P R O J E C T R E P O R T

ISYE 6414 – REGRESSION ANALYSIS

FETAL HEALTH CLASSIFICATION
TEAM NUMBER 10

AHMED RABBANI
ASHISH ARORA
NICOLE FLOWERHILL
VIVIANE FARNUNG

10 / DECEMBER / 2021

TABLE OF CONTENTS

1 INTRODUCTION

Cardiotocograph exams (CTG) measure changes in the fetal heart rate and draw conditional relationships with uterine contractions. The objective of these exams is to assess the fetus health by monitoring fetal heart rate (FHR), fetal movements, uterine contractions, and numerous other related factors. Results from these exams enable health care professionals to perform additional assessments to ensure fetus's wellbeing and determine if the baby needs to be delivered by caesarean section or instrumental vaginal birth.

The objective of our project is to analyse the features collected from multiple cardiotocograph exams and their corresponding results to gain an understanding of the critical factors that determine a fetus's health. The goal of this study is to use this descriptive model to enable healthcare professionals to better predict a fetus's health in a timely fashion and perform necessary actions to prevent child and maternal mortality.

We will be primarily working with Fetal Health data set acquired from Kaggle [1]. The data set includes 2126 training examples and 21 predictors. This data set was collected by examining features of 2126 cardiotocograph exams. The results of these exams were then classified by three expert obstetricians into three categories, briefly explained below:

- Normal: Features **UNLIKELY** to be associated with fetal compromise. No further action required.
- Suspect: Features **MAY** be associated with fetal compromise. Requires further action.
- Pathological: Features are **LIKELY** to be associated with fetal compromise. Immediate action required.

2 DESCRIPTION OF DATASET

The dataset consists of 21 variables and 2126 data points. To correctly use the data, the data needs to be analyzed according to its type.All the predictors were quantitative variables whereas our response variable was a categorical variable. Summary of predicting variables along with their representations in the dataset are shown in Table 1.

Predictors	Representation
Baseline.Value	Baseline Fetal Heart Rate
Accelerations	Number of accelerations per second
Fetal.Movement	Number of Fetal movements per second
Uterine.Contractions	Number of Uterine Contractions per second
Light.Decelerations	Number of light decelerations per second
Severe.Decelerations	Number of severe decelerations per second
Prolongued.Decel	Number of prolonged decelerations per second
Abnormal.SV	Percentage of time with abnormal short-term variability
Mean.SV	Mean value of short-term variability
Percent.LV	Percentage of time with abnormal long-term variability
Mean.LV	Mean value of long-term variability
His.width	Width of histogram
His.Min	Histogram minimum value
His.Max	Histogram maximum value
His.Peak	Number of peaks in histogram
His.Zeros	Number of zeroes in histogram
His.Mode	Mode of histogram
His.Mean	Mean of histogram
His.Median	Median of histogram
His.Var	Variance of histogram
His.Tend	Trend of histogram
Fetal_health	Health of Fetus

Table 1: Predictors in Original Dataset

As explained earlier, our response variable corresponds to fetus's health given the features of its cardiogram. Count of each category of response variable along with their representations in the data are shown in Table 2.

Category of Response Variable	Category Representation	Number of Observations
Normal	1	1655
Suspect	2	295
Pathological	3	176
Total number of observations		2126

Table 2: Response Variable in Original Dataset

3 METHODOLOGY

3.1 Multinomial Regression

The initial approach for this three-class problem was to use multinomial logistic regression. The Multinomial Logistic Regression Model deals with modelling the outcomes of a categorical-dependent variable with more than two categories and predicts the possibilities of different outcomes based on several independent variables [2]. The model estimates the effect of predictors on the probability of success in each $n-1$ binary logistic model for the category it represents respectively, in comparison to a reference category such that n is the number of categories of the response variable [3]. The model can be deployed to define the relationship between a categorical response variable and multiple explanatory variables and identify the effect of each variable.

After conducting multinomial regression in R on a training set and producing the evaluation metrics such as confusion matrix and misclassification error, it was noted that given the nature of the data set, a more extensive analysis beyond our work could not be conducted. Therefore, for further inspection a three - model binary logistic regression model approach was taken.

3.2 Binomial Logistic Regression

In this approach, we made three models for each class of response variable. For each model, the response variable was a binary class. For instance, for Model 1, the response variable classes were normal and not normal (adding the categories of suspect and pathological). For Model 2, response variable classes were suspect and not-suspect (adding the categories of normal and pathological). And for Model 3, response variable classes were pathological and not pathological (adding the categories of normal and suspect). With this approach, exploratory analysis, model fitting, variable selection, analysis and prediction were conducted. The results for the first model are shown in this report in a detailed manner. A similar analysis was conducted with Model 2 and Model 3 respectively. Results of Model 2 and Model 3 will be shown in a concise manner.

