Supervised Learning with Conditional Inference Tree

Cardiotocograms, also known as CTGs, have been instrumental within clinical medicine for a long time. Obstetricians use these measurements and classifications to obtain detailed information and intelligence about newborns and their mother prior and during labor. In 2018, an article presented through the Journal of Clinical Medicine detailed the practicality of CTGs. The same article noted that interpretations of these sensorial readings is mainly attributed to the observer; which creates challenges of consistency of interpretations and defies the human naked-eye. Questions like what happens if /when the interpreter misses a key detail, or what could be the meaning of a combination of diagnostic signals, furthermore, what time-sensitive conditions may the measurements expose, requiring immediate actions are few examples of concerns posed by the continuous practice of merely optical assessments of a CTG (Zhao, Zhang & Deng, 2018).

The following exploration presents an assessment of CTGs using the conditional inference tree (ctree) model. The same shows how the algorithm expedites and enhances the interpretation of CTG readings while appraising multiple fetal readings simultaneously. Moreover, the study aims to identify potential hidden patters which may require further attention.

The dataframe to be analyzed comes for the UCI Machine Learning Repository, and it consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) as identified and recorded by cardiotocograms. The multivariate dataframe contains 2126 observations and 23 variables. Each diagnostic attribute within these CTGs were automatically processed and measured. Ultimately, all CTGs were classified by three subject matter experts, and under unanimity, assigned with response-labels based on the fetal state and/or morphologic detected patterns (Dua, & Graff, 2019).

The following list, obtained from the data dictionary, explains the meaning of the involved codes across these readings.

LB - FHR baseline (beats per minute)
AC - Number of accelerations per second
FM - Number of fetal movements per second
UC - Number of uterine contractions per second
DL - Number of light decelerations per second
DS - Number of severe decelerations per second
DP - Number of prolonged decelerations per second
ASTV - Percentage of time with abnormal short-term variability
MSTV - Mean value of short term variability
ALTV - Percentage of time with abnormal long-term variability
MLTV - Mean value of long -erm variability
Width - Width of FHR histogram
Min - Minimum of FHR histogram
Max - Maximum of FHR histogram
Nmax - Number of histogram peaks
Nzeros - Number 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

As observed, the above list includes unique CTG measurements, statistical attributes as well as observations from some of the recorded variables. The last two variables, CLASS and NSP, represent the previously mentioned classification and response-labeling conducted by the obstetricians.

Exploratory Analysis. The given cardiotocography.csv file was loaded and vectored as `ctg`. A look into the structure of the dataframe confirms some of the variables and information obtained from the repository and .csv file itself. The dataframe contains 2126 observations and 23

variables, formatted as integers and numeric values, as illustrated on figure 1.0. A glimpse over these represented values highlight few transformation options. Case in point, the targeted variable, which is the NSP , will need to get converted to a factor. Others like FM, DP, or ALTV may be representative of asymmetrical distributions. Furthermore, a variable like DS could have only one time of value, making it unusable for classification purposes.

```
> str(ctg)
'data.frame':   2126 obs. of  23 variables:
 $ LB      : int  120 132 133 134 132 134 134 122 122 122 ...
 $ AC      : num  0 0.006 0.003 0.003 0.007 0.001 0.001 0 0 0 .
 $ FM      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ UC      : num  0 0.006 0.008 0.008 0.008 0.01 0.013 0 0.002 
 $ DL      : num  0 0.003 0.003 0.003 0 0.009 0.008 0 0 0 ...
 $ DS      : num  0 0 0 0 0 0 0 0 0 0 ...
 $ DP      : num  0 0 0 0 0 0.002 0.003 0 0 0 ...
 $ ASTV    : int  73 17 16 16 16 26 29 83 84 86 ...
 $ MSTV    : num  0.5 2.1 2.1 2.4 2.4 5.9 6.3 0.5 0.5 0.3 ...
 $ ALTV    : int  43 0 0 0 0 0 0 6 5 6 ...
 $ MLTV    : num  2.4 10.4 13.4 23 19.9 0 0 15.6 13.6 10.6 ...
 $ Width   : int  64 130 130 117 117 150 150 68 68 68 ...
 $ Min     : int  62 68 68 53 53 50 50 62 62 62 ...
 $ Max     : int  126 198 198 170 170 200 200 130 130 130 ...
 $ Nmax    : int  2 6 5 11 9 5 6 0 0 1 ...
 $ Nzeros  : int  0 1 1 0 0 3 3 0 0 0 ...
 $ Mode    : int  120 141 141 137 137 76 71 122 122 122 ...
 $ Mean    : int  137 136 135 134 136 107 107 122 122 122 ...
 $ Median  : int  121 140 138 137 138 107 106 123 123 123 ...
 $ Variance: int  73 12 13 13 11 170 215 3 3 1 ...
 $ Tendency: int  1 0 0 1 1 0 0 1 1 1 ...
 $ CLASS   : int  9 6 6 6 2 8 8 9 9 9 ...
 $ NSP     : int  2 1 1 1 1 3 3 3 3 3 ...
> 
```

Figure 1.0 – Possible irregular distributions across the variables.

Followed the dataframe structure assessment, the `summary(ctg)` command was used to expand and corroborates other aspects within the data.

```
> summary(ctg)
       LB               AC                FM                UC                DL                DS          
 Min.   :106.0   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000e+00  
 1st Qu.:126.0   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.002000   1st Qu.:0.000000   1st Qu.:0.000e+00  
 Median :133.0   Median :0.002000   Median :0.000000   Median :0.004000   Median :0.000000   Median :0.000e+00  
 Mean   :133.3   Mean   :0.003178   Mean   :0.009481   Mean   :0.004366   Mean   :0.001889   Mean   :3.293e-06  
 3rd Qu.:140.0   3rd Qu.:0.006000   3rd Qu.:0.003000   3rd Qu.:0.007000   3rd Qu.:0.003000   3rd Qu.:0.000e+00  
 Max.   :160.0   Max.   :0.019000   Max.   :0.481000   Max.   :0.015000   Max.   :0.015000   Max.   :1.000e-03  
       DP                ASTV            MSTV             ALTV             MLTV            Width       
 Min.   :0.0000000   Min.   :12.00   Min.   :0.200   Min.   : 0.000   Min.   : 0.000   Min.   :  3.00  
 1st Qu.:0.0000000   1st Qu.:32.00   1st Qu.:0.700   1st Qu.: 0.000   1st Qu.: 4.600   1st Qu.: 37.00  
 Median :0.0000000   Median :49.00   Median :1.200   Median : 0.000   Median : 7.400   Median : 67.50  
 Mean   :0.0001585   Mean   :46.99   Mean   :1.333   Mean   : 9.847   Mean   : 8.188   Mean   : 70.45  
 3rd Qu.:0.0000000   3rd Qu.:61.00   3rd Qu.:1.700   3rd Qu.:11.000   3rd Qu.:10.800   3rd Qu.:100.00  
 Max.   :0.0050000   Max.   :87.00   Max.   :7.000   Max.   :91.000   Max.   :50.700   Max.   :180.00  
      Min              Max            Nmax             Nzeros            Mode             Mean           Median      
 Min.   : 50.00   Min.   :122   Min.   : 0.000   Min.   : 0.0000   Min.   : 60.0   Min.   : 73.0   Min.   : 77.0  
 1st Qu.: 67.00   1st Qu.:152   1st Qu.: 2.000   1st Qu.: 0.0000   1st Qu.:129.0   1st Qu.:125.0   1st Qu.:129.0  
 Median : 93.00   Median :162   Median : 3.000   Median : 0.0000   Median :139.0   Median :136.0   Median :139.0  
 Mean   : 93.58   Mean   :164   Mean   : 4.068   Mean   : 0.3236   Mean   :137.5   Mean   :134.6   Mean   :138.1  
 3rd Qu.:120.00   3rd Qu.:174   3rd Qu.: 6.000   3rd Qu.: 0.0000   3rd Qu.:148.0   3rd Qu.:145.0   3rd Qu.:148.0  
 Max.   :159.00   Max.   :238   Max.   :18.000   Max.   :10.0000   Max.   :187.0   Max.   :182.0   Max.   :186.0  
    Variance         Tendency          CLASS            NSP        
 Min.   :  0.00   Min.   :-1.0000   Min.   : 1.00   Min.   :1.000  
 1st Qu.:  2.00   1st Qu.: 0.0000   1st Qu.: 2.00   1st Qu.:1.000  
 Median :  7.00   Median : 0.0000   Median : 4.00   Median :1.000  
 Mean   : 18.81   Mean   : 0.3203   Mean   : 4.51   Mean   :1.304  
 3rd Qu.: 24.00   3rd Qu.: 1.0000   3rd Qu.: 7.00   3rd Qu.:1.000  
 Max.   :269.00   Max.   : 1.0000   Max.   :10.00   Max.   :3.000  
> 
```

