Prediction of Neonatal's Heart Condition Using Cardiotocographic Observations

RAJESH RANJAN

Introduction

To predict the newborn health condition using various heart related observations obtained via Cardiotocography recordings.

Scope of Study:

A.   Conduct Data Mining applied science to estimate the patient's condition (1-Normal, 2-Suspected, 3-Pathological) based on the obtained heart related vital records.

Analysis Methodology (using R):

A.   Decision Tree analysis-

B.   Multinomial Regression Analysis-

C.   Liner Discriminant Analysis-

Dataset- Cardiotocography

Cardiotocography (CTG) –a graphic recording of fetal heart rate and uterine contractions for the assessment of Infant's Heart condition.
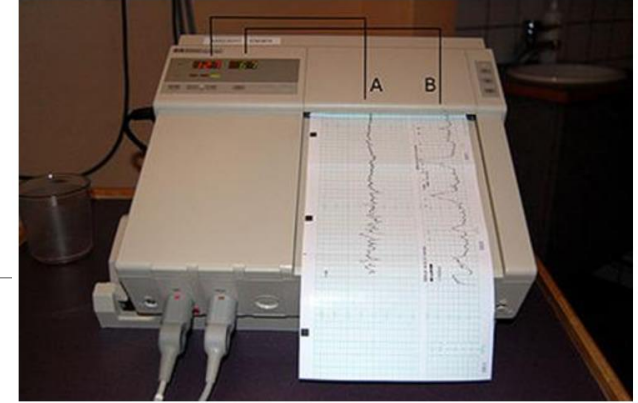
Fig. 1. Cardiotocography
A – curve of fetal heart rate (HR)
B – curve representing alterations in the uterine muscle tone

Vital parameters for the study:
Independent Variable:
A. LB (FHR base Line per minute)
B. AC- Acceleration per second
C. FM- Fetal Movement per second

Categorical Variable
NSP- Fetal State (N- Normal, S-Suspected, P- Pathological)

Attribute Information:

LB - FHR baseline (beats per minute)
AC - # of accelerations per second
FM - # of fetal movements per second
UC - # of uterine contractions per second
DL - # of light decelerations per second
DS - # of severe decelerations per second
DP - # of prolongued decelerations per second
ASTV - percentage of time with abnormal short term variability
MSTV - mean value of short term variability
ALTV - percentage of time with abnormal long term variability
MLTV - mean value of long term variability
Width - width of FHR histogram
Min - minimum of FHR histogram
Max - Maximum of FHR histogram
Nmax - # of histogram peaks
Nzeros - # of histogram zeros
Mode - histogram mode
Mean - histogram mean
Median - histogram median
Variance - histogram variance
Tendency - histogram tendency
CLASS - FHR pattern class code (1 to 10)
NSP - fetal state class code (N=normal; S=suspect; P=pathologic)