4 DATA EXPLORATION AND ANALYSIS - MULTINOMIAL REGRESSION

4.1 Model Fitting

Multinomial regression model with $K = 3$ was used at this step where K denotes the number of classes and π_{iK} denotes the probability of i^{th} case belonging to K^{th} class. This methodology is summarized in the equation below:

$$(Y = y) = \frac{n!}{y_i! \dots . y_K!} \pi_i^{y1} \dots . \pi_K^{yK}$$

Results of multinomial regression are shown in Table 3

Predictor	Coefficients	Column1	Standard Error	Column2
	Class 2	Class 3	Class 2	Class 3
(Intercept)	-17.059	-14.278	2.261	3.479
Baseline.Value	0.068	0.019	0.020	0.024
Accelerations	-735.675	-309.427	0.018	0.011
Fetal.Movement	10.052	15.081		3.720
Uterine.Contractions	-342.138	-325.323	0.044	0.031
Light.Decelerations	-77.269	128.511	0.014	0.054
Severe.Decelerations	-0.190	0.172	0.000	0.000
Prolongued.Decel	38.892	155.550	0.005	0.011
Abnormal.SV	0.042	0.199	0.012	0.024
Mean.SV	-0.618	0.656	0.329	0.305
Percent.LV	0.027	0.055	0.007	0.010
Mean.LV	-0.005	-0.060	0.032	0.059
His.Max	0.036	-0.028	0.012	0.015
His.Peak	-0.003	0.073	0.060	0.090
His.Zeros	-0.100	-0.283	0.169	0.322
His.Var	0.021	0.073	0.009	0.009
His.Tend	-0.088	-1.157	0.238	0.333

Table 3: Results of Multinomial Regression

The log of odds ratio in a multinomial logistic regression model can be estimated as an exponential function of regression coefficients, where A, B and C are classes of the response variable.

$$ln\left[\frac{pA}{pB}\right] = \beta_{0,1} + \beta_{1,1} \cdot X^{(1)} + \beta_{2,1} \cdot X^{(2)} + \dots . + \beta_{m,1} \cdot X^{(m)}$$

$$ln\left[\frac{pB}{pC}\right] = \beta_{0,2} + \beta_{1,2} \cdot X^{(1)} + \beta_{2,2} \cdot X^{(2)} + \dots . + \beta_{m,2} \cdot X^{(m)}$$

The analysis was conducted in two phases. In the first phase, we evaluated the classification performance of the model on the training data. Whereas, in the second phase we evaluated the prediction performance of the model on the test data. For evaluating the performance of the classification model, a confusion matrix was created using the training data set to identify the classification performance of the model for each class of response variable. Confusion matrix for training set is shown in Table 4.

Category	Normal	Suspect	Pathological
Class 1	965	56	6
Class 2	32	100	13
Class 3	7	13	85

Table 4: Confusion Matrix on Training Data

A misclassification error of 0.09945184 was calculated from the above confusion matrix. This value indicated that the proportion of misclassified observations is little or in other words, the model has an accuracy of around 90% i.e., the percentage of a correctly classified instance during classification. A similar analysis was done on the test data set. Confusion matrix for test set is shown in Table 5.

Category	Normal	Suspect	Pathological
Class 1	622	35	5
Class 2	23	86	14
Class 3	6	5	53

Table 5: Confusion Matrix for Test Data

A misclassification rate of 0.1036514 calculated from the above matrix, indicated that the proportion of misclassified observations is less. Based on the training dataset, the number of patients in each category were identified and their ratios were calculated for sensitivity analysis.

Since the misclassification rate of the overall model is approximately 0.104 i.e., 10%, the misclassification rate for each category of the response variable for both the testing and training dataset was computed. Results for this analysis are shown in Table 6 and 7 respectively.

Category	Normal	Suspect	Pathological
Class 1	96.1%	5.6%	0.6%
Class 2	18.9%	59.2%	7.7%
Class 3	6.7%	12.5%	81.7%

Table 6: Class wise Performance on Training Set

Category	Normal	Suspect	Pathological
Class 1	95.5%	5.4%	0.7%
Class 2	18.3%	68.3%	11.1%
Class 3	8.3%	69.5%	73.6%

Table 7: Class wise Performance on Test Set

The model used for training dataset is capable of performing 96% of classification on class 1, and 81.7% on class 3. The accuracy of the model falls to less than 60% for the case of suspected patients. The same model when used for testing dataset is capable of performing 95.5% classification on class 1 and 73.6% classification on class 3. The accuracy of model is slightly better at 68.25% for classification of class 2 i.e. the suspected patients.