Figure 1.1 – NSP, as the response variable, will be converted to a factor

Case in point, judging by the numbers, Width, Min, Max, Nmax, Nzeros, Mode, Mean, Median, Variance, and Tendency appear to be statistical results of a measurement. Also, the variance across distributions is evident, thus, extra steps will have to be taken during the pre-processing phase. Lastly, no NA values were represented to this point, and using the `>colSums(is.na(ctg)` confirms there are no missing values.

```
> colSums(is.na(ctg))
      LB       AC       FM       UC       DL       DS       DP     ASTV     MSTV     ALTV     MLTV    Width      Min 
       0        0        0        0        0        0        0        0        0        0        0        0        0 
     Max     Nmax   Nzeros     Mode     Mean   Median Variance Tendency    CLASS      NSP 
       0        0        0        0        0        0        0        0        0        0 
> 
```

Figure 1.2 – No identified missing values across the dataset.

After reviewing the statistical attributes of the data set, a histogram was built to represent the LB variable. Per the data dictionary, this particular attribute is the most influential attribute of the set. The image in figure 1.3 shows how equally distributed the LB attribute is, and by the featured frequency, the mean will most likely fall between 130-135 beats per minute.

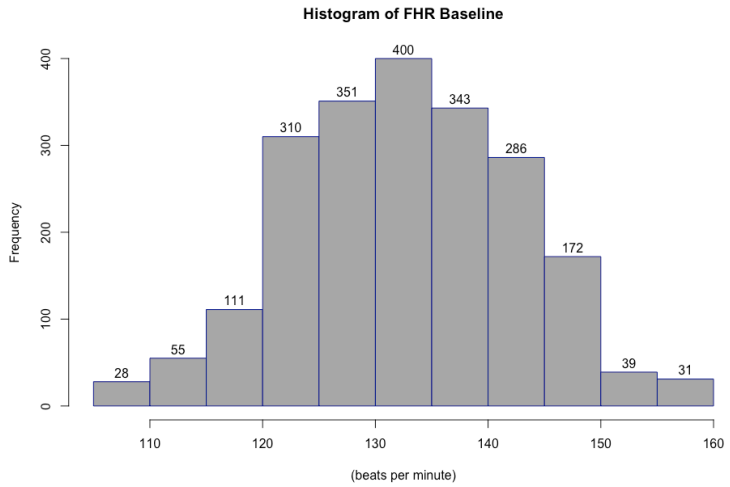


Figure 1.3 – Fetal heart hate baseline (LB variable) fairly distributed.

Further exploration of the LB variable (see figure 1.4) underscores some readings across the measurement particularly distanced from the mean. See the Appendix section for figure 1.4 code.

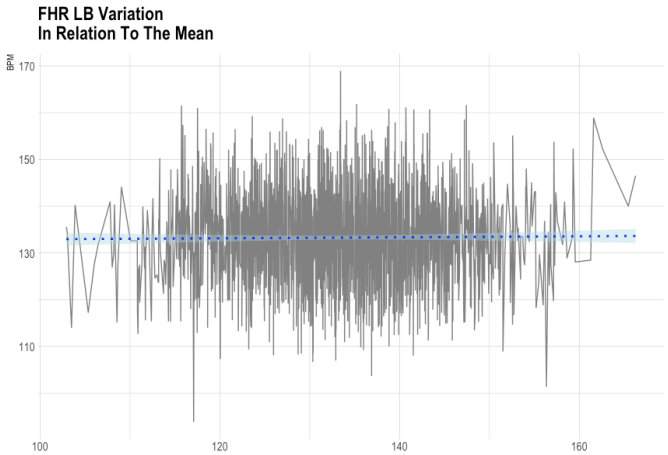


Figure 1.4 – Baseline variation across the dataset.

Motivated by the variability of the line graph, a deeper look was taken to identify out of the represented observations which ones were outside the 2-standard deviation (s.d.) ranges. Figure 1.5, portrays what the developed code (see appendix A2) captured. It shows the values associated to 1 or 2 s.d.s' boundaries, and how some of the readings exceed such limits.

