtain the optimal separating hyperplane between the two classes based on a kernel trick. Firstly, a tuned model was built using a sequence of numbers to find the best values for "epsilon" and "cost" parameters. The best SVM model was taken from the tuned model with radial kernel based function, "epsilon = 0" and "cost = 4" represented by the darker area where less misclassification errors were recorded (refer Figure 13). The resultant support vectors were 197; 94 were for class "0 = Normal" and 103 for class "1= Abnormal".

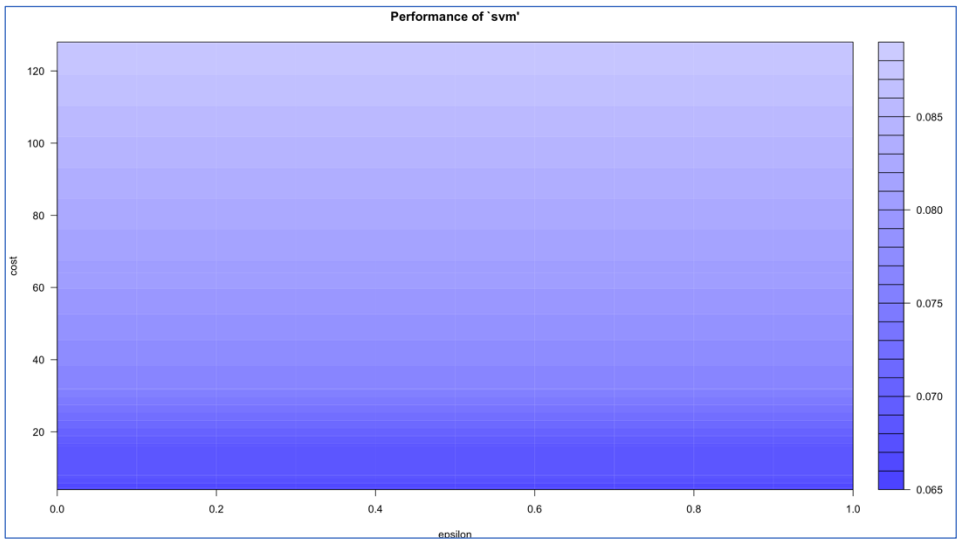

Figure 13: SVM Cost and Epsilon Optimal Values

4.6 Implementation of K-Nearest Neighbour Model

KNN model was built using "caret" package (Kuhn, 2020) with tuning parameters "tuneLength = 20". Additionally, to create a level plane field data points were standardised by subtracting each data point from mean and dividing by standard deviation using "centre" and "scale" properties. KNN returned the optimal k value equals 7 based on accuracy as a performance metric (refer Figure 14).