Data Preparation

- 2126 Observations, 22 parameters
- Convert Numerical categorical variable to Factor type
- Data Slicing [Training, Test] with 80:20 rule
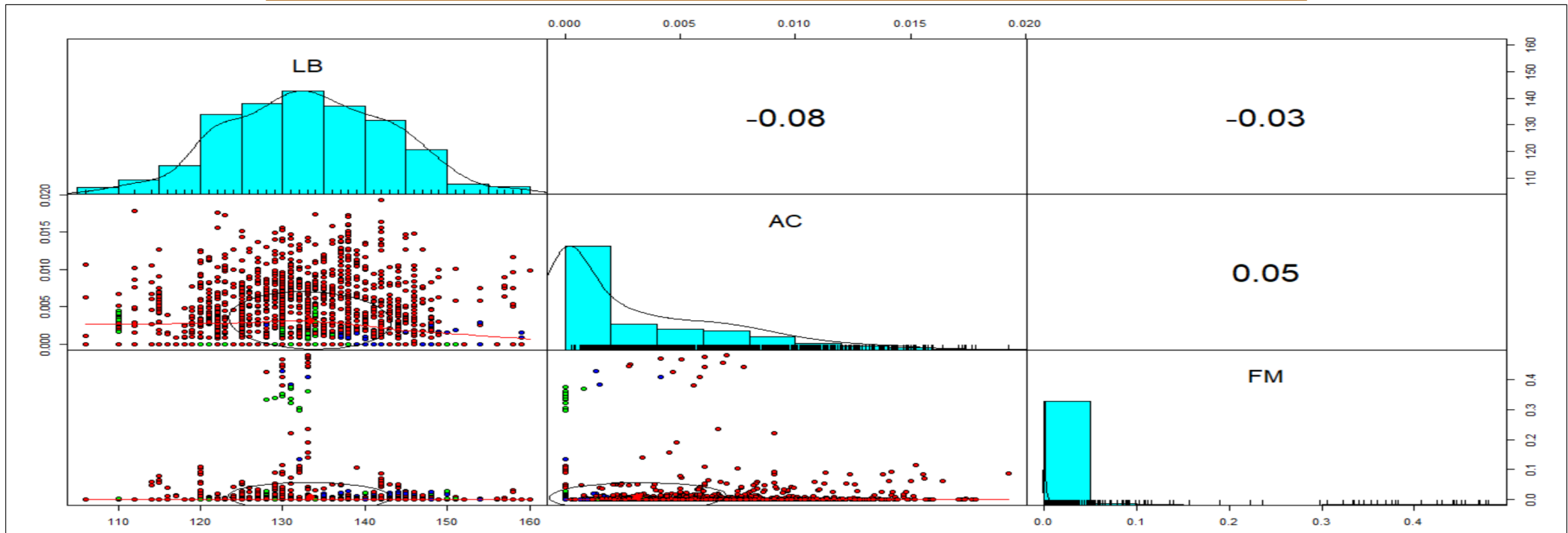
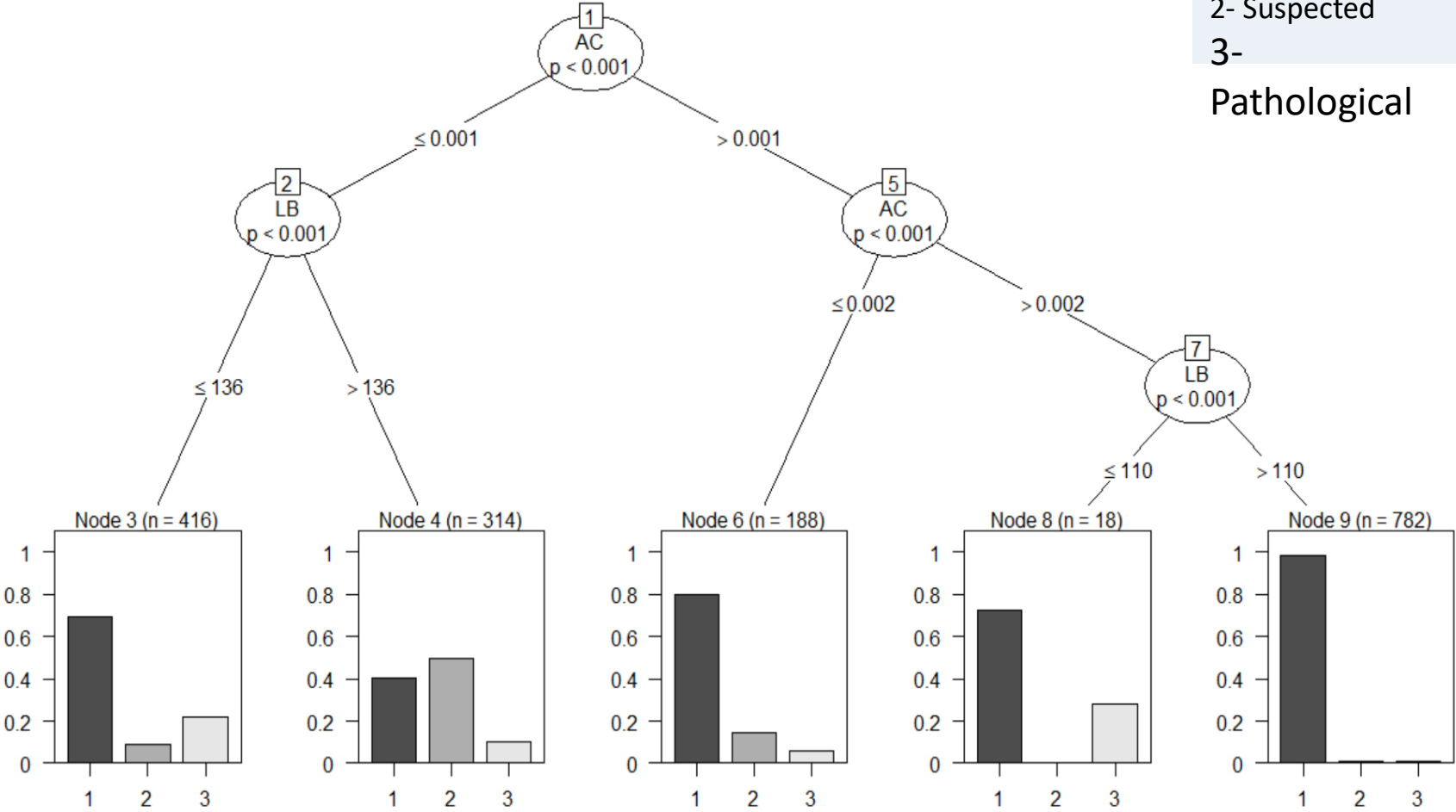# First Analysis - Decision Tree on Training Dataset