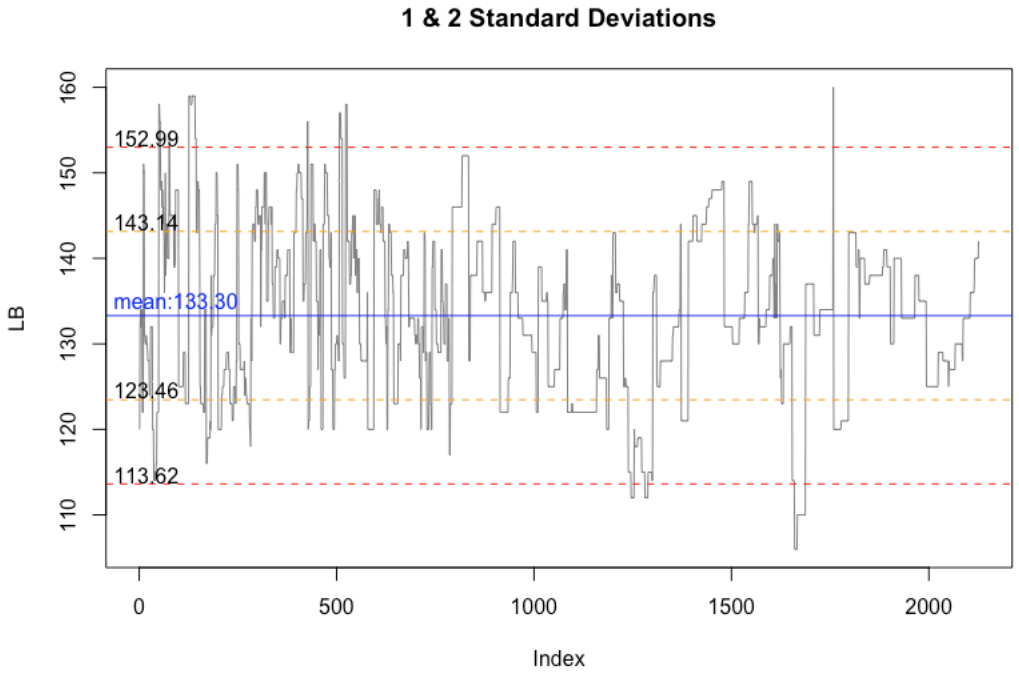


Figure 1.5 – Graphical display of LB's 1 & 2 standard deviations from the mean

Further calculation, like the one below, shows how critical this reading resulted, given that those readings within 2 s.d. account for over 96% of the studied observations. That said, this does not imply the contained readings are one way or the other (normal, suspect, or pathologic), but instead shows the usability of these quantified observations.

```
> # LB Observations higher than 2-s.d.
>  lba<-(sum(ctg$LB >152.99)) #39
> # LB Observations lower than 2-s.d.
>  lbb<-(sum(ctg$LB <113.62)) #44
> lba+lbb #=83 obs outside of 2-s.d.
[1] 83
> sum(between(ctg$LB, 113.62, 152.99))/nrow(ctg) # of obs within 2-s.d.
[1] 0.9609595
>
```

Figure 1.6 – Observations within 2 s.d. of the LB variable account for 96% of the dataset.

Considering the findings through the EDA process, and explanatory details from the data source, variables that were not part of the original measurements were excluded. Specifically, these variables were: width, min, max, nmax, nzeros, mode, mean, median, variance, and tendency. After further inquiry, these characteristics and calculations were related to the FHR baseline histogram made by the analytical software used by the dataset originators (Zhao, Zhang & Deng, 2018). Initially, the CLASS variable was converted to a factor and all of its 10 levels renamed. Unfortunately, this variable appears to be also a classification by the medical professionals, and contributed to overfitting of the model, thus forcing to reassess the modeling approach. That said, the CLASS variable was used as a guide for interpretation, but not included on the final set of variables to test within the model. Besides, the NSP variable was transformed to a factor, and its levels renamed to *normal*, *suspect*, and *pathologic*.

```
> # b. Exclude non-original measurements, rename targeted values
> ctg[12:22] <- NULL
> ctg$NSP<-as.numeric(ctg$NSP)
> ctg$NSP<-factor(ctg$NSP, levels= 1:3, labels = c("Normal","Suspect", "Pathologic"))
> # ctg$CLASS<-as.numeric(ctg$CLASS)
> # ctg$CLASS<-factor(ctg$CLASS, levels=1:10, labels= c('A', 'B', 'C', 'D', 'SH', 'AD', 'DE', 'LD', 'FS', 'SP'))
```

Figure 1.7 – Transformation steps prior to creating a model.

After transforming the mentioned variables, a graphical representation of the target variable was created. Figure 1.8 displays a basic barplot and tabulation of the number of observations, including the NSP renamed values and their respective counts.

```
# Visualization of original NSP
plot(ctg$NSP, main="Original NSP Distribution",
     xlab="Fetal State Classification",
     ylab="Frequency", col=c(3, 7, 2))
text(ctg$NSP, labels=as.character(tabulate(ctg$NSP)), adj=3, pos=3)
```

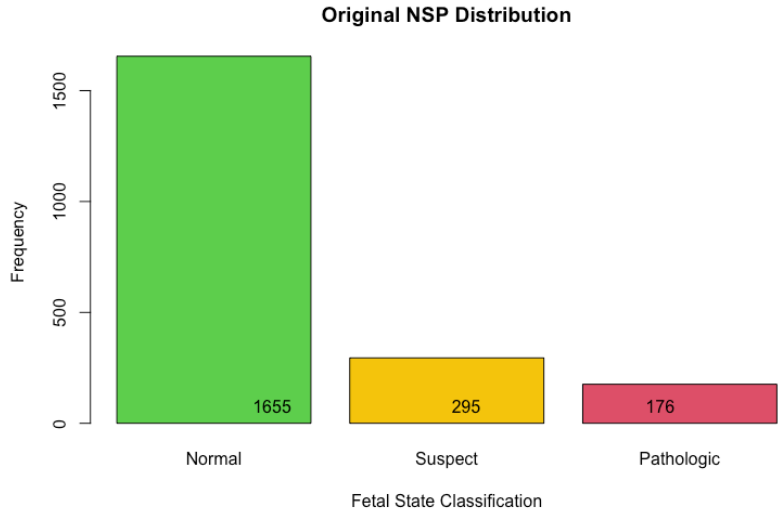


Figure 1.8 – Original NSP labeled observations.

At this point, the dataset was better organized and ready for the model portion. Bellow there is a representation of the final selected variables and their respective distributions.