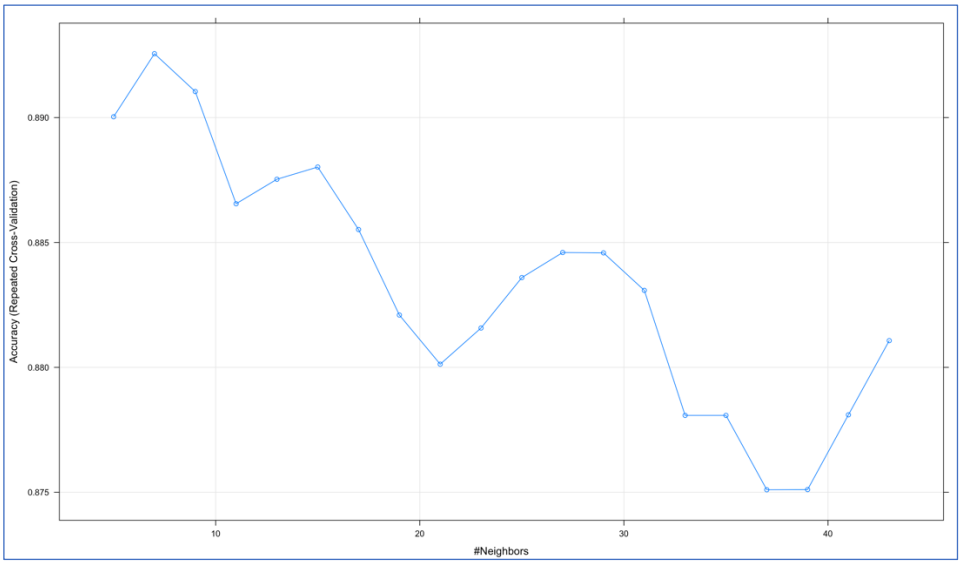

Figure 14: KNN Optimal K Value

4.7 Implementation of Generalised Linear Model

The GLM model was developed using built-in R statistical package. The model was built with "family = binomial". Class predictions were based on "type = response" and a cut-off value of 0.5 or lower being classified as "0 = Normal" and above 0.5 being classified as "1= Abnormal".

4.8 Implementation of Extreme Gradient Boosting Model

The package "xgboost" (Chen et al., 2020) was used to develop the XGBoost model. Train and test datasets were converted to a matrix form using one hot encoding technique. The model was optimised by a list of tuning parameters to enhance performance, "objective = multi:softprob", evaluation metric "eval_metric = mlogloss", number of classes "num_class = 2", "eta = 0.01" to avoid overfitting the model, "max_depth = 5", number of iteration "nrounds = 1000" and a watch list containing the new matrix-shaped train and test datasets which allows to watch how much error exists at each iteration.

4.9 Implementation of Deep Learning Multilayer Perceptron Neural Network Models

Three different MPNN models were developed using "Keras and TensorFlow" packages (Falbel et al., 2020) to obtain the best performing model. 1st model consisted of 21 variables input layer, one hidden layer of 8 neurons and an output layer. 2nd and 3rd models were built with similar input and output layers but with one hidden with 50 neurons for 2nd model and two hidden layers with 50 and 8 neurons for 3rd model. All models used a rectified linear units "relu" activation function for hidden layer and "softmax" activation function for output layer which keeps range between 0 and 1 that can be used as probabilities. Since the research problem is a binary class one, models were compiled with "loss = binary_crossentropy", "optimizer = adam", "metrics = accuracy". Models were fitted with normalised and matrix-formed training dataset and "epocs = 200", "validation_split = 0.2" where 20% of data points are used and "class_weight" to resolve class unbalance problem existed in the response variable.

4.10 Conclusion

Machine learning and deep learning models were time-consuming to develop due to efforts exerted to identify the best hyperparameters to control and enhance the classification behaviour of each model, allowing for a fast-computational process and a better performance relating to accuracy, recall and specificity being the primary evaluation metrics for this research. This chapter extensively addresses objective 5, incorporating 5(a) to 5(j), from table 1: Research Objectives. The developed algorithms and the extracted results will be evaluated in next chapter.

5 Evaluation and Results

5.1 Introduction

The developed machine learning and deep learning models were evaluated based on performance metrics; accuracy, recall and specificity to identify the best performing model. The amended Table 4 from the typical classification model advises the following formulas;

Table 4: Confusion Matrix

	Actual	
Prediction	0 = Normal	1 = Abnormal
0 = Normal	A	B
1 = Abnormal	C	D

$$Accuracy = \frac{A+D}{A+B+C+D} \quad Recall = \frac{D}{D+B} \quad Specificity = \frac{A}{A+C}$$

The above mentioned formulas and table were derived from "confusionMatrix" function available within "caret" package (Kuhn, 2020).