5 DATA EXPLORATION AND ANALYSIS – LOGISTIC REGRESSION – MODEL 1

5.1 Exploratory Analysis

The dataset was initially inspected for null values, which were not present. After ensuring that the data set was complete, correlation analysis was done on the whole data to check for possible multicollinearity. Correlation matrix of the original data is shown in Figure 1. As per Figure 1, there appears to be strong correlation among the variables His Width, His Min, His Mode, His Mean and His Median. We decided to remove these variables from the dataset at this stage. As per literature, it is counterintuitive to remove these predictors in this manner, however, if we continue our analysis with the whole set of predictors then we would encounter Matrix Rank error in R during the model fitting. Checking for multicollinearity at that stage, will indicate high multicollinearity among the variables which are being removed at this stage.

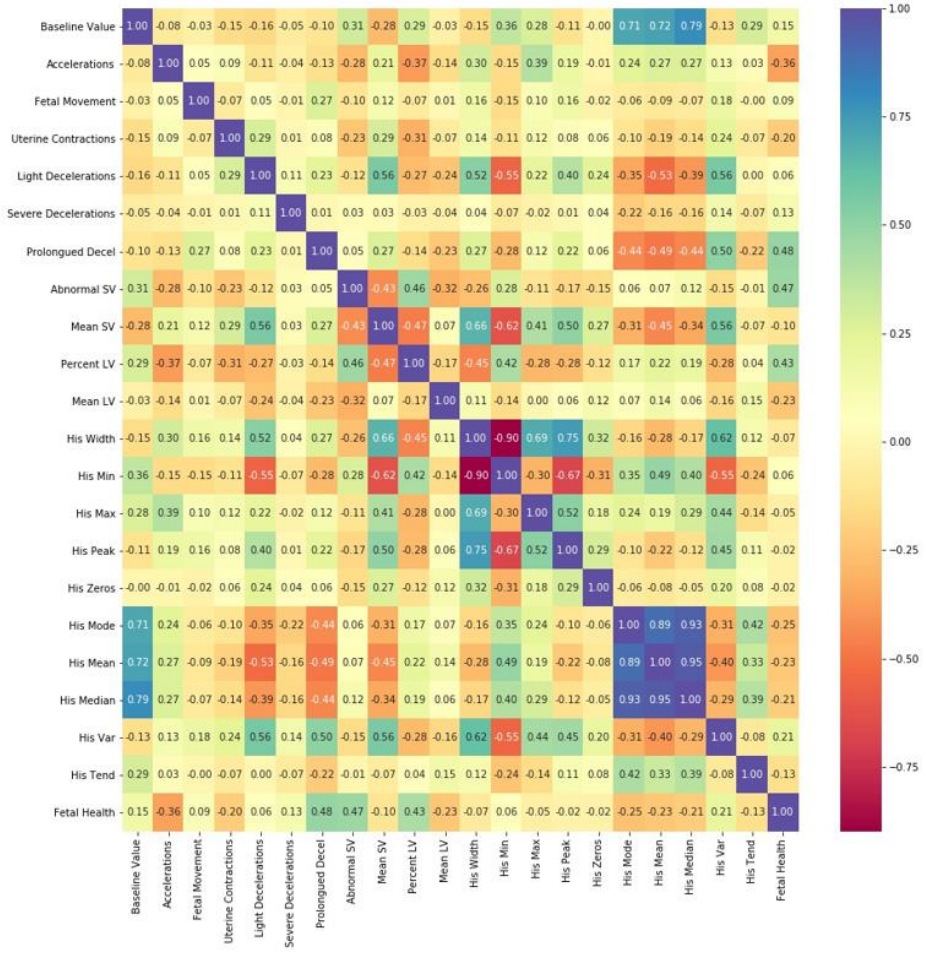

Figure 1: Correlation Matrix For Original Dataset

Correlation of predictors with respect to response variable in the modified dataset is shown in Figure 2.