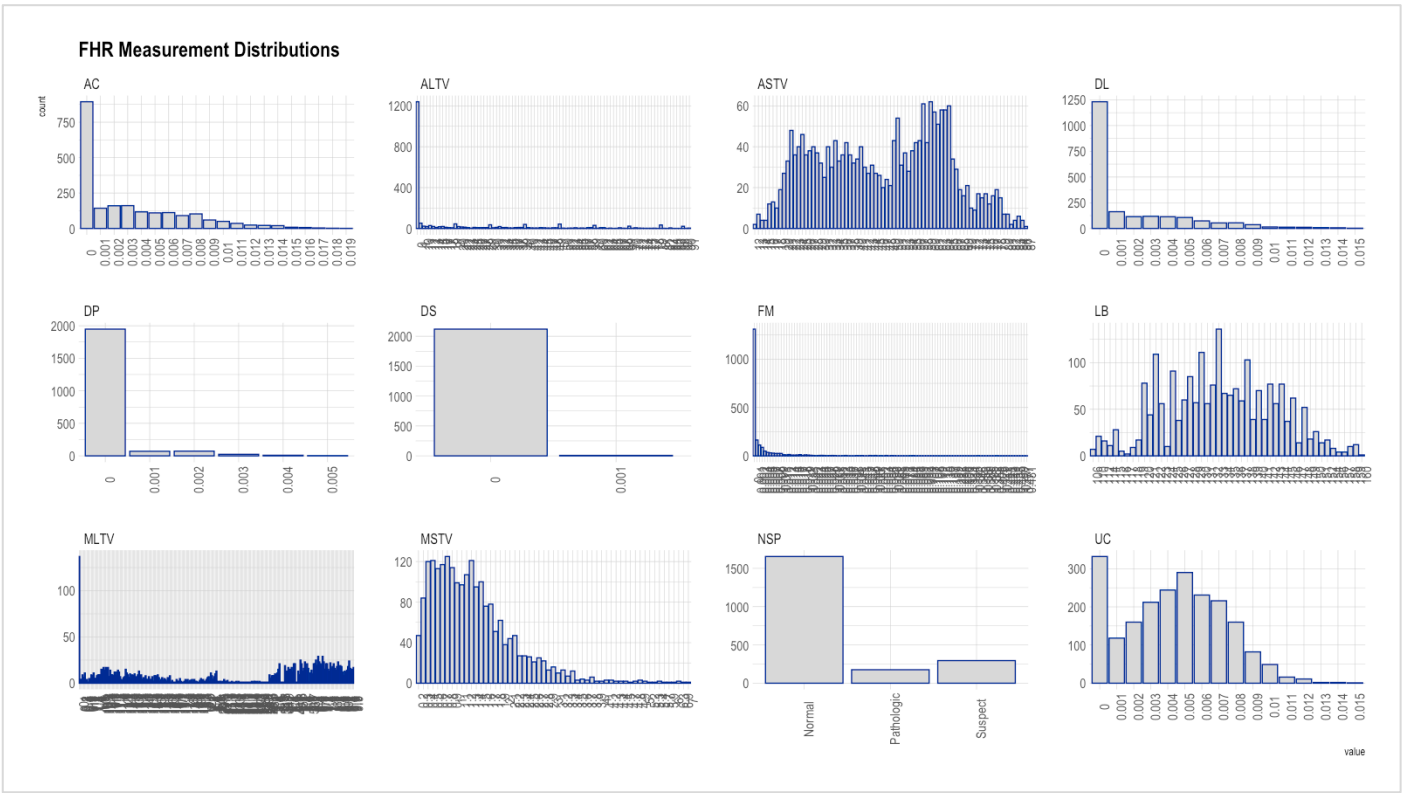


Figure 1.9 – Dataset transformed.

Algorithm Intuition: As previously stated, the conditional inference tree (ctree) is the algorithm to employ for this dataset classification. The objective is to use the model for identification of those independent variables with the greatest influence against the response variable. Author, Torsten Hothorn, summarizes the algorithm behind the ctree as follow, "A conditional inference trees estimate a regression relationship by binary recursive partitioning in a conditional inference framework (2006). Hothorn explains that first "the algorithm tests the sample for hypothesis of independence between the independent variables and the response". If the hypothesis cannot be rejected the process stops. Conversely, it selects the variable with the strongest measured association (measured by the corresponding p-value). Secondly, it implements a binary decision and split in the selected variable, and recursively continues this process until all observations get evaluated (Hothon, 2006).

The first step developing the model was setting a seed and splitting the dataset into different samples to train and test the model. After numerous evaluations, comparison of different simulations, and accuracy appraisals, the dataset splitting proportions were set to .70 and .30. Per figure 2.0, the probability weights for these subsets were set to 70%, and 30% for the training and testing subsets, respectively.

```
> #5. Split the data into a training and test set
> set.seed(1234)
> ind <- sample(2, nrow(ctg), replace = T, prob = c(0.70, 0.30))
> train.data <- ctg[ind == 1, ]
> test.data <- ctg[ind == 2, ]
```

Figure 2.0 – Subsets probability weights.

Appendix A4 portrays the structure of the developed training model. It consist of 25 inner nodes and 26 terminal nodes for a total of 51 nodes. The first split or branching to the

right occurs at the root where driven by DP “prolonged deceleration” > 0.001, it groups 149 observations, with a majority of pathologic cases distributed through nodes 47, 48, & 49. To the left of the root, DP readings of <0.001 decelerations are furthered split by ALTV “percentage of time with abnormal long term variability” of <68 to the left, and >68 to the right to node 44. The footprint of the tree is significantly large, illustrated on figure 2.1 Also, although the greater concentration of pathologic cases are located at the right side of the tree, additional pathologic cases are noticed on the left limbs, namely node 22, 21, and 15. As an observation, this is an example of irregular behaviors potentially hidden to the human sight.

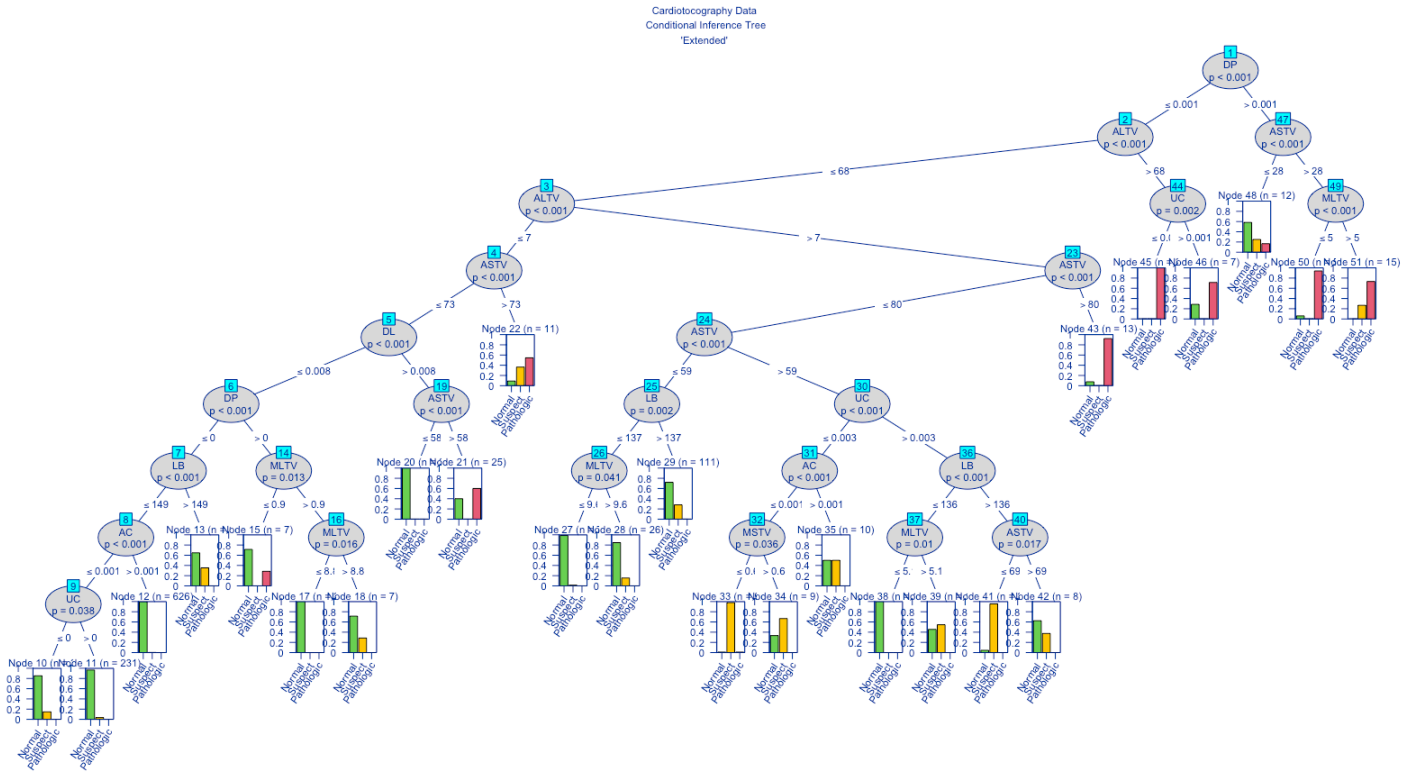


Figure 2.1 – Structure of training dataset model.