Plot Zoom

Infant's health prediction using decision tree

1- Normal
2- Suspected
3- Pathological

Conditional inference tree with 5 terminal nodes

Response: NSP_Factor
Inputs: LB, AC, FM
Number of observations: 1718

```
1) AC <= 0.000834028; criterion = 1, statistic = 263.403
  2) LB <= 136; criterion = 1, statistic = 131.511
    3)*  weights = 416
  2) LB > 136
    4)*  weights = 314
1) AC > 0.000834028
  5) AC <= 0.002209945; criterion = 1, statistic = 52.155
    6)*  weights = 188
  5) AC > 0.002209945
    7) LB <= 110; criterion = 1, statistic = 18.889
      8)*  weights = 18
    7) LB > 110
      9)*  weights = 782
```

First Analysis- Decision Tree on Training Dataset cont....

Normal   Suspected   Pathological

```
  [1] 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2
 [49] 2 2 2 1 2 1 1 2 1 1 1 2 1 1 1 1 2 2 1 1 1 1 2 1 1 1 2 2 1 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1
 [97] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[145] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[193] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
[241] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 1 1 1 2 2 2 2 2 2 2 2 1 1 1 2 2 2 2 2 2 2 2 2 2
[289] 2 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2 2 2 2
[337] 2 1 1 2 2 1 1 2 2 2 2 1 1 1 1 1 1 2 2 2 2 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2
[385] 2 2 2 2 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
[433] 2 2 1 2 2 2 1 2 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2
[481] 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 2 1 1 1 1 2 2 2 2 1 1 2 2 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2
[529] 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1
[577] 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2 2 2 2 1 1 1 1 1 1 1
```

First Analysis- Decision Tree on Training Dataset cont....

Infant's health calculation prediction count of Training dataset
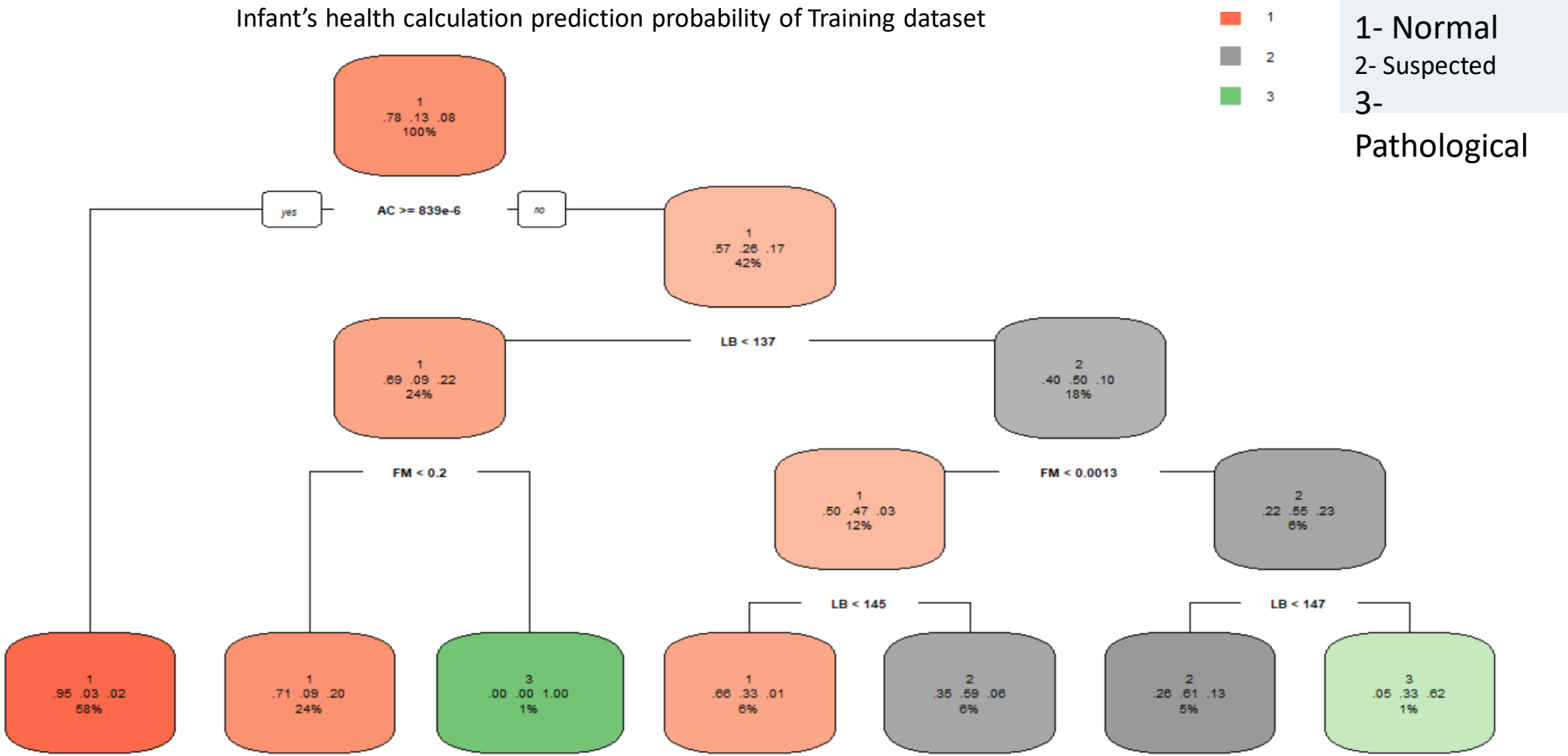