5.2 Evaluation and Results

Evaluation metrics; accuracy, recall and specificity obtained from the confusion matrix for each model were represented for informative insights (refer Figure 15). For healthcare providers dealing with cardiotocographs, the potential task is to classify abnormal foetuses providing useful indicators to experts to act on time. Considering this, recall could be seen as the most important evaluation metric for abnormal foetuses' classification and must be prioritised over specificity. Recall measures what proportion of all positive values were correctly classified (a positive outcome 1 = abnormal). During the preliminary screening, it is essential to maximise the detection of abnormal foetuses (true positives), even at the cost of incorrectly detecting more false positives. False positives can be identified later upon closer examination of the foetus.

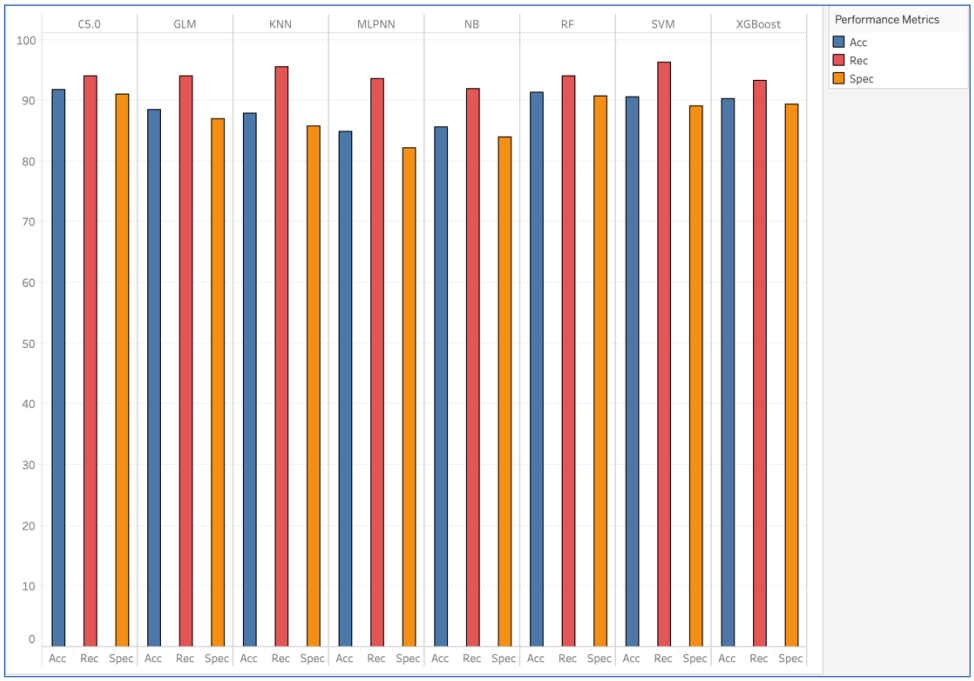

Figure 15: Accuracy, Recall and Specificity of the Developed Models

5.3 Comparison of Developed Models

Table 5 illustrates that SVM recall is significantly higher than the KNN algorithm at 96.32% compared to 94.12% for RF and C5.0 algorithms. Specificity will also need to be high enough to make utilising the model viable. Specificity prioritises true negative classes instead of true positives. As can be seen, the C5.0, RF and XGBoost algorithms have slightly better specificity in comparison to SVM (91% to 89%) – a trade-off for lower recall. With regards to Accuracy, SVM holds up well to the decision tree algorithms. SVM is only outperformed by the C5.0 algorithm by about 1.10%.

Table 5: Comparison of the Developed Models Performance (%)

Model	C5.0	GLM	KNN	MLPNNs	NB	RF	SVM	XGBoost
Accuracy	91.76	88.58	87.96	84.90	85.73	91.44	90.65	90.33
Recall	94.12	94.11	95.59	93.70	91.91	94.12	96.32	93.38
Specificity	91.11	87.07	85.86	82.17	84.04	90.71	89.09	89.49