The following confusion matrix illustrates how the model performed against the training data. Out of 1523 total observations, the model properly categorized 1413 for a predicted classification accuracy of ~93%.

```
> table(predict(model), train.data$NSP, dnn=c("PREDICTED", "ACTUAL"))
            ACTUAL
PREDICTED    Normal Suspect Pathologic
  Normal       1168      70          4
  Suspect        10     123          1
  Pathologic     17       8        122
```

Figure 2.2 – Training model confusion matrix.

Output. Completed the training portion, the test sample was evaluated against the model. As mentioned before, this set contained the remaining 30% of the sampled population. The same branched out to 16 inner nodes and 17 leaf nodes, for a total of 33. Nodes 33, 32, 30, 28, 18, 17, and 15 were distinguished by containing observations identified as *pathologic*. Alternatively, node 27, 24, and 14 predominant classification was *suspect*. Refer to appendix A5 for a full depiction of this model in "simple" mode.

```
Model formula:
NSP ~ LB + AC + FM + UC + DL + DS + DP + ASTV + MSTV + ALTV +
    MLTV

Fitted party:
[1] root
|   [2] DP <= 0.001
|   |   [3] ALTV <= 68
|   |   |   [4] ALTV <= 15
|   |   |   |   [5] DP <= 0
|   |   |   |   |   [6] DL <= 0.01
|   |   |   |   |   |   [7] ALTV <= 1: Normal (n = 291, err = 1.0%)
|   |   |   |   |   |   [8] ALTV > 1
|   |   |   |   |   |   |   [9] LB <= 143
|   |   |   |   |   |   |   |   [10] ASTV <= 59: Normal (n = 78, err = 0.0%)
|   |   |   |   |   |   |   |   [11] ASTV > 59: Normal (n = 14, err = 42.9%)
|   |   |   |   |   |   |   [12] LB > 143
|   |   |   |   |   |   |   |   [13] ASTV <= 45: Normal (n = 7, err = 14.3%)
|   |   |   |   |   |   |   |   [14] ASTV > 45: Suspect (n = 13, err = 15.4%)
|   |   |   |   |   [15] DL > 0.01: Normal (n = 10, err = 30.0%)
|   |   |   |   [16] DP > 0
|   |   |   |   |   [17] MSTV <= 1.7: Normal (n = 12, err = 25.0%)
|   |   |   |   |   [18] MSTV > 1.7: Normal (n = 12, err = 50.0%)
|   |   |   [19] ALTV > 15
|   |   |   |   [20] ASTV <= 59
|   |   |   |   |   [21] LB <= 147
|   |   |   |   |   |   [22] FM <= 0: Normal (n = 30, err = 3.3%)
|   |   |   |   |   |   [23] FM > 0: Normal (n = 7, err = 28.6%)
|   |   |   |   |   [24] LB > 147: Suspect (n = 13, err = 46.2%)
|   |   |   |   [25] ASTV > 59
|   |   |   |   |   [26] MSTV <= 0.5
|   |   |   |   |   |   [27] ASTV <= 77: Suspect (n = 51, err = 2.0%)
|   |   |   |   |   |   [28] ASTV > 77: Suspect (n = 9, err = 44.4%)
|   |   |   |   |   [29] MSTV > 0.5: Normal (n = 11, err = 27.3%)
|   |   [30] ALTV > 68: Pathologic (n = 14, err = 14.3%)
|   [31] DP > 0.001
|   |   [32] ASTV <= 57: Normal (n = 10, err = 50.0%)
|   |   [33] ASTV > 57: Pathologic (n = 21, err = 0.0%)

Number of inner nodes:    16
Number of terminal nodes: 17
```

Figure 2.4 – Structure of test data model.

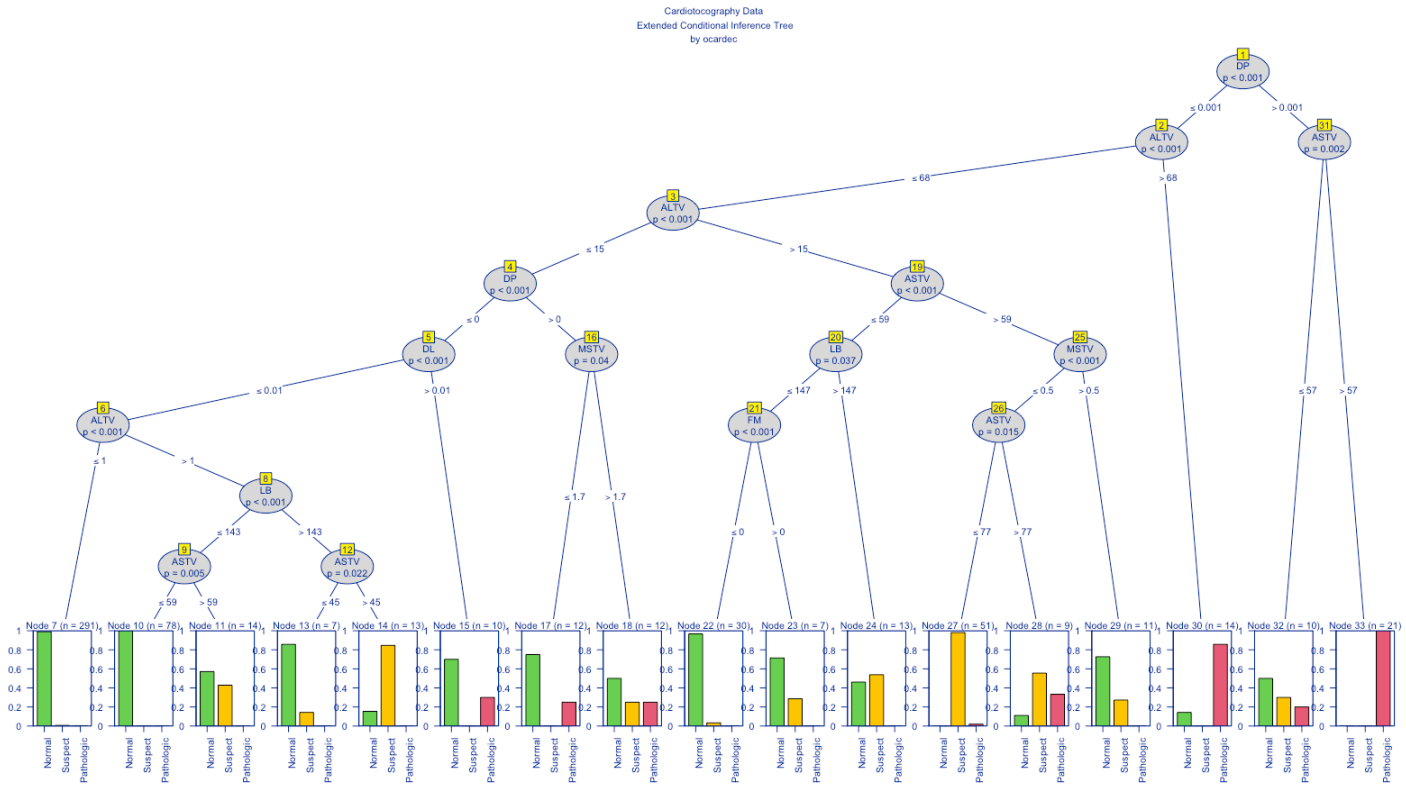


Figure 2.5 – Test data conditional inference tree.

