

Analysis Of Cardiotocogram Data For Fetal Distress Determination

Dr.J.Thamil Selvi¹
Abinaya.R², Amrita Varshini.S³, and Divya.V⁴
Department of Electronics And Communication Engineering
Sri Sai Ram Engineering College, Chennai.

Abstract-

Cardiotocography is one of the most widely used technique for recording changes in fetal heart rate (FHR) and uterine contractions. Assessing cardiotocography is crucial in that it leads to identifying fetuses which suffer from lack of oxygen, i.e. hypoxia. This situation is defined as fetal distress and requires fetal intervention in order to prevent fetus death or other neurological disease caused by hypoxia. Fetal heart rate and uterine contraction are 2 vital parameters that require continuous monitoring during the intrapartum period. These two parameters are used to detect the fetal distress condition. Cardiotocography is the most widely used tool for monitoring FHR and UC. It gives output in graphical format which is interpreted by health care professionals to detect fetal distress cases. Currently these interpretations are subjective and vary from expert to expert and guideline to guideline. We propose to develop an application that takes time series CTG data that classifies them as normal and distress as output.

Keywords- Fetal Heart Rate, Cardiotocograph signals, Supervised Learning, fetal distress, intrapartum, Random Forest.

Interpretation of CTG signal

Interpretation of a CTG tracing requires both qualitative and quantitative description of: Uterine activity (contractions), Baseline fetal heart rate (FHR), Baseline FHR variability, Presence of accelerations, Periodic or episodic decelerations, Changes or trends of FHR patterns over time.

Cardiotocography for fetal monitoring, the system comprised of discriminant analysis (DA), decision tree (DT), and artificial neural network (ANN) in their study to evaluate fetal distress by the same CTG data. This reached the results showing that the accuracies of DA, DT and ANN are 82.1%, 86.36% and 97.78% respectively, and 80%, 10%, and the remaining 10% of the whole dataset which were randomly used for training, testing, and validation respectively. This was implemented a supervised ANN which can classify the CTG data, the results are evaluated with respect to rand index, precision, recall and f-Score.

In the proposed method data set was trained using software and classification was carried out. **Fine Gaussian Support Vector Machine (FGSVM). The features were given as input to the classifier.**

I. INTRODUCTION

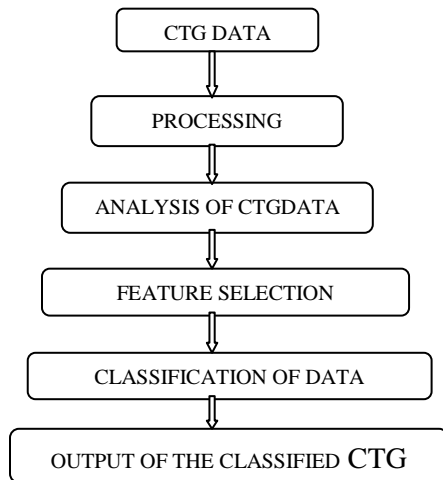
Cardiotocography is a monitoring technique providing important and vital information on fetal status during antepartum and intrapartum periods. It is one of the most widely used technique for recording changes in fetal heart rate (FHR) and uterine contractions. Assessing cardiotocography is crucial in that it leads to identifying fetuses which suffer from lack of oxygen, i.e. hypoxia. This situation is defined as fetal distress and requires fetal intervention in order to prevent fetus death or other neurological disease caused by hypoxia.

According to the World Health Organisation, there are over 2 million neonatal deaths during the intrapartum period worldwide every year. Fetal heart rate and uterine contraction are 2 vital parameters that require continuous monitoring during the intrapartum period.

II. METHODOLOGY

a) Database

The data was extracted from the intrapartum repository. The cardiotocography data set used is publicly available at "The Data Mining Repository of University of California Irvine (UCI)". Out of the 2127 samples, 259 samples were removed. 21 attributes were extracted, out of which 5 attributes were selected.



In [1], the cardiogram data is extracted from the Physionet Repository .

In [2], the data collected is processed and imported.