1- Normal
2- Suspected
3- Pathological

First Analysis- Decision Tree on Training Dataset cont....

Infant's health calculation prediction probability of Training dataset

1- Normal
2- Suspected
3- Pathological
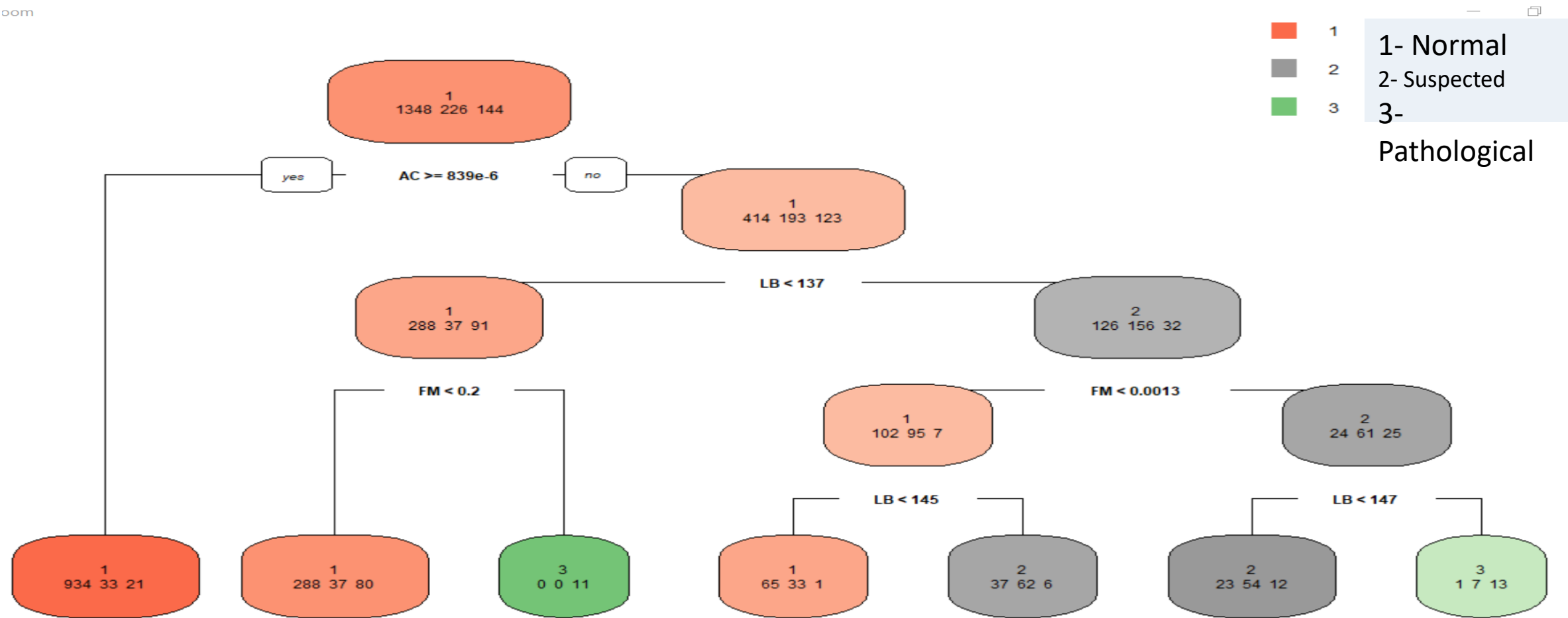
First Analysis- Decision Tree on TEST Dataset

