Influence of Feature Selection Methods on Cardiotocography Data: A Quantitative Investigation

M. Ramla

Abstract*: Cardiotocography is the most powerful tool to monitor the fetal health state during both antenatal and postnatal periods. Using the EFM machine, the fetal heart beat pace and uterine contractions of pregnant mother are recorded simultaneously. This signal data will help doctors to assess wellness of the fetus and classify them. Feature selection is of utmost importance to potentially uplift the predictive model for CTG analysis. The objective of Feature Selection is to keep more pertinent features that capture the hidden insights and ignoring the redundant variables to improve the classification accuracy of the predictive model. A walkthrough, on the UCI-CTG dataset using different feature selection methods such as filters, wrappers, and embeds are made and then compared. This paper attempts to analyze the inherent nature of the data using methods such as Linear Discriminant Analysis, Recursive Elimination, Forward and Backward Elimination and Lasso Regression. The extracted features are used for classification and the performance evaluation of the classifier is observed using Accuracy as the performance metrics.

Keywords: Feature Selection, CTG, Fetal Heart Rate, LDA, Lasso, Recursive Feature Elimination.

I. INTRODUCTION

Feature Selection is the crucial procedure of discarding excess or unrelated features to reduce the runtime of the classifier. [4] The extraneous data drastically disturbing the classification accuracy must be removed. The process of Feature selection behaves like a filter. It aims at opting out features that are not useful. The potential benefits of Feature Selection are to make the machine learning algorithm to learn quicker, easier to understand, figuring out the complexity of the model and above all to shoot up the accuracy of the model. The problem of overfitting is also reduced considerably. Filters, Wrappers, and Embeds are three common types of feature selection algorithms [4].

A. Filter Methods

Filter methods quantifies each attribute with a statistical measure and gives weightage to each predictor. These predictors are then ordered by the score and either chosen or detached from the feature set. The methods are mostly univariate. They judge the features separately, or with respect to the target class. ANOVA, Chi squared test, T-Test and correlation coefficient are few filter selection approaches.

B. Wrapper Methods

Wrapper methods, deems feature selection as a greedy hunt for best features, where several combination of subsets are collected, evaluated and matched up with other combinations. In detail, it picks a portion of the feature set, assesses the performance of the classifiers on that portion.[4] This is then repeatedly done until desired performance is reached. The feature subset for that yields the maximum accuracy is selected. [4] The search process may be done in a methodically or stochastically or by applying heuristics. [17]. Forward Selection, Backward Elimination, Recursive Feature Elimination (RFE) are few wrapper methods.

C. Embedded Methods

Embedded methods measure the usefulness of the feature subset guided by the learning process. Those features that predominantly contribute to the accuracy of the model are arrived at while the model is being developed. Feature selection is built in within the algorithm used for prediction. This performs a kind of optimization in the Predictive model developed. The most familiar type of embedded class of feature selection is regularization methods. Regularization methods are popularly referred as penalization methods. They bring in additional constraints into the optimization of a predictive algorithm and bias the model towards lower complexity. LASSO and Ridge Regularization, Elastic Net, Genetic Algorithms are some embedded methods of feature selection.

II. BACKGROUND

Cardiotocography is a technical means of recording the fetal heart beat rate and Uterine Contractions of mother. CTG is a monitoring technique used routinely during pregnancy and labor. This non-invasive test is conducted to assess the wellness of the fetal. [1]. The subjective interpretation of CTG traces has conflicts among inter observers. Automated FHR analysis confining to the International Guidelines has been agreed upon as the most promising solution. [2] Advances in the field of Machine Learning and AI have made it possible for machines to learn from data and make predictions based on learning. CTG is commonly used in third trimester and during labor to monitor the fetal well being and detect the fetal hypoxia. The key features used by physicians are Baseline FHR, Baseline Variability, Accelerations and Decelerations. [3] Accelerations, shows the signs of healthy baby and Decelerations are a sudden dip in the fetal heart rate by 15bpm. [18]

Revised Manuscript Received on April 25, 2019.

***** Correspondence Author

M. Ramla*, Department of Computer Applications, Faculty of Science and Humanities, SRM IST, Kattankulathur, Chennai. E-mail: ramzsami@gmail.com.