In [3], the attributes and their characteristics with respect to classification as normal and pathologic CTG signals are analysed.

In [4], features are extracted from the data and the best features are selected.

In [5], the cardiogram data is classified using classifiers.

In [6], the outputs of the classified signals are verified and the most accurate classifier is chosen.

b) Feature Extraction

Feature extraction is frequently used as a preprocessing step to machine learning, it is a process of choosing a subset of original features so that the feature space is optimally reduced according to certain evaluation criterion. The main goal is to increase the accuracy .

By using 21 given attributes data can be classified according to FHR pattern class or fetal state class code. In this study, fetal state class code is used as target attribute instead of FHR pattern class code and each sample is classified into one of three groups normal, suspicious or pathologic. The dataset includes a total of 2126 samples of which is 1655 normal, 295 suspicious and 176 pathologic samples which indicate the existing of fetal distress.

Symbol	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 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-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

The following two feature classification methods to classify the data set are given below:

PCA

Principal component analysis (PCA) is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components .PCA reduces attribute space from a larger number of variables to a smaller number of factors and as such is a "non-dependent" procedure (that is, it does not assume a dependent variable is specified).

Genetic Algorithm

Genetic Algorithms(GAs) are adaptive heuristic search algorithms that belong to the larger part of evolutionary algorithms. Genetic algorithms are based on the ideas of natural selection and genetics. These are intelligent exploitation of random search provided with historical data to direct the search into the region of better performance in solution space. They are commonly used to generate high-quality solutions for optimization problems and search problems.

c) Classification

Classification is the task of generalizing known structure to apply to new data. Data classification may be supervised and unsupervised. The supervised classification method requires the presence of training data set typically defined by the expert-the teacher. Each class of objects is characterized by the basic statistical parameters (mean values vector, covariance matrix), which are values vector, covariance matrix), which are computed from the training set. These parameters guide the discrimination process. The unsupervised classification is also known as classification without the teacher.

The following types of classifiers are used for data classification.

Naive Bayes Classifier (NB)

Naive Bayes is a statistical classifier which assumes no dependency between attributes. It attempts to maximize the posterior probability in determining the class. By theory, this classifier has minimum error rate but it may not be case always. According to Bayesian theorem

$$P(A|B) = P(A) * P(B|A) / P(B)$$

Where, $P(B|A) = P(A \cap B) / P(A)$

Decision Tree (DT)

Decision trees are powerful and popular tools for classification and prediction. It represents rules, which can be understood by humans and used in knowledge system such as database. It is a popular classifier which is simple and easy to implement. It requires no domain knowledge or parameter setting and can handle high dimensional data.

Support vector Machine (SVM)

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples. In two dimensional space this hyperplane is a line dividing a plane in two parts where in each class lay in either side.

Linear SVM

Linear SVM is the newest extremely fast machine learning (data mining) algorithm for solving multiclass classification problems from ultra large data sets that implements an original proprietary

version of a cutting plane algorithm for designing a linear support vector machine.

Random Forest

It is a class for constructing a forest of random trees. RFs are forests of decision trees generated using random sampling, which can be used for both, classification and regression problems. While decision trees comprise of only one tree for testing, RFs comprise of multiple decision trees in the testing phase, thus making it a better option as compared to decision trees. Since, RFs are a group of different decision trees, they are also called an ensemble method. This method groups together multiple classifiers to form a strong classifier and improves performance using divide-and-conquer.

Prediction Accuracy

Classifier accuracy or prediction accuracy is the strength of system for accurate prediction. Accuracy is the number of correct predictions over the entire data set number.

Confusion Matrix

Confusion matrix is a visual tool for distribution of predicted class and actual class. Confusion matrix is square matrix with n rows and n columns which are indexed as actual class of samples and predicted class of samples respectively.

III. RESULTS AND DISCUSSION

The CTG dataset contains 2126 instances and 21 attributes with 1 class attribute. The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC). The CTGs were classified by expert obstetricians and a consensus classification label has been assigned to each of them. The classification is with respect to a fetal heart rate class code (N-Normal, S-Suspect and P-Pathologic). Therefore this dataset can be used for 3-class experiment.