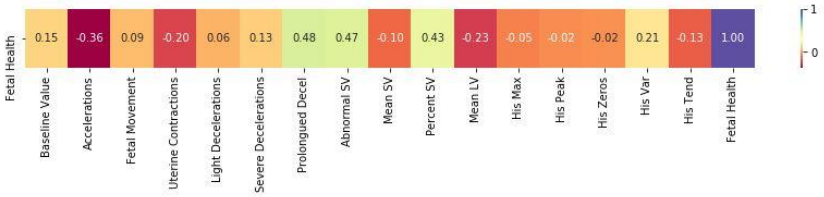

Figure 2: Correlation Between Predictors and Response - Modified Data Set

As per Figure 2, response variable Fetal health is independent of the predictors since no correlation is stronger than 0.48. To check if data transformation is necessary, the predicting variables were displayed in box plots and histograms. Box plots checked how the data was distributed and highlighted potential outliers. They also showed the distribution of binary response variable across each predictor. Histograms were used to see the individual distributions of each predictor. Since the values for different predictors use different data ranges, the data had to be normalized to the same range. If data is not normalized, a slight change in one variable output is more impactful on the result than another. For this dataset, all variables were scaled with respect to their mean and standard deviation. Figure 3 and Figure 4 show box plots and histograms of selected predictors after normalization.

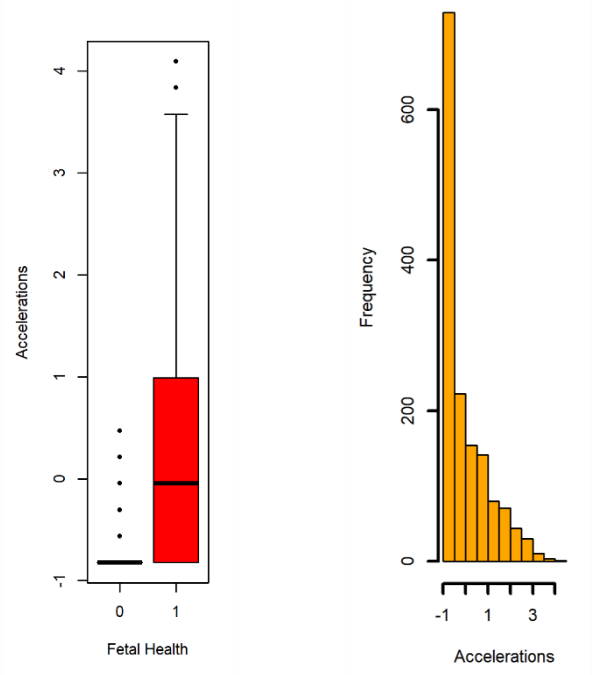

Figure 3: Histogram Box Plot of Acceleration After Normalization

The fetal health value of 0 indicates that a fetus is “Not Normal”, while a value of 1 indicates that a fetus is “Normal”. As per Figure 3, accelerations, values higher than –1 predict that a fetus might be healthy. The average value for a healthy fetus is around 0. Since most data points are of value –1 and the amount is slowly decreasing until 4.5, it can be assumed that a value higher than –1 is an indicator for a healthy fetus, marginally speaking. A value of –1, however, does not lead to the conclusion that a fetus is unhealthy.

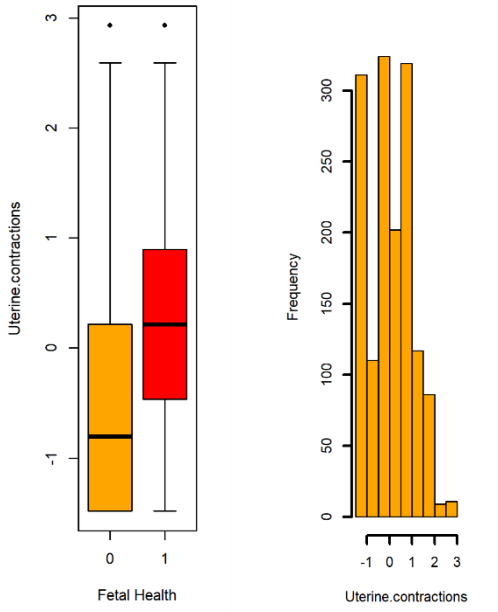

Figure 4: Box Plot of Uterine Contractions After Normalization

The boxplot in Figure 4 shows that a “Normal” fetus tends to have Uterine Contractions values between –0.5 and 1, while an “Not Normal” fetus tends to have Uterine Contractions values between –1.5 and 0.2. This means that there is an overlap between –0.5 and 0.2. As the histogram shows, there are many data points which have a value between –0.5 and 0 and fewer values between 0 and 0.5. Therefore, it can be assumed that the health of a fetus cannot be determined between –0.5 and 0, but values below that data range indicate an “Not Normal” fetus and values above a “Normal” one, marginally speaking.

In the last step of data preparation, the dataset was split into a training and test dataset, where the training dataset contains 70% and the test dataset contains 30% of the data.

5.2 Model Fitting

Our modified data set had 16 predictors and 1 response variable. At first, we fitted a model including all the predictors. Table 8 represents the coefficients of the predictors, as well as their respective standard error, z-value and p-value for the first model (normal and not-normal).