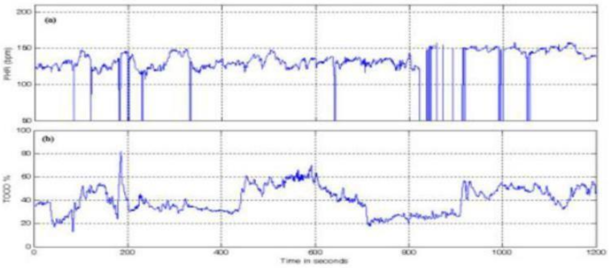


Fig: 1. A typical CTG Signals- FHR & UC

Plethora of research has reviewed the performance of multitude of machine learning algorithms by applying it on CTG data.[3] But the effectiveness of the predictive model relies on the primary feature selection step. The paramount success of the predictive model is quite often dictated by the good features. Hence, a walkthrough on the CTG data using various feature selection methods have been attempted in this paper. The rest of the paper is organized as follows: Data Set Description, Feature Selection Methods on CTG data, experimental outputs, Graphical Visualizations, Results and Discussions.

III. DATA SET DESCRIPTION

Source : UCI CTG DataSet
No. of Instances :2126 measurements and classifications of foetal heart rate (FHR) signals

FEATURES	INFORMATION (per second)
LB	FHR baseline (bpm)
AC	No. of accelerations
FM	No. of fetal movements
UC	No. of Uterine contractions
DL	No. of light decelerations
DS	No. of severe decelerations
DP	# of prolonged decelerations
ASTV	Percentage of time with abnormal short term variability
MSTV	Mean Value of short 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	No. of Histogram Peaks
Nzeros	No. of histogram zeros
Mode	Histogram Mode
Mean	Histogram Mean
Class	N-Normal, S-Suspicious, P-Pathological

The fetal state is interpreted by physicians based on the International Federation of Gynecology and Obstetrics (FIGO) as follows,

	Normal CTG[a]	Suspicious CTG	Pathological CTG
Baseline[b]	110-160 bpm	Lacking at least one of normal characteristics, but with no pathological features	<100 bpm
Variability[c,d,j]	5-25 bpm		Reduced/increased variability[c,d]; sinusoidal pattern[i]
Decelerations[e,f,g,h,i]	No repetitive* decelerations		Repetitive* late or prolonged decelerations for >30 min (or >20 min if reduced variability); one deceleration >5 min
Interpretation	No hypoxia/acidosis	Low probability of hypoxia/acidosis	High probability of hypoxia/acidosis

IV. FEATURE SELECTION ALGORITHMS

A. Linear Discriminant Analysis

This filter approach of Feature Selection initiates by finding directions to maximum separation between classes. The basic assumptions are, the predictors are Gaussian distribution and it is a multiclass problem. [5] LDA computes the group means and then computes for each individual the probability of belonging to different groups. The individual is then affected to the group with high probability.

LDA emphasizes on finding a subspace which clusters the samples from the same class while expanding the margin of samples from different classes. [5] Linear Discriminant Analysis also works as a dimensionality reduction algorithm, it means that it reduces the number of dimension from original to C — 1 number of features where, C denotes then number of classes.

The results of Linear Discriminant Analysis performed on CTG dataset are tabulated.

Prior probabilities of groups:

1	2	3
0.77845720	0.13875823	0.08278457

Features	LD1	LD2
LB	6.30-02	4.471480e-02
AC	2.594040e+01	2.033782e+02
FM	2.106918e-01	-9.134641e-01
UC	-7.018283e+01	6.879580e+01
DL	-1.798974e+01	-1.445936e+01
DP	1.156131e+03	1.640411e+02
ASTV	2.766136e-02	-1.005738e-02
MSTV	-1.091363e-02	2.014578e-02
ALTV	3.813500e-02	-6.789201e-03
MLTV	2.019927e-02	1.258203e-02
Width	-1.789725e-03	-2.611049e-04
Min	8.485110e-03	-3.051089e-03
Max	1.328779e-02	-8.840726e-03
NMax	-1.729994e-02	-6.561280e-02
Nzeros	4.662333e-02	-6.835874e-03
Mode	-1.654057e-02	-6.808454e-03
Median	-4.016835e-02	-9.133098e-03
Variance	1.063108e-02	1.983105e-03
Tendency	2.764795e-01	-2.412240e-02

The histogram and density plot for CTG observations in each of the three groups on first linear discriminant are also plotted.