Fig. 1. Sample raw data containing 21 features

Feature Selected Data Sample

Due to the presence of redundant features in the raw data sample, the scatter plot was cluttered and analysis could not be done. Thus, feature selection to remove the same redundant features was carried out.

1	SeqFile	LB	FM	UC	ASTV	MSTV	NSP
2	CTG0002.t	132	0	4	17	2.1	
3	CTG0003.t	133	0	5	16	2.1	
4	CTG0004.t	134	0	6	16	2.4	
5	CTG0005.t	132	0	5	16	2.4	
6	CTG0006.t	134	0	10	26	5.9	
7	CTG0007.t	134	0	9	29	6.3	
8	CTG0008.t	122	0	0	83	0.5	
9	CTG0009.t	122	0	1	84	0.5	
10	CTG0010.t	122	0	3	86	0.3	
11	CTG0013.t	131	57	6	28	1.4	
12	CTG0014.t	131	147	4	28	1.5	
13	CTG0015.t	130	489	5	21	2.3	
14	CTG0016.t	130	273	3	19	2.3	
15	CTG0017.t	130	290	3	24	2.1	
16	CTG0019.t	130	317	4	23	1.9	
17	CTG0020.t	130	557	6	29	1.7	
18	CTG0021.t	129	304	4	30	2.1	
19	CTG0022.t	128	272	2	26	1.7	
20	CTG0023.t	128	219	2	34	2.5	
21	CTG0024.t	128	0	0	80	0.5	
22	CTG0025.t	128	0	2	86	0.3	
23	CTG0026.t	124	0	0	86	0.3	
24	CTG0027.t	124	0	0	86	0.4	
25	CTG0028.t	124	0	0	87	0.2	
26	CTG0030.t	132	65	0	26	6	
27	CTG0031.t	132	129	2	26	4.5	
28	CTG0032.t	132	79	3	22	6.9	

Fig.2.Sample data after feature extraction, containing 5 features

Out of the 21 features that were present in the raw data, 5 features were extracted. The selected features include **baseline heart rate, fetal movement, uterine contractions, short term variability(ASTV, MSTV) in fetal heart rate.**

Features like **long term variability, acceleration in heart rate, deceleration in heart rate** was found to be redundant.

The features are fed to the classifiers like **Decision Tree, Complex Tree, Gaussian Support Vector Machine, and Random Forest Classifier.** The scatter plot for the classified data was generated.

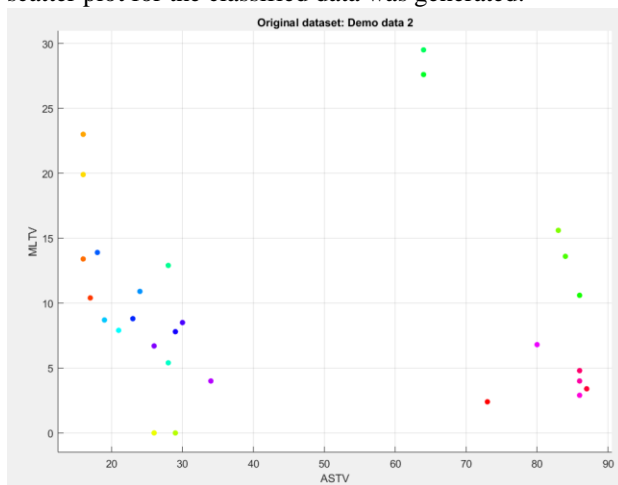


Fig. 3 .Classifier output after the random forest classifier

The data was classified using the classifiers and was trained. The accuracy of each classifier was then noted.