Predictor	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	4.174	0.372	11.233	0.000
Baseline.Value	-0.359	0.161	-2.238	0.025
Accelerations	3.553	0.511	6.960	0.000
Fetal.Movement	-0.246	0.129	-1.906	0.057
Uterine.Contractions	0.490	0.122	4.022	0.000
Light.Decelerations	0.272	0.198	1.373	0.170
Severe.Decelerations	-0.088	0.070	-1.255	0.209
Prolongued.Decel	-1.366	0.174	-7.852	0.000
Abnormal.SV	-1.432	0.170	-8.410	0.000
Mean.SV	0.930	0.233	3.996	0.000
Percent.LV	-0.582	0.124	-4.690	0.000
Mean.LV	-0.161	0.191	-0.842	0.400
His.Max	-0.724	0.179	-4.035	0.000
His.Peak	-0.035	0.155	-0.228	0.820
His.Zeros	-0.070	0.102	-0.683	0.495
His.Var	-1.082	0.218	-4.974	0.000
His.Tend	0.301	0.215	1.398	0.162

Table 8 Summary of Model 1

The estimated logit transformation of the probability of a fetus being normal given the predictors is represented by the following equation:

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) = {} & 4.17394 - 0.35940X_{baseline.value,i} + 3.55320_{Accelerations,i} \\ & - 0.24581_{Fetal.Movement,i} + 0.48997_{Uterine.Contractions,i} \\ & + 0.27190_{Light.Decelerations,i} - 0.08842_{Severe.Decelerations,i} \\ & - 1.36619_{Prolongued.Decel,i} - 1.43224_{Abnormal.SV,i} \\ & + 0.92992_{Mean.SV,i} - 0.58188_{Percent.SV,i} - 0.72359_{His.Max,i} \\ & - 0.03540_{His.Peak,i} - 0.06968_{His.Zeros,i} - 1.08245_{His.Var,i} \\ & + 0.30072_{His.Tend,i}\end{aligned}$$

For a one unit increase in the baseline fetal heart rate (baseline.value), the log-odds of the fetal health being normal decreases by 0.35940, holding all other variables constant. Each of the predictors in this model and equation can be interpreted in this fashion.

The ANOVA Chi-square test (similar to the partial F-test) was conducted for each predictor to see whether they add any explanatory power to the model The results for this analysis are shown in Table 9 below. Light Decelerations, Histogram number of peaks (His.Peak), histogram number of zeros (His.Zeros), and histogram trend/tendencies (His.Tend) all have high p-values, indicating they do not add explanatory power to the model, given that they are added to the model when all other predictors are already present in the model.

Predictor	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
Baseline.Value	1	98.115	1487.000	1477.120	0.000
Accelerations	1	354.983	1486.000	1122.137	0.000
Fetal.Movement	1	20.085	1485.000	1102.052	0.000
Uterine.Contractions	1	44.891	1484.000	1057.162	0.000
Light.Decelerations	1	0.863	1483.000	1056.299	0.353
Severe.Decelerations	1	10.429	1482.000	1045.869	0.001
Prolongued.Decel	1	223.606	1481.000	822.263	0.000
Abnormal.SV	1	158.279	1480.000	663.984	0.000
Mean.SV	1	4.434	1479.000	659.551	0.035
Percent.LV	1	28.329	1478.000	631.222	0.000
Mean.LV	1	5.830	1477.000	625.391	0.016
His.Max	1	28.335	1476.000	597.056	0.000
His.Peak	1	0.116	1475.000	596.940	0.734
His.Zeros	1	1.234	1474.000	595.706	0.267
His.Var	1	22.831	1473.000	572.875	0.000
His.Tend	1	1.962	1472.000	570.913	0.161

Table 9: Anova Test for Each Predictor - Model 1

5.3 Model Evaluation

Model 1 was checked for outliers using Cook's distance, as displayed below in Figure 5. Since all of the Cook's distances are approximately below 0.1, Model 1 does not demonstrate outliers when using a threshold of 1. This was also demonstrated by Figure 6 which demonstrates the standard residuals of Model 1 for both categories, visually indicating no outliers.