Per the above graphic (figure 2.5), variables prolonged decelerations (DP), percentage of time with abnormal long term variability (ALTV), and percentage of time with abnormal short term variability (ASTV) were key splitting nodes affecting all other child-nodes. The data does not provide the causes of these irregularities, but it does highly the important of such percentages as triggers to constantly observe. Importantly, node 5 portrays a split based on FHR baseline changes (DL). It segregates readings based on those bellow or above the 1% variation. Those greater than or equal to 0.01 branch out to terminal node 15, including *pathologic* readings. On the other side, those less than or equal to 0.01 branch to the left. This node 5 splitting point, captures 403 out of the 637 sample for a total of 64.5% (including node 15) based on the DL reading, highlighting the previously mentioned assumption on the importance of those DL readings within 2 standard deviations.

The confusion matrix for the test.data model captures a total of 555 correct predictions for a classification accuracy rate of 92% with confidence intervals between 89.5% and 94%.

```
> # Confusion matrix and stats
> testPred2 <- predict(model2, newdata = test.data, method="NSP")
> confusionMatrix(testPred2, test.data$NSP)
Confusion Matrix and Statistics

            Reference
Prediction   Normal Suspect Pathologic
  Normal        449      21         12
  Suspect         9      73          4
  Pathologic      2       0         33

Overall Statistics

               Accuracy : 0.9204
                 95% CI : (0.8958, 0.9407)
    No Information Rate : 0.7629
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7809

 Mcnemar's Test P-Value : 0.001165

Statistics by Class:

                     Class: Normal Class: Suspect Class: Pathologic
Sensitivity                 0.9761         0.7766           0.67347
Specificity                 0.7692         0.9745           0.99639
Pos Pred Value              0.9315         0.8488           0.94286
Neg Pred Value              0.9091         0.9594           0.97183
Prevalence                  0.7629         0.1559           0.08126
Detection Rate              0.7446         0.1211           0.05473
Detection Prevalence        0.7993         0.1426           0.05804
Balanced Accuracy           0.8727         0.8755           0.83493
```

Figure 2.6 – Test data sample confusion matrix.

Conclusion. Based on the above model's descriptions, and inferred findings, one may argue and advocate for the use of supervised learning algorithms, like the ctree. This can definitely serve as supporting mechanisms through the medical industry, especially, as evaluated here neonatal, or labor and delivery medical departments. Per the described intended goals, this study employed the ctree methodology to assess and classified independent variables and their influence toward the response variable of NSP. As previously alluded, the model enhances

uncovering inclinations, branches, or tree nodes not typically perceivable by the human naked-eye, thus discourages the practice of CTG interpretation solely based on the medical practitioner experience.

Although this is a basic assessment on the usability of the ctree algorithm, supplemented with other data mining techniques including, but not limited to, clustering, classification and dimensionality reduction, could boost the performance accuracy of the model while reducing error variances. Some encountered challenges while conducting the study were related to identifying best transformation techniques or interpretation of the data, given the limited background of the medial domain. As a recommendation, any future development of this algorithm in relation to CTGs would highly benefit from having sustainable knowledge of this medial area. For best accuracy outcomes, better transformation techniques and variable selection process is essential.

References

Bruce, B., & Bruce, A. (2017). Practical statistics for data science. O'Reilly Media, Inc.

Dietrich, D., Heller, B., & B. Yang. (2015). Data science & big data analytics: Discovering, analyzing, visualizing and presenting data. John Wiley & Sons, Inc.

Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Hothorn, T. (2006). Conditional Inference Tree. https://www.rdocumentation.org/packages/party/versions/1.35/topics/Conditional%20Inference%20Trees

Zhao, Z., Zhang, Y., & Deng, Y. (2018). A Comprehensive Feature Analysis of the Fetal Heart Rate Signal for the Intelligent Assessment of Fetal State. Journal of clinical medicine, 7(8), 223. https://doi.org/10.3390/jcm7080223 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6111566/

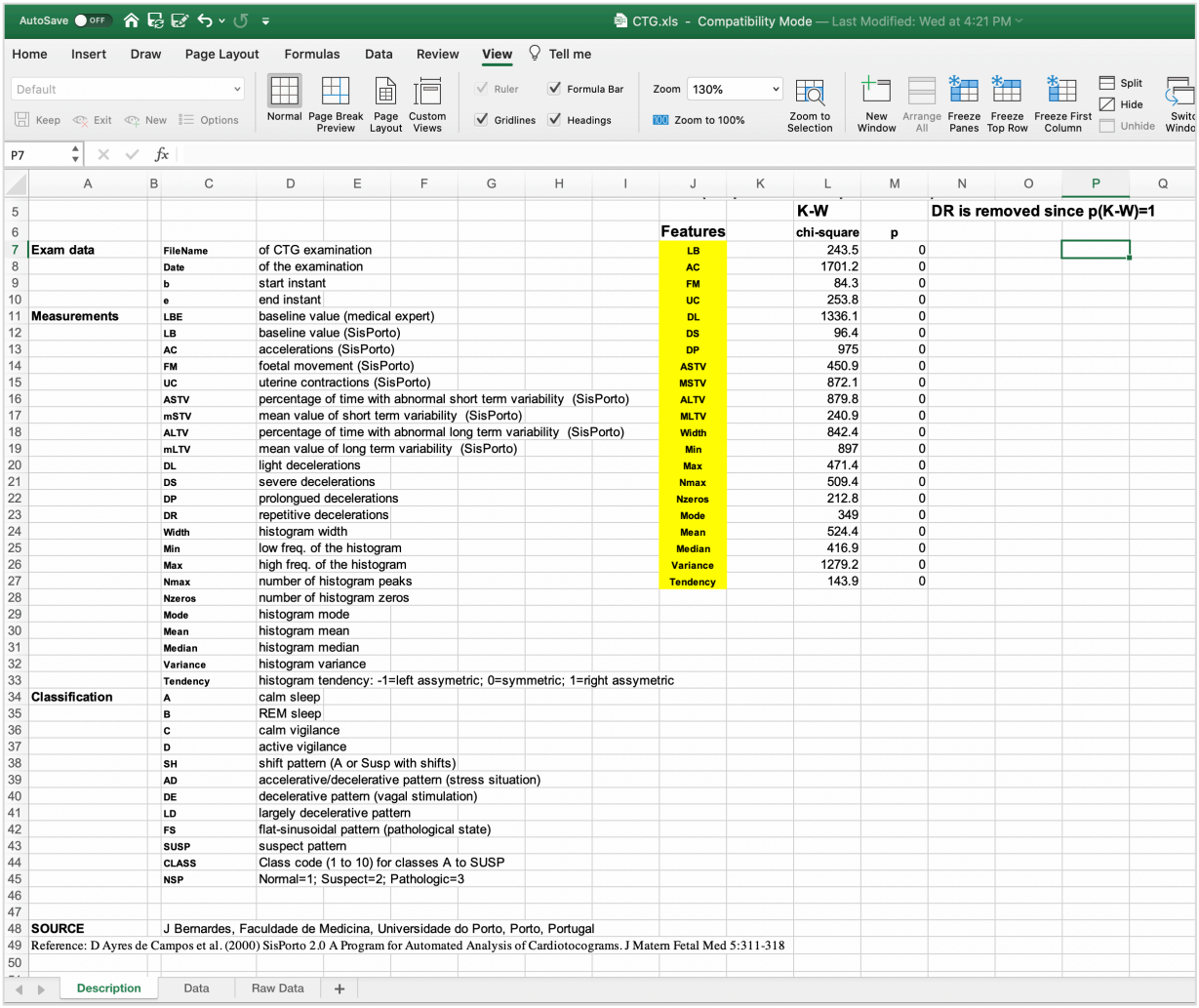
Appendix A

```
# Stats of LB
t.test(ctg$LB)
m<-mean(ctg$LB)
std<-sd(ctg$LB)
upr=m+std
lwr=m-std
lbdf <- data.frame(ctg,my_x = 0 + rnorm(length(ctg$LB), mean=m,
sd=std),my_y = 0 + rnorm(length(ctg$LB), mean=m, sd=std))
# Show LB Variation
print(pltlb<-ggplot(lbdf, xlab=FALSE, aes(x=(my_x), y=my_y))+
geom_line(col="grey51",linemitre=1)+
geom_smooth(method=lm,color="blue",lty=3,fill="lightblue",se=T)+
labs(x=NULL,y="BPM",title="FHR LB Variation\nIn Relation To The Mean")+
theme_ipsum())
```