5.4 Comparison of Developed Models vs. Existing Models

Based on a comparison in Table 2 (chapter 2, sub-section 2.4), (Lee et al., 1999) achieved the highest sensitivity of 90% for abnormal foetuses using ANN. This research project could achieve an advancement regarding (Lee et al., 1999) ANN by applying MLPNN algorithms with multiple different layers and a collection of performance-effective tuning parameters to get an accuracy of 84.90% and a sensitivity of 93.70% as shown in Table 6. Another advancement this research project could achieve was the ability to classify abnormal foetuses by using the SVM algorithm which returned an accuracy of 90.65% and a sensitivity of 96.32% compared to (Krupa et al., 2011) SVM accuracy of 81.5% and (Magenes et al., 2004) SVM accuracy of 78.26% and a sensitivity of 78%.

Table 6: Comparison with Existing Models (%)

Author	Model	Accuracy	Recall
(Lee et al., 1999)	ANN		90.0
(Krupa et al., 2011)	SVM	81.5	
(Magenes et al., 2004)	SVM	78.26	78.0

5.5 Conclusion

Results from machine learning and deep learning classifiers could be seen as promising and helpful regarding clinical data mining applications. Due to the differences in accuracy between the decision tree and SVM algorithms being small enough, and because SVM has a higher recall, (a recall high enough to account for dozens more actual foetus abnormalities in prenatal screening), SVM should be pursued as a relevant and useful addition when combined with clinical diagnostics for foetal abnormality detection. This addresses the evaluation of results and the best model selection which is objective 6 in Table 1: Research Objectives (chapter 1, sub-section 1.3). The research findings are discussed in the following.

6 Discussion

Several elements are considered in a comprehensive review of this research project which fall in the following categories: post-hoc analysis and pre-processing methods, machine learning and deep learning models' selection and research deployment and skills learnt.

6.1 Post-hoc Analysis and Pre-processing Methods

The significant difference between the means of the two different groups in NSP response represented by of chi-square goodness of fit with p-value < 0.05, and the high observed power value for each independent variable returned by MANOVA technique proved the power of this research and formed a stepping stone towards the next steps of data pre-processing and the application of machine learning and deep learning algorithms.

In accordance with answering the research question, the first method in data pre-processing stage was transforming CTG data into two-class classification problem by coding Normal foetuses' class to "0" and suspect or pathologic foetuses' class to "1". This made it easier for classifiers and led to a shorter running time. Fortunately, the data was complete and did not have missing values. The generated correlation matrix played an essential role in deciding the features to be selected, thus excluding multicollinearity amongst predictor variables.

Since medical errors form a significant problem in the healthcare sector, outliers in CTG data were handled by winsorizing technique. This step was crucial as it reduces bias and misclassification errors of developed models. In order to avoid classification bias, the class imbalance problem has been resolved by under-sampling technique to gain a balanced training data that helps in building balanced models and avoids the domination of the higher class, which can cause a low predictive model's accuracy for the infrequent class.

6.2 Machine Learning and Deep Learning Models Selection

Implementation chapter represented a crucial step on what specific classifiers have to be selected among a variety of supervised machine learning and deep learning algorithms to

address the research problem. Since this research project dealt with two-group classification problem and with a data without missing values, GLM was the first candidate model to be applied to discover the influence of predictors on the predicted variable and to gain the probability of abnormal foetus classification. To reduce the developed model's overfitting and to get an estimate of important predictors, RF method was selected. The number of trees of 500 were used for quicker run times in model training.

Because of the ability to deal with numerical and categorical variables, The C5.0 Decision Tree algorithm was selected to classify abnormal foetuses. NB classifier was among other candidates because of the fast execution time and the ability to run in case of small training data. Due to easy implementation and robustness, the KNN algorithm was selected where the classification of abnormal foetuses was computed based on majority votes of k-nearest neighbours to each data point.