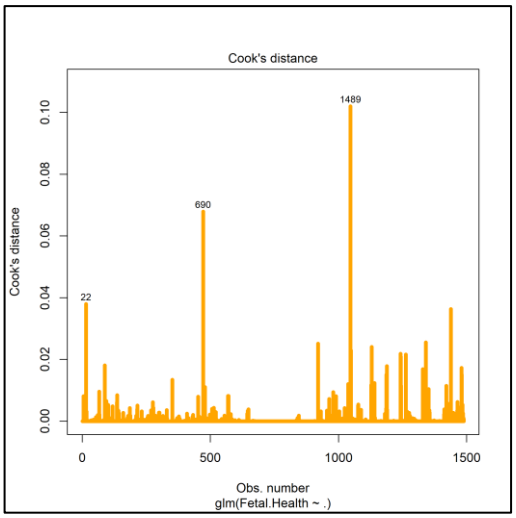

Figure 3: Cook's Distance - Model 1

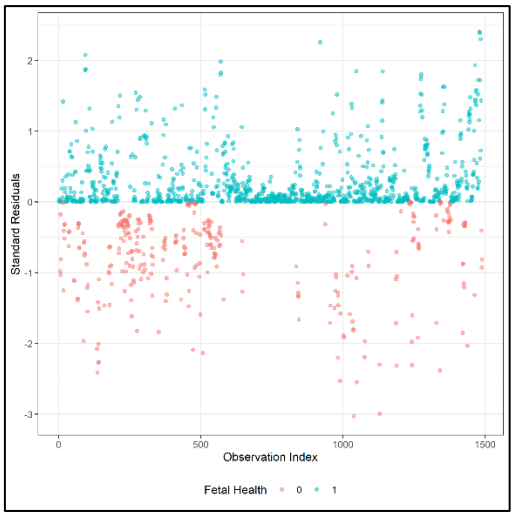

Figure 4: Standardized Residuals - Model 1

To diagnose whether there is multicollinearity between the predictors, the variance inflation factor (VIF) was computed for each predicting variable. As shown in Table 10, since all of the VIFs are below 10, the variables do not indicate multicollinearity.

Predictor	VIF
Baseline.Value	2.179
Accelerations	1.785
Fetal.Movement	1.231
Uterine.Contractions	1.216
Light.Decelerations	3.727
Severe.Decelerations	1.190
Prolongued.Decel	2.090
Abnormal.SV	1.394
Mean.SV	4.611
Percent.LV	1.750
Mean.LV	2.190
His.Max	3.062
His.Peak	2.309
His.Zeros	1.384
His.Var	3.233
His.Tend	1.684

Table 10: VIF values for Model 1

6 MODEL SELECTION

6.1 Normal and Not Normal – Model 1

To reduce high dimensionality and create a simpler model without compromising explanatory power, three different variable selection methods were used. In the first method, forward-backward stepwise regression was used which performed variable selected based on AIC criteria. This method selected all variables from the full model except for Mean of Long-Term Variability (Mean.LV), Histogram Number of Peaks (His.Peak), Histogram Number of Zeros (His.Zeros), and Histogram Tendency/Trends (His.Tend) predictors. Second method used the Lasso regression model, which selected all the predictors from the full model except for Histogram Number of Peaks (His.Peak) predictor. Lastly, Elastic Net model was used for variable selection, however this model selected all of the variables of the full model, which is understandable given that it uses a combination of L1 and L2 penalties. For both elastic net and lasso, optimal value of lambda was calculated using cross validation. Summary of variable selection is shown in Table 11.

Predictor	Full Model	Stepwise Model	Lasso Model	Elastic Net Model
Baseline.Value	X	X	X	X
Accelerations	X	X	X	X
Fetal.Movement	X	X	X	X
Uterine.Contractions	X	X	X	X
Light.Decelerations	X	X	X	X
Severe.Decelerations	X	X	X	X
Prolongued.Decel	X	X	X	X
Abnormal.SV	X	X	X	X
Mean.SV	X	X	X	X
Percent.LV	X	X	X	X
Mean.LV	X		X	X
His.Max	X	X	X	X
His.Peak	X			X
His.Zeros	X		X	X
His.Var	X	X	X	X
His.Tend	X		X	X

Table 11: Summary of Variable Selection

Earlier, the threshold probability value for recognising the positive class (Normal Fetus) was set as 0.5. In order to evaluate the effect of this threshold value on the model performance metrics, the models obtained from various variable selection methods were evaluated using different performance metrics on the test set. Results of this analysis are shown in Table 12.