Appendix A1. Figure 1.4 Code.

```
# What is 2-standard deviations from the mean?
upr2=m+(std*2)
lwr2=m-(std*2)
# Plot LB distribution boundaries
plot.new()
plot(ctg$LB, type="l", col="grey51", ylab="LB", main="1 & 2 Standard
Distributions")
abline(h = m, col = "blue")
abline(h = upr, col = "orange", lty=2)
abline(h = lwr, col = "orange", lty=2)
abline(h = upr2, col = "red", lty=2)
abline(h = lwr2, col = "red", lty=2)
text(-65,134, "mean:133.30", col = "blue", adj = c(0, -.1))
text(-65,upr, round(upr, 2), col = "black", adj = c(0, -.1))
text(-65,lwr, round(lwr, 2), col = "black", adj = c(0, -.1))
text(-65,upr2, round(upr2, 2), col = "black", adj = c(0, -.1))
text(-65,lwr2, round(lwr2, 2), col = "black", adj = c(0, -.1))
```

Appendix A2. Figure 1.5 Code.



5															K-W			DR is removed since p(K-W)=1						
6														Features	chi-square	p								
7	**Exam data**		**FileName**	of CTG examination										**LB**	243.5	0								
8			**Date**	of the examination										**AC**	1701.2	0								
9			**b**	start instant										**FM**	84.3	0								
10			**e**	end instant										**UC**	253.8	0								
11	**Measurements**		**LBE**	baseline value (medical expert)										**DL**	1336.1	0								
12			**LB**	baseline value (SisPorto)										**DS**	96.4	0								
13			**AC**	accelerations (SisPorto)										**DP**	975	0								
14			**FM**	foetal movement (SisPorto)										**ASTV**	450.9	0								
15			**UC**	uterine contractions (SisPorto)										**MSTV**	872.1	0								
16			**ASTV**	percentage of time with abnormal short term variability (SisPorto)										**ALTV**	879.8	0								
17			**mSTV**	mean value of short term variability (SisPorto)										**MLTV**	240.9	0								
18			**ALTV**	percentage of time with abnormal long term variability (SisPorto)										**Width**	842.4	0								
19			**mLTV**	mean value of long term variability (SisPorto)										**Min**	897	0								
20			**DL**	light decelerations										**Max**	471.4	0								
21			**DS**	severe decelerations										**Nmax**	509.4	0								
22			**DP**	prolongued decelerations										**Nzeros**	212.8	0								
23			**DR**	repetitive decelerations										**Mode**	349	0								
24			**Width**	histogram width										**Mean**	524.4	0								
25			**Min**	low freq. of the histogram										**Median**	416.9	0								
26			**Max**	high freq. of the histogram										**Variance**	1279.2	0								
27			**Nmax**	number of histogram peaks										**Tendency**	143.9	0								
28			**Nzeros**	number of histogram zeros																				
29			**Mode**	histogram mode																				
30			**Mean**	histogram mean																				
31			**Median**	histogram median																				
32			**Variance**	histogram variance																				
33			**Tendency**	histogram tendency: -1=left assymetric; 0=symmetric; 1=right assymetric																				
34	**Classification**		**A**	calm sleep																				
35			**B**	REM sleep																				
36			**C**	calm vigilance																				
37			**D**	active vigilance																				
38			**SH**	shift pattern (A or Susp with shifts)																				
39			**AD**	accelerative/decelerative pattern (stress situation)																				
40			**DE**	decelerative pattern (vagal stimulation)																				
41			**LD**	largely decelerative pattern																				
42			**FS**	flat-sinusoidal pattern (pathological state)																				
43			**SUSP**	suspect pattern																				
44			**CLASS**	Class code (1 to 10) for classes A to SUSP																				
45			**NSP**	Normal=1; Suspect=2; Pathologic=3																				
46																								
47																								
48	**SOURCE**		J Bernardes, Faculdade de Medicina, Universidade do Porto, Porto, Portugal																					
49	Reference: D Ayres de Campos et al. (2000) SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J Matern Fetal Med 5:311-318																							
50																								

Appendix A3. Dataset dictionary.