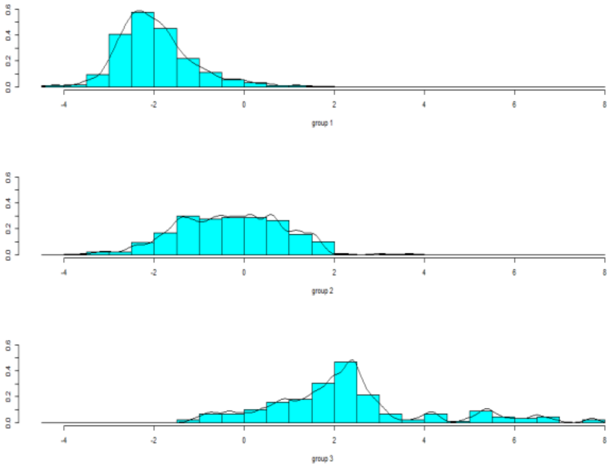


Fig: 2 Histogram distribution for Class Groups

Proportion of Trace

LD1	LD2
0.8006	0.1994

The plots for each observation in the space of the first two Linear Discriminants are plotted. Points are identified using Group ID.

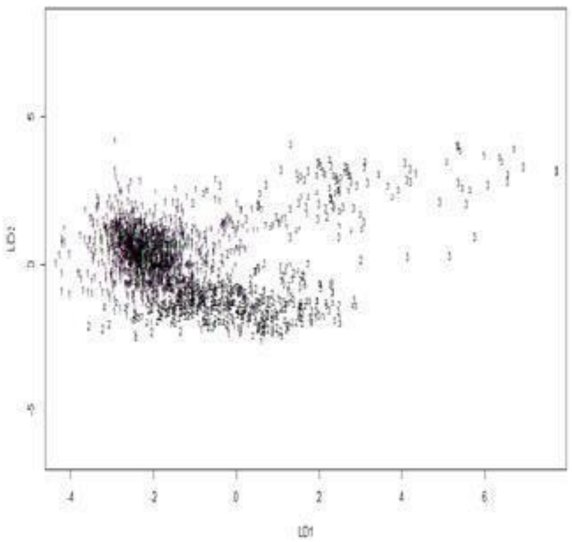


Fig: 3 Density Plot of Linear Discriminants

B. Wrapper Methods

Wrapper methods use a subset of features and train the model using them. Based on the deductions that we draw from the previous model, we fix on inclusion or omission of features from the subset. In this way, the problem essentially reduces to a search problem. Popular and most common list of wrapper methods includes forward feature selection, backward feature elimination, recursive feature elimination, etc.

Boruta package in R was used for implementing feature selection with wrapper methods. This package finds the importance of a feature by creating shadow features. [13]

Boruta is a wrapper algorithm that outputs the Variable Importance Measure and by default uses random forest classification algorithm. This algorithm confines all the important, interesting features present in the dataset with respect to an outcome variable.

Step in Boruta:

(a) Add shadow features using permuted values.
(b) Train a random forest classifier on the extended data set
(c) At every iteration, check if the real feature has higher Z-score than the shadow feature. If so, it is called as a HIT.
(d) Algorithm terminates when all features are either accepted or rejected

The plot uncovers the significance of each of the features of CTG dataset. The top six confirmed features of the dataset are listed below.

Features	Mean Imp	Decision
ASTV	31.82234	Confirmed
ALTV	30.66349	Confirmed
Mean	29.45711	Confirmed
MSTV	28.22882	Confirmed
AC	25.38928	Confirmed
DP	25.05904	Confirmed

The general idea behind the algorithm is to reduce the misleading impact of randomness in the original feature sets.

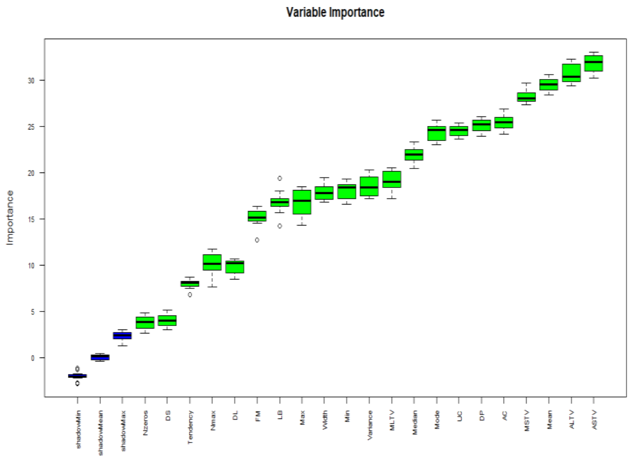


Fig: 4 Box Plot of Variable Importance

C. Forward & Backward Elimination Method

This method searches for the best possible regression model by repeatedly choosing and removing variables to land at a model with the lowest possible Akaike Information Criterion (AIC). This method
deals with variables with high p-value.
The shortlisted features in the step wise algorithm in both the directions (forward & backward) are listed below.