Model 1 - Normal				
Metric	Full Model	Stepwise Model	Lasso Model	Elastic Net Model
0.3 Threshold				
Accuracy	90.1%	90.0%	90.1%	90.9%
Sensitivity	84.8%	84.7%	84.2%	90.3%
Specificity	91.2%	91.1%	91.4%	91.0%
0.5 Threshold				
Accuracy	90.0%	90.0%	90.0%	89.8%
Sensitivity	75.5%	74.8%	75.5%	75.3%
Specificity	94.4%	94.8%	94.4%	94.3%
0.7 Threshold				
Accuracy	88.4%	88.5%	88.4%	88.4%
Sensitivity	67.9%	68.3%	67.9%	67.9%
Specificity	96.9%	96.9%	96.9%	96.9%

Table 12: Probability Threshold Value Effect on Model Performance

As per Table 12, as the probability threshold value increases, there is an increase in specificity and decrease in sensitivity. This implies that the model is able to identify “Not Normal Fetus” more correctly than “Normal Fetus” as the threshold value increases.

As per Model 1, the positive class is a “Normal Fetus” whereas the negative class is a “Not Normal Fetus”. In the context of the problem, it is very important to correctly identify a “Not Normal Fetus” so that corrective medical measures can be immediately taken. However, it is also important to correctly identify a “Normal Fetus” to ensure that no such swift medical action is taken that may jeopardize the fetus’s health. Therefore, for Model 1 both specificity and sensitivity are important however, specificity is slightly more important than sensitivity. Therefore, we decide to choose Stepwise Regression Model 1 with 0.5 probability threshold value since it depicts high sensitivity and specificity scores. The chosen model is also highlighted in cyan in Table 12.

6.2 Suspect and Not Suspect – Model 2

All the analysis which has been documented until now was repeated for Suspect and Not Suspect case. Table 13 shows the model selection stage for Model 2.

Model 2 - Suspect				
Metric	**Full Model**	**Stepwise Model**	**Lasso Model**	**Elastic Net Model**
0.3 Threshold				
Accuracy	88.4%	88.5%	88.2%	87.9%
Sensitivity	96.8%	96.9%	96.8%	96.6%
Specificity	55.4%	55.8%	55.0%	54.2%
0.5 Threshold				
Accuracy	87.6%	87.9%	87.3%	87.4%
Sensitivity	92.0%	92.3%	91.8%	91.7%
Specificity	55.8%	57.0%	54.5%	55.4%
0.7 Threshold				
Accuracy	87.0%	87.4%	87.1%	87.1%
Sensitivity	88.6%	88.9%	88.7%	88.2%
Specificity	57.6%	61.8%	58.8%	61.5%

Table 13: Model Selection for Model 2

As per Model 2, the positive class is a “Suspect Fetus” whereas the negative class is a “Not Suspect Fetus”. In the context of the problem, it is very important to correctly identify a “Suspect Fetus” so that corrective medical measures can be immediately taken. Therefore, for Model 2,sensitivity is an important criterion. Therefore, we decided to choose Stepwise Regression Model 2 with 0.3 probability threshold value since it depicts the highest sensitivity score. The chosen model is also highlighted in cyan in Table 13.

6.3 Pathological and Not Pathological – Model 3

All the analysis which has been documented until now was repeated for Pathological and Not Pathological case. Table 14 shows the model selection stage for Model 3.

Model 3 - Pathological			
Metric	**Full Model**	**Stepwise Model**	**Elastic Net Model**
0.3 Threshold			
Accuracy	95.8%	95.8%	96.2%
Sensitivity	98.8%	98.8%	99.0%
Specificity	69.2%	69.2%	71.9%
0.5 Threshold			
Accuracy	97.0%	96.7%	97.0%
Sensitivity	98.3%	98.0%	98.0%
Specificity	82.4%	81.6%	85.1%
0.7 Threshold			
Accuracy	96.2%	96.2%	96.2%
Sensitivity	97.0%	96.8%	96.8%
Specificity	85.0%	86.8%	86.8%

Table 14: Model Selection for Model 3

As per Model 3, the positive class is a “Pathological Fetus” whereas the negative class is a “Not Pathological Fetus”. In the context of the problem, it is extremely critical to correctly identify a “Pathological Fetus” so that corrective medical measures can be immediately taken. Therefore, for Model 3 sensitivity is an important criterion. Therefore, we decided to choose Elastic Net Model 3 with 0.3 probability threshold value since it depicts the highest sensitivity score. The chosen model is also highlighted in cyan in Table 14.

For pathological case, since the number of observations belonging to “Pathological Fetus” were so low, the fitted lasso model gave an warning indicating that the probabilities of one ore more observations are indistinguishable from 0 or 1. Ideally, we should increase sample size or remove outliers to address this warning. We did the latter, but it did not change the error. Therefore, the performance metrics for the lasso model were discarded from Table 14.