The SVM was selected because of the kernel trick and regularisation parameter which transformed CTG data to the required form to find the optimal boundaries between CTG data points and to avoid the overfitting problem. Extreme Gradient Boosting has been chosen over regular gradient boosting for the ability to do parallel computations, take different types of input data and minimize the loss function that reduces the misclassification error in the resulted confusion matrix (Chen et al., 2020). The selection of Multi-layer Perceptron Neural Networks lies in the backpropagation technique which reduces the errors made by the network by distinguishing the not linearly separable data and the ability to do feature selection automatically. Thus, developed models were not randomly selected but with a specific purpose in mind.

6.3 Research Deployment and Skills Learnt

This research project can be deployed in healthcare sector incorporating hospitals, expert obstetricians and clinicians as beneficiaries to alert them about foetuses suffering foetal abnormality, enabling them to combine a medical diagnostic expertise with a view to taking appropriate actions and to provide better outcomes before it is too late. Inaccurate indications for abnormal foetuses will result in hospitals losing their money, exposure to litigation, damage to reputations and the prioritisation of their efforts and time that could better be allocated for a really deserved abnormal cases.

New data analytics skills were acquired during this research journey; constructing a post-hoc analysis for testing the power of the study and the significance of independent features, handling outliers and class imbalance problem and, most importantly, the skill of tuning machine learning and deep learning algorithms for a faster execution and delivering consistent results.

This research provides positive insights and confidence that could lead to further exploration on the topic, with the availability of larger data set regarding foetal heart rate and uterine contraction. This additional data can lead to an improved classification performance, providing synergistic benefits by the combination of technology, data mining and clinical skills that may enhance and provide a new methodology by which similar or related research questions are tackled.

7 Conclusion and Future Work

The aim of this research was to train machine learning and deep learning algorithms in order to decide the best performing classifier capable of classifying abnormal foetuses based on CTG recordings that helps in preventing foetal problems by a proper and on-time intervention by expert obstetricians.

All research objectives (chapter 1, sub-section 1.3) have been implemented. Based on developed solutions in this research project within the context of the data available, machine learning overweighed its subset deep leaning through the performance of SVM algorithm which scored the best performing results among all other seven developed and evaluated classifiers addressing Research Question (chapter 1, sub-section 1.2) comprehensively. This project also identified the power of this study by following post-hoc test analysis approach returning significant p-values and high observed power values for independent variables which confirmed the suitability of CTG data for abnormal foetuses' classification. Additionally, a correlation matrix showing no multicollinearity confirmed the effectiveness of independent features on the models' performance addressing research sub-question (chapter 1, sub-section 1.2).

The limitation of this research is centred on providing greater volume of CTG data which, in turn, supplements the volume of the training data that could enhance the performance of developed models regarding the targeted business application. Since C5.0 and SVM models showed a promising performance in classifying abnormal foetuses in terms of accuracy and recall, a hybrid application combining SVM and C5.0 "SVM-DT" along with a large volume of CTG data could be explored in future research.

Overall, by applying multiple machine and deep learning algorithms along with data pre-processing and the use of tuning parameters for each model, this research project for classifying abnormal foetuses could push forward health and medical data mining applications allowing clinics, hospitals and experts to make more timely and accurate intervention decisions that will lead to better outcomes for a foetus and perhaps save lives.

8 Acknowledgement

Foremost, I would like to express my sincere thanks to Dr Catherine Mulwa for her guidance and support throughout this research project. I also wish to acknowledge UIC machine learning repository for allowing CTG dataset to be publicly accessed. My deep gratitude goes to my family for their continuing support to complete this research project.

References

Almström, H., Ekman, G., Axelsson, O., Ulmsten, U., Cnattingius, S., Maesel, A., Maršál, K., Årström, K., 1992. Comparison of umbilical-artery velocimetry and cardiotocography for surveillance of small-for-gestational-age fetuses. The Lancet 340, 936–940. https://doi.org/10.1016/0140-6736(92)92818-Z

Ayres-de-campos, D., Bernardes, J., Garrido, A., Marques-de-sá, J., Pereira-leite, L., 2000. SisPorto 2.0: A Program for Automated Analysis of Cardiotocograms. J. Matern. Fetal Med. 9, 311–318. https://doi.org/10.3109/14767050009053454

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., 2020. Extreme Gradient Boosting [WWW Document]. R Package Version 1002. URL https://CRAN.R-project.org/package=xgboost (accessed 6.12.20).