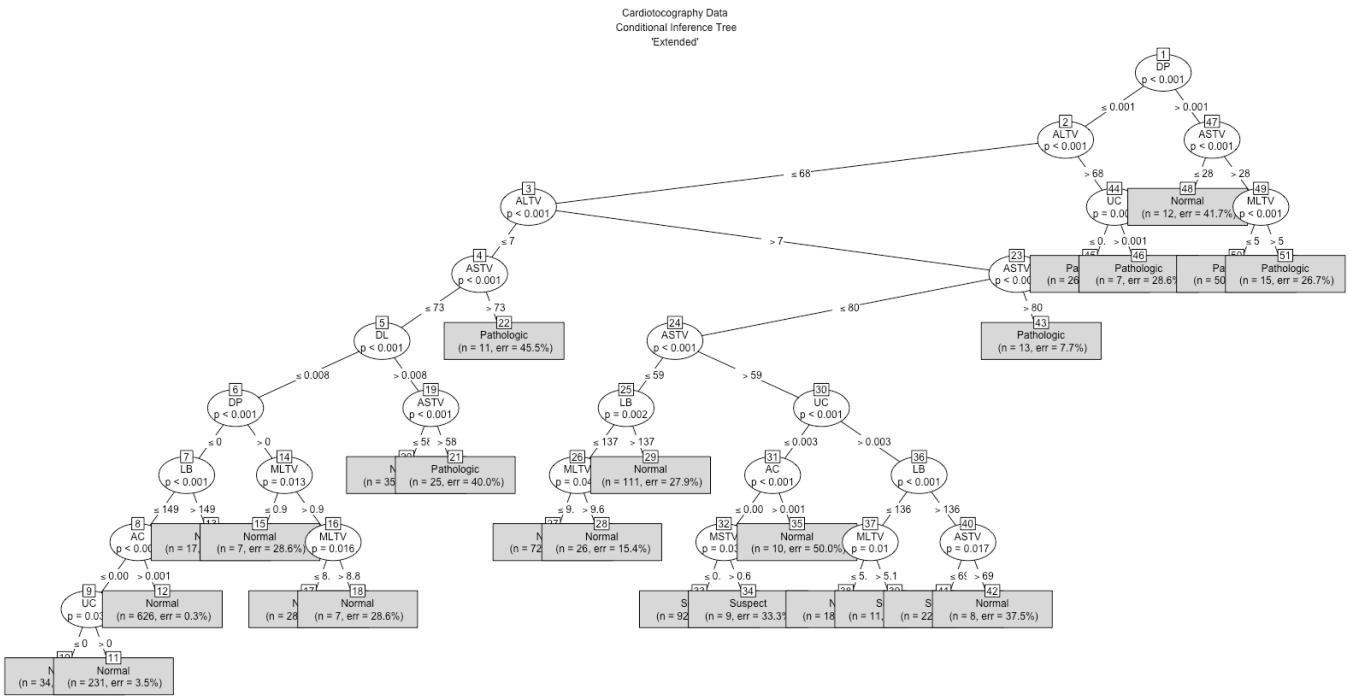
Model formula:
NSP ~ LB + AC + FM + UC + DL + DS + DP + ASTV + MSTV + ALTV +
MLTV

Fitted party:
[1] root
| [2] DP <= 0.001
| | [3] ALTV <= 68
| | | [4] ALTV <= 15
| | | | [5] ASTV <= 58
| | | | | [6] LB <= 143
| | | | | | [7] DP <= 0
| | | | | | | [8] UC <= 0.01
| | | | | | | | [9] UC <= 0: Normal (n = 85, err = 3.5%)
| | | | | | | | [10] UC > 0: Normal (n = 633, err = 0.2%)
| | | | | | | [11] UC > 0.01: Normal (n = 14, err = 7.1%)

```
|  |  |  |  |  |  [12] DP > 0: Normal (n = 30, err = 13.3%)
|  |  |  |  |  [13] LB > 143
|  |  |  |  |  |  [14] AC <= 0.003: Normal (n = 56, err = 25.0%)
|  |  |  |  |  |  [15] AC > 0.003: Normal (n = 54, err = 0.0%)
|  |  |  |  [16] ASTV > 58
|  |  |  |  |  [17] ALTV <= 6
|  |  |  |  |  |  [18] DL <= 0.008
|  |  |  |  |  |  |  [19] ASTV <= 76
|  |  |  |  |  |  |  |  [20] DP <= 0
|  |  |  |  |  |  |  |  |  [21] UC <= 0: Normal (n = 8, err = 37.5%)
|  |  |  |  |  |  |  |  |  [22] UC > 0
|  |  |  |  |  |  |  |  |  |  [23] MLTV <= 7.7: Normal (n = 90, err = 0.0%)
|  |  |  |  |  |  |  |  |  |  [24] MLTV > 7.7: Normal (n = 49, err = 14.3%)
|  |  |  |  |  |  |  |  [25] DP > 0: Normal (n = 12, err = 16.7%)
|  |  |  |  |  |  |  [26] ASTV > 76: Pathologic (n = 7, err = 42.9%)
|  |  |  |  |  |  [27] DL > 0.008: Pathologic (n = 27, err = 44.4%)
|  |  |  |  |  [28] ALTV > 6
|  |  |  |  |  |  [29] ASTV <= 69: Suspect (n = 27, err = 22.2%)
|  |  |  |  |  |  [30] ASTV > 69: Normal (n = 11, err = 9.1%)
|  |  |  [31] ALTV > 15
|  |  |  |  [32] ASTV <= 79
|  |  |  |  |  [33] ASTV <= 59
|  |  |  |  |  |  [34] MLTV <= 9.3
|  |  |  |  |  |  |  [35] FM <= 0.003: Normal (n = 97, err = 10.3%)
|  |  |  |  |  |  |  [36] FM > 0.003: Normal (n = 7, err = 42.9%)
|  |  |  |  |  |  [37] MLTV > 9.3: Suspect (n = 19, err = 42.1%)
|  |  |  |  |  [38] ASTV > 59
|  |  |  |  |  |  [39] MSTV <= 0.6
|  |  |  |  |  |  |  [40] UC <= 0.006: Suspect (n = 114, err = 4.4%)
|  |  |  |  |  |  |  [41] UC > 0.006: Suspect (n = 10, err = 40.0%)
|  |  |  |  |  |  [42] MSTV > 0.6
|  |  |  |  |  |  |  [43] LB <= 139: Normal (n = 15, err = 13.3%)
|  |  |  |  |  |  |  [44] LB > 139: Suspect (n = 7, err = 14.3%)
|  |  |  |  [45] ASTV > 79: Pathologic (n = 7, err = 14.3%)
|  |  [46] ALTV > 68
|  |  |  [47] UC <= 0.001: Pathologic (n = 20, err = 0.0%)
|  |  |  [48] UC > 0.001: Pathologic (n = 7, err = 28.6%)
|  [49] DP > 0.001
|  |  [50] ASTV <= 24: Normal (n = 11, err = 36.4%)
|  |  [51] ASTV > 24: Pathologic (n = 59, err = 8.5%)

Number of inner nodes:    25
Number of terminal nodes: 26
```

Appendix A4. Training set model formula and fitted structure.



Appendix A5. Conditional Inference Tree, training dataset, "simple" mode.

```
> #8. visualize the tree
> plot(model, main="Cardiotocography Data\n Conditional Inference Tree\n'Extended'",
+       type="simple",ep_args = list(justmin = 8), drop_terminal = F,
+      gp = gpar(fontsize = 9), margins = c(4,4, 4, 4))
> plot(model, type="extended", ep_args = list(justmin =8), drop_terminal=F, tnex=1.5,
+      gp=gpar(fontsize = 8, col="dark blue"),
+      inner_panel = node_inner(model, fill=c("light grey","cyan"), pval=T),
+      terminal_panel=node_barplot(model, fill=c(3,7,2), beside=T, ymax=1, rot = 75,
+      just = c(.95,.5), ylines=T, widths = 2, gap=0.05, reverse=F, id=T),
+      margins = c(5,3, 4, 3),
+      main ="Cardiotocography Data\n Conditional Inference Tree\n'Extended'")
```

Appendix A6. Figure 2.1 code.