7 MEDICAL INTERPRETATION AND MODEL DEPLOYMENT

Summary of selected model for each class of response variable along with their chosen predictors and their estimated coefficients are shown in Table 15:

Feature	Model 1 - Normal	Model 2 - Normal	Model 3 - Pathological
Baseline.Value	-0.256	0.363	-0.437
Accelerations	3.663	-3.292	-2.734
Fetal.Movement	-0.245	X	0.394
Uterine.Contractions	0.493	-0.466	-0.225
Light.Decelerations	0.349	-0.858	0.900
Severe.Decelerations	-0.100	X	0.199
Prolongued.Decel	-1.340	X	1.636
Abnormal.SV	-1.410	0.380	2.874
Mean.SV	0.886	-0.657	0.200
Percent.LV	-0.549	X	0.995
Mean.LV	X	-0.339	0.566
His.Max	-0.836	0.895	0.150
His.Peak	X	0.276	-1.188
His.Zeros	X	X	0.186
His.Var	-1.062	X	1.319
His.Tend	X	0.432	-0.256

Table 15: Summary of All Models

In the context of the problem, certain features such as baseline heart rate value, decelerations and abnormal short-term variability can be very critical in determining a fetus's health. For instance, a lower baseline heart rate, higher decelerations, higher abnormal short-term variability and higher prolonged deceleration imply that a fetus is not healthy. Whereas, higher accelerations and lower abnormal short term variability imply that a fetus is healthy.

The above models can be deployed as such that that they can be used in conjunction to make a classification prediction on a new observation. For instance, for a new observation that is actually a Normal Fetus, Model 1 should classify it as a "Normal Fetus" whereas Model 2 and Model 3 should classify it as "Not Suspect Fetus" and "Not Pathological Fetus" respectively. Predictions for "Suspect Fetus" and "Pathological Fetus" can be made in a similar manner. Model predictions for certain test observations are shown in Table 16:

Model Deployment					
S.no	Index	True Label	Model 1	Model 2	Model 3
1	257	Normal	Normal	Not Suspect	Not Pathological
2	516	Normal	Normal	Not Suspect	Not Pathological
3	398	Suspect	Not Normal	Suspect	Not Pathological
4	138	Suspect	Not Normal	Suspect	Not Pathological
5	751	Pathological	Not Normal	Suspect	Pathological
6	1490	Pathological	Not Normal	Not Suspect	Pathological

Table 16: Model Predictions on New Observations

From Table 16, it appears that all models predicted the correct class given the true label of a new observation, with the exception of Model 2's prediction on observation number 5. This observation is actually a pathological case and model 2 is recognising it as a suspect case. We can opt for another model 2 from the model selection table that can recognise this case as the "Not Suspect Fetus" but that model is also likely to recognise an actual suspect case as "Not Suspect Case". The consequences of the latter possibility are much higher than the current scenario of recognizing a "pathological case" as a "suspect case". In short, it would be far better to recognise a "pathological case" as a suspect case than to recognise a "suspect case" as not a suspect case.

8 CONCLUSION

In this study we used a data set to predict a fetus's health based on the features of its cardiotocograph exams. We started our analysis with a multinomial logistic regression which is followed by a binomial logistic regression for each class. During exploratory analysis, we removed certain predictors with high correlations to avoid multicollinearity in our fitted model. After fitting an initial full model we checked for outliers and multicollinearity in the model. We performed variable selection by using stepwise regression, elastic net regression and lasso regression. Then we evaluated the effect of changing the probability threshold based on the model performance as per different metrics. After choosing models based on the criteria suitable for the class of each response variable, we used all three models to make predictions on a new set of observations. Our models can be used by medical professionals to assess a fetus's health and perform corrective measures to ensure the fetus's wellbeing. This can in turn help health care professionals to prevent child or maternal mortality.

One of the major challenges in this data set is that the class of response variables were not evenly distributed. Most observations belonged to the "Normal Fetus". Having a data set with equal number of classes can help to better recognise the pathological cases, which were heavily outnumbered in this data set. Secondly, in actual conditions, the interpretation of fetus's CTC varies with respect to the baby's trimester and labor's stage. If multiple data sets can be collected for different set of these features, then different models can be fitted to assess a fetus's health based on its trimester and labor.

9 REFERENCES

[1] Larxel. (2020, October 12). Fetal Health Classification. Kaggle. Retrieved December 9, 2021, from https://www.kaggle.com/andrewmvd/fetal-health-classification.

[2] Home. IDRE Stats. (n.d.). Retrieved December 9, 2021, from https://stats.idre.ucla.edu/r/dae/multinomial-logistic-regression/.

[3] Mohamad, N. A., Ali, Z., Noor, N. M., & Baharum, A. (2016). Multinomial logistic regression modelling of stress level among secondary school teachers in Kubang Pasu District, Kedah. https://doi.org/10.1063/1.4954623