Chung, T.K.H., Mohajer, M.P., Yang, Z.J., Chang, A.M.Z., Sahota, D.S., 1995. The prediction of fetal acidosis at birth by computerised analysis of intrapartum cardiotocography. BJOG Int. J. Obstet. Gynaecol. 102, 454–460. https://doi.org/10.1111/j.1471-0528.1995.tb11317.x

Costa, A., Ayres-de-Campos, D., Costa, F., Santos, C., Bernardes, J., 2009. Prediction of neonatal acidemia by computer analysis of fetal heart rate and ST event signals. Am. J. Obstet. Gynecol. 201, 464.e1-464.e6. https://doi.org/10.1016/j.ajog.2009.04.033

Czabański, R., Jeżewski, J., Horoba, K., Jeżewski, M., 2013. Fetal state assessment using fuzzy analysis of fetal heart rate signals—Agreement with the neonatal outcome. Biocybern. Biomed. Eng. 33, 145–155. https://doi.org/10.1016/j.bbe.2013.07.003

Drummond, C., Holte, R.C., 2003. C4.5, Class Imbalance, and Cost Sensitivity: Why Under-Sampling beats Over-Sampling. Citeseerx 8.

Falbel, D., Allaire, J., Chollet, F., 2020. R Interface to "Keras" [WWW Document]. R Package Version 2300. URL https://CRAN.R-project.org/package=keras (accessed 6.12.20).

Gamage, J., Mathew, T., Weerahandi, S., 2004. Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem and MANOVA. J. Multivar. Anal. 88, 177–189. https://doi.org/10.1016/S0047-259X(03)00065-4

GraciaJacob, S., Geetha Ramani, R., 2012. Data Mining in Clinical Data Sets: A Review. Int. J. Appl. Inf. Syst. 4, 15–26. https://doi.org/10.5120/ijais12-450774

Grivell, R.M., Alfirevic, Z., Gyte, G.M., Devane, D., 2015. Antenatal cardiotocography for fetal assessment. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.CD007863.pub4

Hoo, K.A., Tvarlapati, K.J., Piovoso, M.J., Hajare, R., 2002. A method of robust multivariate outlier replacement. Comput. Chem. Eng. 26, 17–39. https://doi.org/10.1016/S0098-1354(01)00734-7

Huang, M.-L., Hsu, Y.-Y., 2012. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. J. Biomed. Sci. Eng. 05, 526–533. https://doi.org/10.4236/jbise.2012.59065

Ishtake, S.H., Sanap, S.A., 2013. Intelligent Heart Disease Prediction System Using Data Mining Techniques. Int. J Healthc. Biomed. Res. 1.

Kim, J.-H., 2009. Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Comput. Stat. Data Anal. 53, 3735–3745. https://doi.org/10.1016/j.csda.2009.04.009

Krupa, N., Ma, M., Zahedi, E., Ahmed, S., Hassan, F.M., 2011. Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine. Biomed. Eng. OnLine 10, 6. https://doi.org/10.1186/1475-925X-10-6

Kuhn, M., 2020. Classification and Regression Training [WWW Document]. R Package Version 60-86. URL https://CRAN.R-project.org/package=caret (accessed 6.12.20).

Kuhn, M., Weston, S., Culp, M., Coulter, N., 2020. C5.0 decision trees and rule-based models for pattern recognition [WWW Document]. R Package Version 0131 R Quinlan. URL https://CRAN.R-project.org/package=C50 (accessed 6.12.20).