	1	2	3
5	0.94534413	0.03340081	0.02125506
14	0.94534413	0.03340081	0.02125506
16	0.94534413	0.03340081	0.02125506
26	0.71111111	0.09135802	0.19753086
28	0.71111111	0.09135802	0.19753086
29	0.71111111	0.09135802	0.19753086
39	0.94534413	0.03340081	0.02125506
40	0.94534413	0.03340081	0.02125506
60	0.94534413	0.03340081	0.02125506
61	0.94534413	0.03340081	0.02125506
72	0.94534413	0.03340081	0.02125506
81	0.25842697	0.60674157	0.13483146
86	0.94534413	0.03340081	0.02125506
90	0.94534413	0.03340081	0.02125506
92	0.35238095	0.59047619	0.05714286
113	0.94534413	0.03340081	0.02125506
116	0.71111111	0.09135802	0.19753086
117	0.71111111	0.09135802	0.19753086

Decision Tree Model Evaluation

Training Dataset

Actual

Prediction	1	2	3
1	1222	70	112
2	126	156	32
3	0	0	0

Standard Error-
(19.75%)

Test Dataset

Actual

Prediction	1	2	3
1	274	21	28
2	33	48	4
3	0	0	0

Standard Error-
(21.07%)
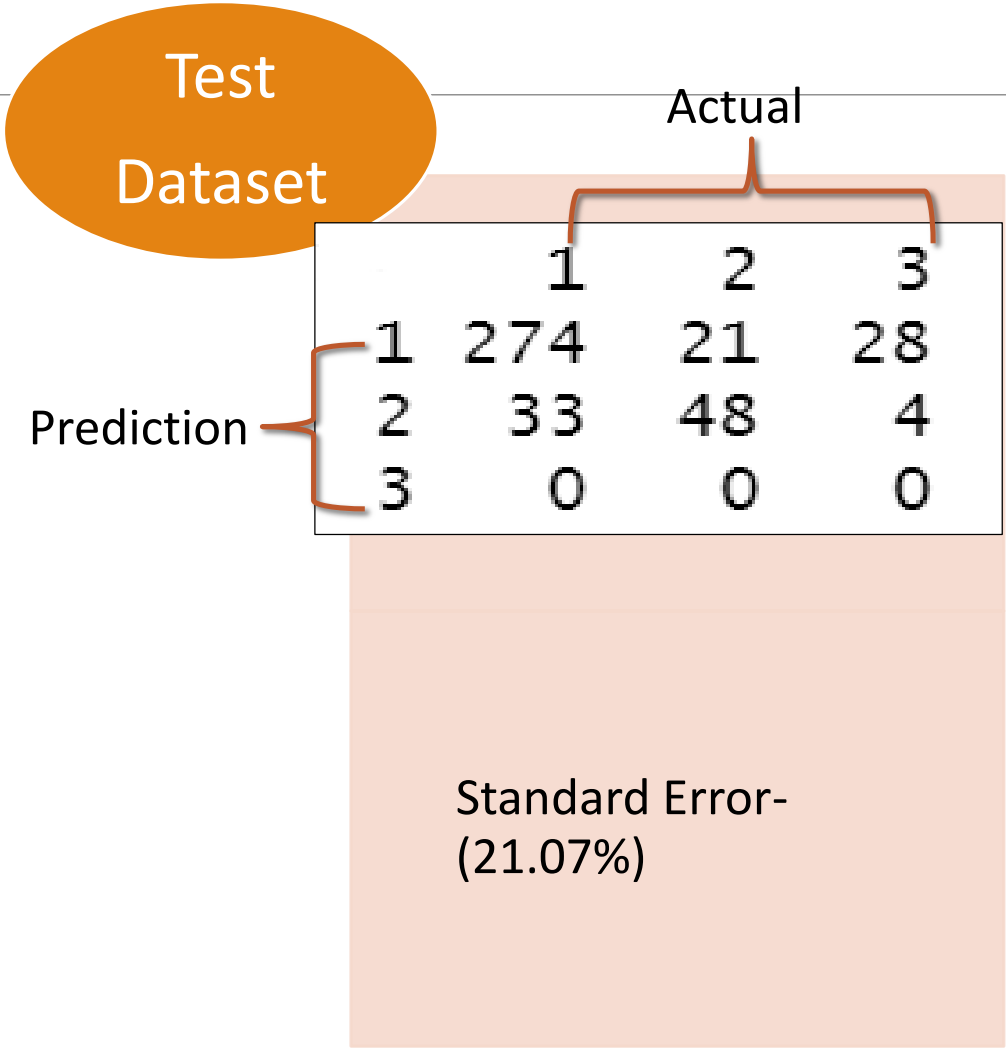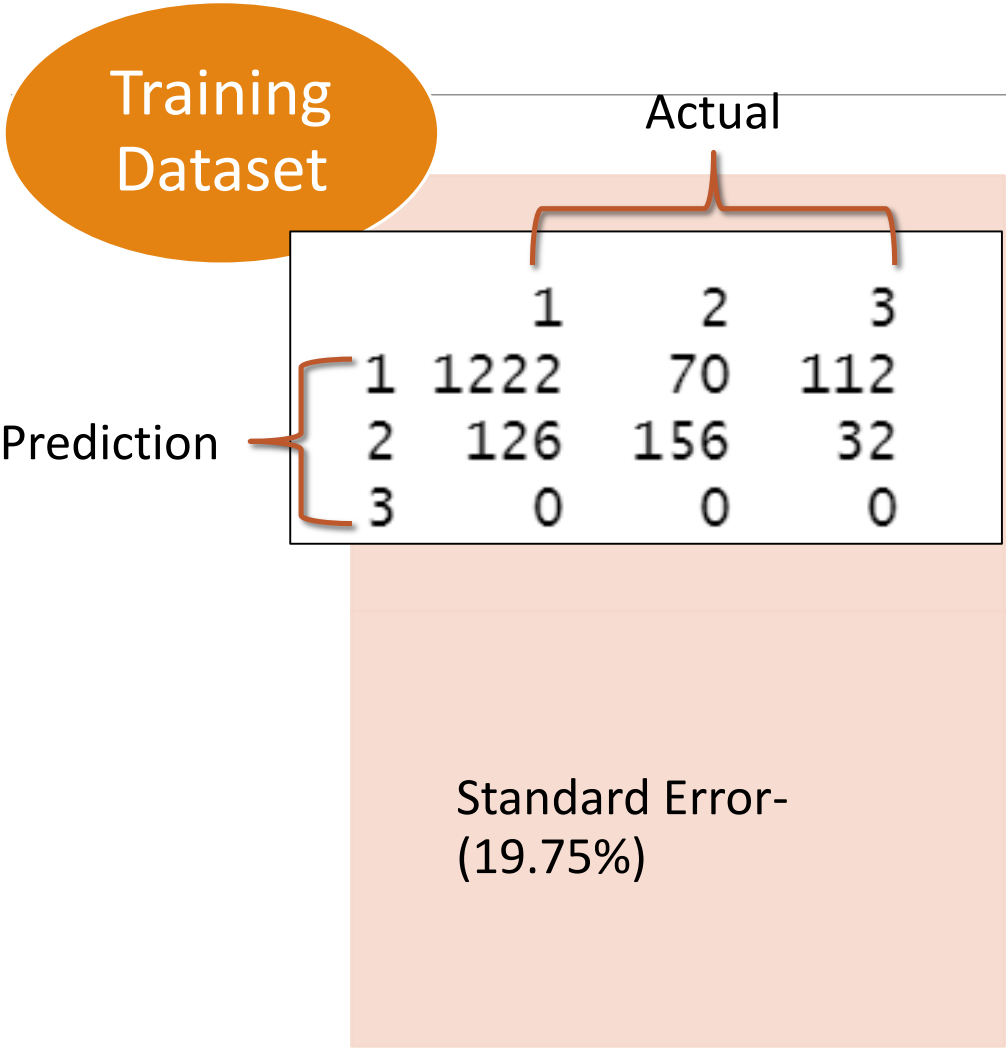
Second Analysis- Multinomial Regression on Training

Second Analysis- Multinomial Regression on Test Dataset

Multinomial Regression Model Evaluation

2- Suspected
3- Pathological

```
Std. Errors:
   (Intercept)           LB              AC          FM
2     1.261066  0.009050217  0.005518354  1.746013
3     1.212057  0.009170115  0.009872315  1.372880
```