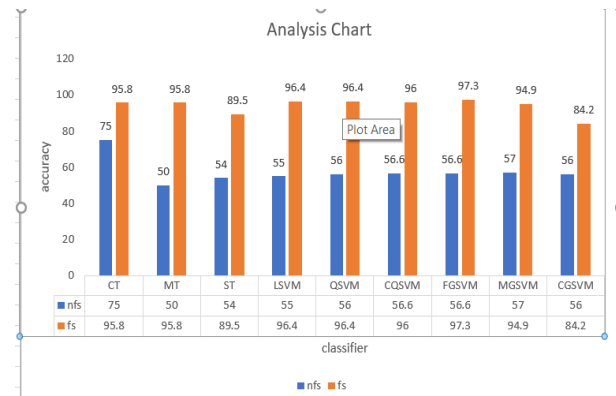


Fig. 4. Comparison of classification techniques before feature selection and after feature selection

Key

Nfs - Non feature selection dataset

Fs - feature selected dataset

The overall performance was analyzed between full CTG dataset (21 attributes) and feature reduced CTG dataset (5 attributes). The comparison result of the classification techniques using full CTG dataset with 21 attributes and reduced CTG dataset with 5 attributes is shown in Figure.

From the table, it can be inferred that with feature extraction, **Fine Gaussian Support Vector Machine** is the most accurate, having an accuracy of **97.3%**.

However, without feature extraction, it can be inferred that **Complex Tree Classifier** is the most accurate, having an accuracy of **75%**. Comparing the two, it can be inferred that feature extraction yields the most accuracy.

IV. CONCLUSION

Nine classification techniques (CT ,MT ,ST ,LSVM, QSM, CSVM, FGSVM, MGSVM, CGSVM) are compared using original CTG dataset with 21 attributes. The DT classifier yields more performance than other classifiers in terms of accuracy. For improving the accuracy, feature selection method is applied. Here, Gain Ratio feature selection method is applied to the original CTG dataset to obtain reduced CTG dataset that has 15 attributes. The performance is analysed using the classification techniques between original CTG dataset with 21 attributes and reduced CTG dataset with 15 attributes. The classification with reduced CTG data set yields more performance than original dataset. In future, other classification methods and feature selection techniques can be applied to this dataset so that performance can be improved.

REFERENCES

- [1] UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml/datasets/Cardiotocography>
- [2] Antonia Costa, MD; Diogo Ayres-de-Campos, PhD; Fernanda Costa, MD; Cristina Santos, MS; Joao Bernardes, PhD, "Prediction of neonatal academia by Computer analysis of fetal heartrate and ST event sibnals" AJOG – American Journal of Obstetrics and Gynecology, 2009.
- [3] J. Bernardes, C. Moura, J. P. de Sa, and L. P. Leite, "The Porto system for automated cardiotocographic signal analysis," J.Perinat. Med., pp.61–65, 1991.
- [4] N. Krupa , M.A.MA, E.Zahedi, S. Ahmed and F.M.Hassan, "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine," BioMed. Eng. OnLine, Vol.10, No.61, 2011.
- [5] R. Mantel, H. P. van Geijn, F. J. Caron, J. M. Swartjes, E. E. van Woerden, and H. W. Jongsma, "Computer analysis of antepartum fetal heart rate and Detection of accelerations and decelerations," Int. J. Biomed. Comput., Vol.25, No.4, 1990, pp.273-286.
- [6] Mei-Ling Huang and Yung-Yan Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network", J. Biomedical Science and Engineering, Vol.5, 2012, pp.526-533.
- [7] G. Rooth, A. Huch, and R. Huch, "Guidelines for the use of fetal monitoring," Int. J. Gynaecol.Obstet., vol.25, 1987, pp.159-167.
- [8] ShahadNidhal, M. A. Mohd. Ali1 and Hind Najah, "A novel cardiotocography fetal heartrate baseline estimation algorithm", Scientific Research and Essays Vol.5, No.24, 2010, pp.4002-4010.
- [9] Sundar, C., M. Chitradevi and G. Geetharamani, "An Overview Of Research Challenges for Classification Of Cardiotocogram Data" Journal of Computer Science, Vol.9, No.2, 2013, pp.198-206.
- [10] SundarChinnasamy, ChitradeviMuthusamy andGeetharamaniGopal, "An Outlier Based Bi-Level Neural Network Classification System for Improved Classification of Cardiotocogram Data", Life Science Journal, Vol.10, No.1, 2013, pp.244-251.