Lee, A., Ulbricht, C., Dorffner, G., 1999. Application of artificial neural networks for detection of abnormal fetal heart rate pattern: a comparison with conventional algorithms. J. Obstet. Gynaecol. 19, 482–485. https://doi.org/10.1080/01443619964256

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest [WWW Document]. R News. URL http://CRAN.R-project.org/doc/Rnews/ (accessed 6.12.20).

Magenes, G., Pedrinazzi, L., Signorini, M.G., 2004. Identification of fetal sufferance antepartum through a multiparametric analysis and a support vector machine, Presented at the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, San Francisco, CA, USA, pp. 462–465. https://doi.org/10.1109/IEMBS.2004.1403194

Majka, M., 2020. High Performance Implementation of the Naive Bayes Algorithm in R. R Package Version 097.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2019. Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [WWW Document]. R Package Version 17-3. URL https://CRAN.R-project.org/package=e1071 (accessed 6.12.20).

Nidhal, S., Mohd. Ali, M.A., Zaidan, A.A., Zaidan, B.B., Najah, H., 2011. Computerized Algorithm for Fetal Heart Rate Baseline and Baseline Variability Estimation based on Distance Between Signal Average and a Value. Int. J. Pharmacol. 7, 228–237. https://doi.org/10.3923/ijp.2011.228.237

Nilsson, J., Ohlsson, M., Thulin, L., Höglund, P., Nashef, S.A.M., Brandt, J., 2006. Risk factor identification and mortality prediction in cardiac surgery using artificial neural networks. J. Thorac. Cardiovasc. Surg. 132, 12–19. https://doi.org/10.1016/j.jtcvs.2005.12.055

Ocak, H., Ertunc, H.M., 2013. Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems. Neural Comput. Appl. 23, 1583–1589. https://doi.org/10.1007/s00521-012-1110-3

Parer, J.T., King, T., 2000. Fetal heart rate monitoring: Is it salvageable? Am. J. Obstet. Gynecol. 182, 982–987. https://doi.org/10.1016/S0002-9378(00)70358-9

Sahin, H., Subasi, A., 2010. Classification of Fetal State from the Cardiotocogram Recordings using ANN and Simple Logistic. Presented at the ISSD, pp. 499–505.

Spencer, J.A.D., 1993. Clinical overview of cardiotocography. BJOG Int. J. Obstet. Gynaecol. 100, 4–7. https://doi.org/10.1111/j.1471-0528.1993.tb10626.x

Spitz, M.R., Hong, W.K., Amos, C.I., Wu, X., Schabath, M.B., Dong, Q., Shete, S., Etzel, C.J., 2007. A Risk Model for Prediction of Lung Cancer. JNCI J. Natl. Cancer Inst. 99, 715–726. https://doi.org/10.1093/jnci/djk153

Sundar, C., M.Chitradevi, M.C., Geetharamani, G., 2012. Classification of Cardiotocogram Data using Neural Network based Machine Learning Technique. Int. J. Comput. Appl. 47, 19–25. https://doi.org/10.5120/7256-0279

Thongkam, J., Xu, G., Zhang, Y., Huang, F., 2009. Toward breast cancer survivability prediction models through improving training space. Expert Syst. Appl. 36, 12200–12209. https://doi.org/10.1016/j.eswa.2009.04.067

Williams, B., Arulkumaran, S., 2004. Cardiotocography and medicolegal issues. Best Pract. Res. Clin. Obstet. Gynaecol. 18, 457–466. https://doi.org/10.1016/j.bpobgyn.2004.02.005

Wirth, R., Hipp, J., 2000. CRISP-DM: Towards a standard process model for data mining, in: Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining. Springer-Verlag, London, UK, pp. 29–39.

Yılmaz, E., Kılıkçıer, Ç., 2013. Determination of Fetal State from Cardiotocogram Using LS-SVM with Particle Swarm Optimization and Binary Decision Tree. Comput. Math. Methods Med. 2013, 1–8. https://doi.org/10.1155/2013/487179