[1] "DP" [2] "ALTV" [3] "ASTV" [4] "DS" [5] "Variance" [6] "UC" [7] "Mode" [8] "LB" [9] "Median" [10] "Min" [11] "MLTV" [12] "Width" [13]"Tendency"[14] "Mean"

D. Recursive Feature Elimination

The RFE method provides a scrupulous way to find out the important variables before feeding it into a Machine Learning algorithm. It works on the Feature Ranking system and removes the low ranking features. The recursive feature elimination algorithm is supplied with the number of features to iterate and the variable evaluation algorithm. The performance of the method over CTG data is tabularized.

Cross-Validated (10 fold, repeated 5 times) Performance over subset size

F	RMS E	R2	MAE	RMSE SD	R2SD	MAES D
1	0.6 8	0.037 21	0.46 85	0.047 49	0.021 22	0.028
2	0.53 13	0.252 83	0.39 27	0.045 42	0.072 8	0.026 T 2:00 PM
3	0.51 48	0.297 3	0.38 32	0.042 45	0.069 8	0.026 in 04 as
4	0.49 62	0.347 17	0.37 15	0.041 5	0.065 19	0.026 99 th
5	0.49 58	0.348 5	0.37 14	0.041 32	0.064 75	0.026 f 69 d
10	0.4 35	0.565 54	0.28 3	0.036 3	0.065	0.024 F. 61
12	0.4 32	0.566 33	0.28 35	0.036 79	0.064 68	0.024 C 93 d

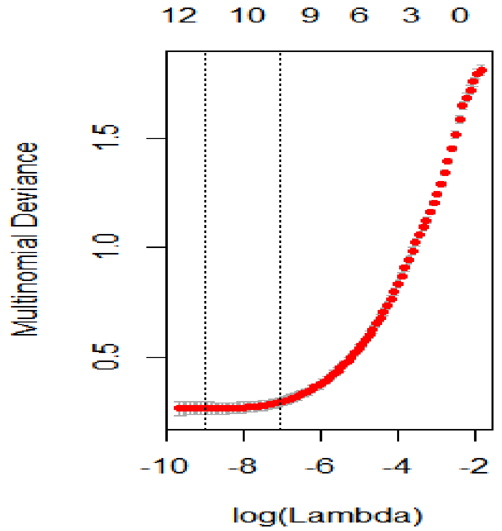
The best 5 influential variables (out of 12): are DS, DP, DL, AC, FM

E. Embedded Method - Lasso Regression

Least Absolute Shrinkage and Selection Operator is the expansion of LASSO. It is a regression analysis method. LASSO uses the regularization method that penalizes with L1-norm.It shrinks the complexity of the model and prevents overfitting as well. LASSO imposes a cost to having large coefficient values and forces it to be less than a threshold value, in a way shrinking certain coefficients to be 0. And it's called L1 regularization, as the cost added, is proportional to the absolute value of weight coefficients. Thus it eliminates the unnecessary variables totally.

(Intercept)	1.000342
LB	-0.05138
DP	2436.275
MSTV	-0.02358
ALTV	-0.01958
MLTV	0.039171
MIN	-0.70528
MAX	-1.44693
MEAN	0.061079
MEDIAN	2.201823
VARIANCE	0.085736

A variable is considered to be important if it has high positive or low negative coefficients.



The numeric value at the top of the plot shows the number of redictors added in the model. The red dots along the Y-axis dicate the Area Under ROC Curve obtained when you add many features as made known on the top x-axis.

The first vertical line on the left, points to the lambda with the lowest MSE. The second line indicates the number of features with the highest deviance within 1 standard deviation.

F. Experimental Results

To compare the efficacy of the feature selection, lassification was performed on the features selected by the ifferent feature selection process. When a classifier is applied on the output, the prediction becomes more accurate and time efficient.

G. Classifier Output

J48 Tree Classifier with 10-Fold Cross Validation on CTG datasets with the selected attributes from various feature selection algorithm was tested. The Classification Accuracy is tabulated.

Feature Selection	Classification Accuracy
FULL Features	75.0769%
Boruta	77.84%
Forward Selection & Backward Elimination	76.30%
Recursive Feature Elimination	82.16%
Lasso Regression	86.46%
LDA	75.07%

V. CONCLUSION

The effects of the various feature selection methods were tested and compared on the no-feature selection of the CTG data. According to the table presented, the machine learning algorithm, J48 Tree yields good accuracy when the Lasso Regression method of variable selection was used.

In this study, the influence of feature selection on classifier decision tree J48 was investigated, and it was observed that Lasso Regression yields good accuracy compared with other methods on CTG dataset. Ultimately, dimensionality reduction through feature selection leads to the high success rate of the predictive model.