LB- FHR baseline (beats per minute)

AC- # of accelerations per second

FM- # of fetal movements per second

| 2- | ~99.9% |
| 3- | ~99.9% |

| 2- | ~95% |
| 3- | ~92% |

Actual

Prediction

```
          1        2        3
1  1592      165      137
2     61      128       27
3      2        2       12
```
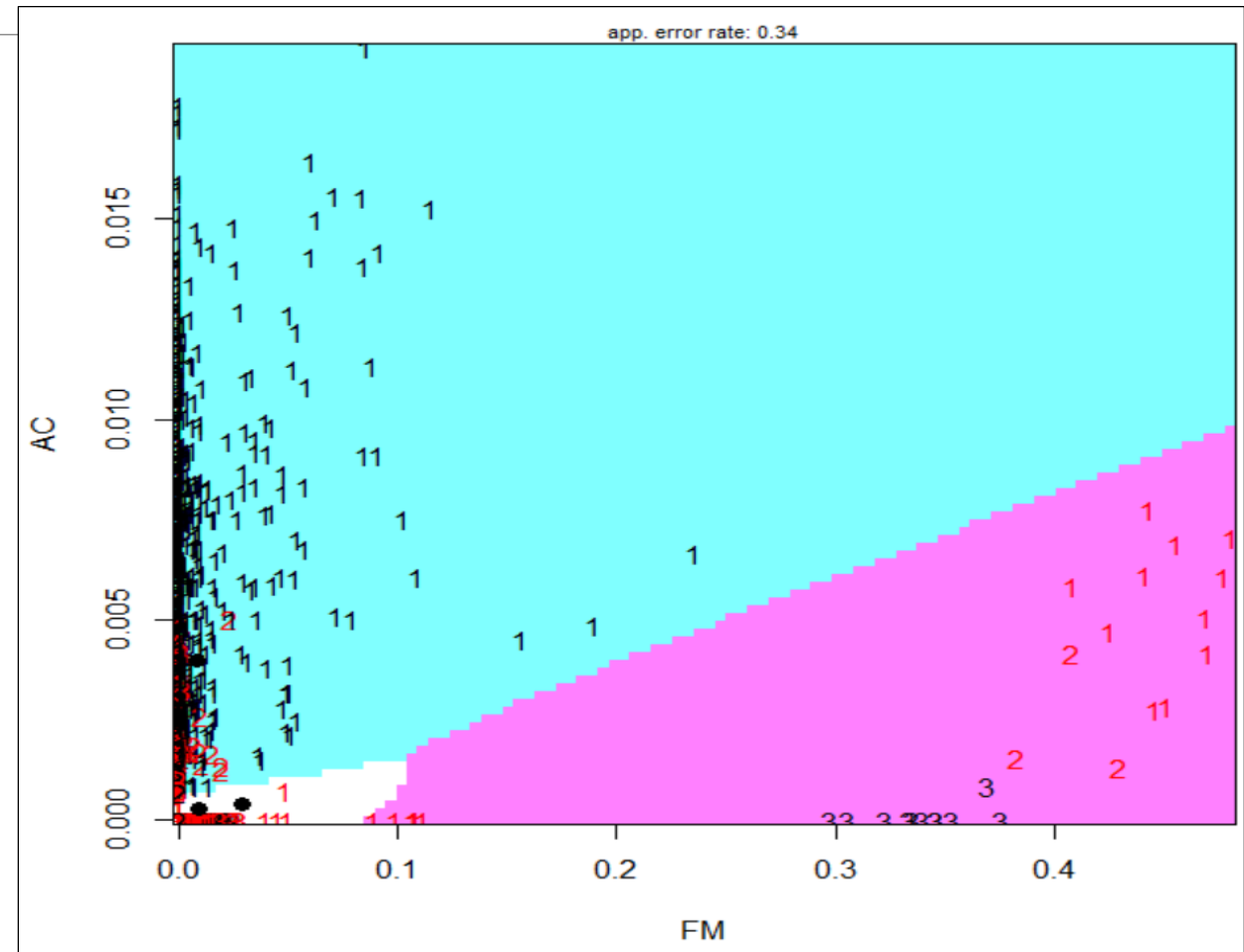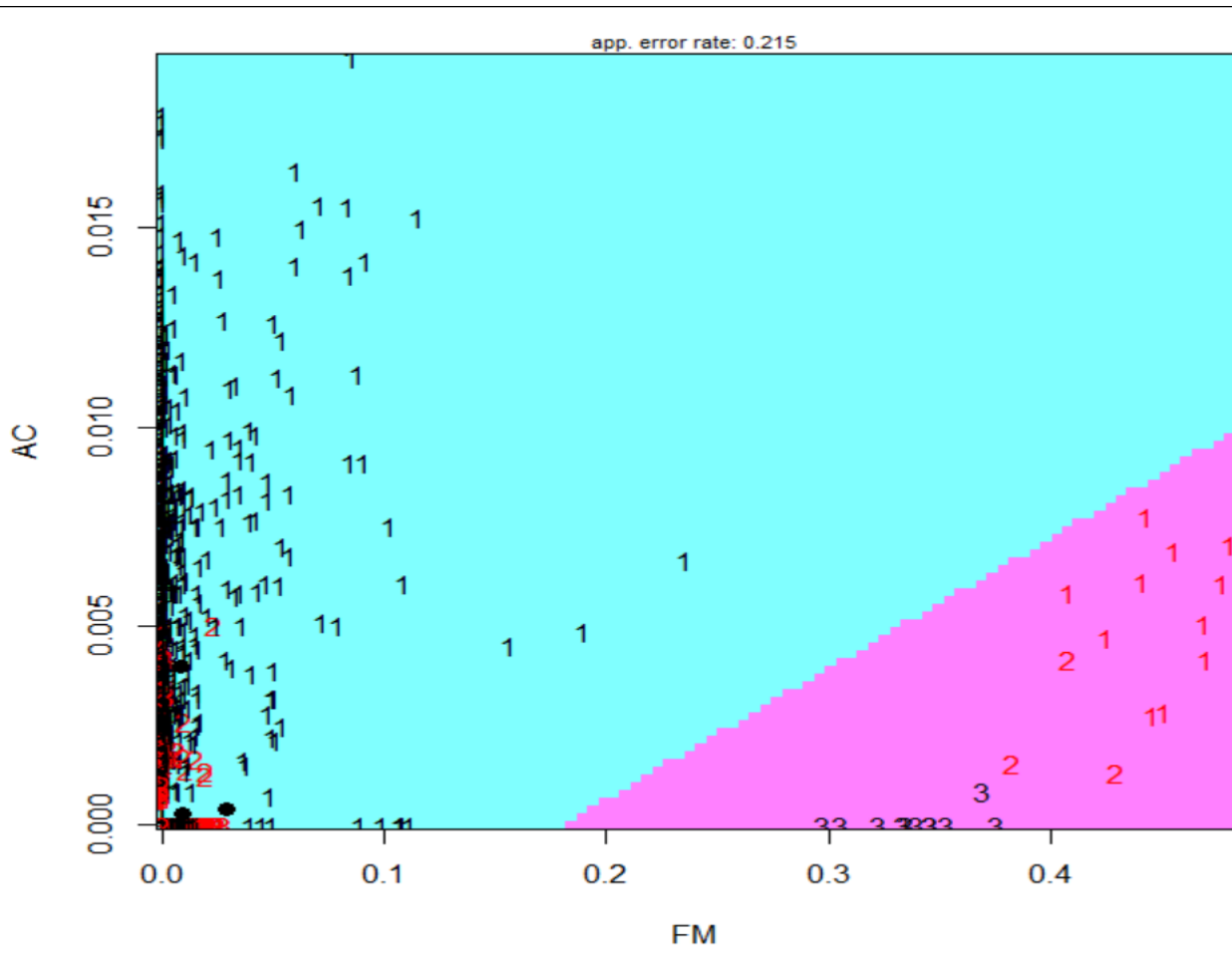
Standard Error- 18.53%

Accuracy Rate- 81.47%

Third Analysis- Liner Discriminant Analysis

Linear Separation vs Quadratic Plot

LDA Model Evaluation

Training Dataset

```
            actual
predicted     1      2      3
        1  1311    169    115
        2    26     54     18
        3    11      3     11
```

Accuracy~80%

Test Dataset

```
            actual
predicted     1      2      3
        1   295     55     28
        2    11     14      3
        3     1      0      1
```

Accuracy~ 76%

Conclusion

Decision Tree model observed as ~80% accurate and precise in predicting the infant's heart condition based on several executions of model on training and test dataset.

The independent variables (LB and AC) used in multinomial regression analysis calculated with ~99% confidence while performing the foresaid regression analysis.

LDA model is evaluated with 80% accurate on train and 76% on test dataset however the plot could not clearly classify the categories for patients.