REFERENCES

1. Zafer Comert and Adnan Fatih Kocamaz, A Study Based on Gray Level Co-Occurance MAtrix and Neural Network Community for Determination of Hypoxic Fetuses, International Artificial Intelligence and DAta Processing Symposium, 2016
2. Zafer Comert and Adnan Fatih Kocamaz,Fetal Hypoxia Detection Based on Deep Convolutional Neural Network with Transfer Learning Approach, Springer International Publishing AG, part of Springer Nature 2019, doi:10.1007/978-3-319-91186-1_25
3. M.Ramla, S.Sangeetha, S.Nickolas, Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements, ICCS IEEE Conference Proceedings , 2016
4. Esra Mahsereci Karabuluta*, Selma Ayşe Özelb, Turgay İbrikçib,c, Procedia Technology, Elseivier, A comparative study on the effect of feature selection on classification accuracy,doi:10.1016/j.protcy.2012.02.068, 2011
5. Zhen Lei Shengcai Liao Stan Z. Li, Efficient Feature Selection for Linear Discriminant Analysis and Its Application to Face Recognition
6. Shahad Alyousif1*,MA Mohd2, Bilal B3, M Sheikh and M Algunaidi4, Rule-Based Algorithm for Intrapartum Cardiotocograph Pattern Features Extraction and Classification,Health Science Journal,DOI:10.21767/1791809X.1000468,Vol.10 No.6:468,2016
7. Bertha Guijarro-Berdiñas, Amparo Alonso-Betanzos,Oscar Fontenla-Romero , Intelligent analysis and pattern recognition in cardiotocographic signals using a tightly coupled hybrid system, Artificial Intelligence 136 (2002) 1–27,Elseivier, 2002
8. Nurul Chamidah, Ito Wasito,Fetal State Classification from Cardiotocography Based on Feature Extraction Using Hybrid K-Meansand Support Vector Machine, ICACSIS 2015 IEEE
9. Vinayaka Nagendra, Steve Corns, 'Evaluation of Support Vector Machines and Random Forest Classifiers in a Real time fetal monitoring system based on Cardiotocography data', IEEE, 2017
10. Anish Batra, Ananya Chandra, Vishal Matoria,' Cardiotocography analysis using conjunction of machine learning algorithms', International Conference on Machine vision and information technology, IEEE,2017
11. Z.Comert and A.F. Kocamaz , 'Comparison of Machine earning Techniques for Fetal Heart Rate Classification ', Vol . 132(2017), Special issue of the third international conference on computational and experimental science and engineering, DOI:: 10.12693/APhysPolA.132.451
12. Vipin Kumar and Sonajharia Minz, Feature Selection-A Literature Review, Smart Computing Review, vol. 4, no. 3, June 2014, DOI: 10.6029/smartcr.2014.03.007
13. Miron B. Kursa University of Warsaw Witold R. Rudnicki University of Warsaw, Feature Selection with the Boruta Package, Journal of Statistical Software,Vol 36 Issue 11, 2010
14. Hasan Ocak , Huseyin Metin Ertunc , "Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems", DOI 10.1007/s00521-012-1110-3
15. F. Jiméneza, , G. Sáncheza , J.M. Garcíaa , G. Sciaviccob , L. Miralles, "Multi-objective evolutionary feature selection for online sales forecasting", Neurocomputing 234 (2017) 75–92, Elsevier.
16. Bryan Johnson , Alex Bennett , Myungjae Kwak, Anthony Choi, "Automated Evaluation of Fetal Cardiotocograms using Neural Network", IEEE International Conference on Systems, Man, and Cybernetics, 2012 .
17. G. Pérez-Caballero, J.M. Andrade, P. Olmos, Y. Molina, I. Jiménez, J.J. Durán, C. Fernandez-Lozano, F. Miguel-Cruz. "Authentication of tequilas using pattern recognition and supervised classification", TrAC Trends in Analytical Chemistry, 2017
18. Kodama, Yuki, Hiroshi Sameshima, Rie Yamashita, Masanao Oohashi, and Tsuyomu Ikenoue. "Intrapartum fetal heart rate patterns preceding terminal bradycardia in infants (> 34 weeks) with poor neurological outcome: A regional population-based study in Japan : FHR patterns preceding bradycardia", Journal of Obstetrics and Gynaecology Research, 2015.

AUTHORS PROFILE

Ramla is working as Assistant Professor in the Department of Computer Applications, SRM IST. Currently pursuing her Ph.D from National Institute of Technology, Thiruchy. her research interest are Data Mining, Predictive Analytics in Health Care. She has published articles related to Cardiotocography for fetal health state monitoring.