Classification of Cardiotocography Data with WEKA

[1] Divya Bhatnagar, [2] Piyush Maheshwari

[1,2] Department of Computer Science and Engineering,
Sir PadampatSinghania University,
Bhatewar, Udaipur, Rajasthan, India

Abstract - Cardiotocography (CTG) records fetal heart rate (FHR) and uterine contractions (UC) simultaneously. Cardiotocography trace patterns help doctors to understand the state of the fetus. Even after the introduction of cardiotocograph, the capacity to predict is still inaccurate. This paper evaluates some commonly used classification methods using WEKA. Precision, Recall, F-Measrue and ROC curve have been used as the metric to evaluate the performance of classifiers. As opposed to some of the earlier research works that were unable to identify Suspicious and Pathologic patterns, the results obtained from the study in this paper could precisely identify pathologic and Suspicious cases. Best results were obtained from J48, Random Forest and Classification via Regression.

Keywords - *Cardiotocography, CTG, Fetal Heart Rate, Uterine Contractions, Data Mining, Classification, Data Analysis.*

1. Introduction

Data Mining has brought many new developments, methods, and technologies in the recent decade. Also the improvement of integration of techniques and the application of data mining techniques had contributed in handling of new kinds of data types and applications. However, the field of data mining and its application in medical domain is still young enough so that the possibilities of the application are still limitless. The major challenges in medical domain are the extraction of intelligible knowledge from medical diagnosis data such as Cardiotocography CTG data. The use of classification and recognition systems has improved with effectiveness to help medical experts in diagnosing diseases.

This paper analyses and generates classification rules from cardiotocography data to identify normal, suspicious and pathologic cases. Section 2 presents the problem and section 3 deals with the basic data mining, Cardiotocography and classification. In section 4, an elaborate survey on Cardiotocography data classification has been presented. Section 5 presents the methodology adopted for analysis and classification of Cardiotocogram Data. The results have been presented in section 6. Section 7 concludes with the scope for future research..

2. Problem Definition

Cardiotocography (CTG) is a simultaneous recording of Fetal Heart Rate (FHR) and Uterine Contractions (UC) and it is one of the most common diagnostic techniques to evaluate maternal and fetal well-being during pregnancy and before delivery. By observing the Cardiotocography trace patterns doctors can understand the state of the fetus. There are several signal processing and computer programming based techniques for interpreting a typical CTG data. Even a few decades after the introduction of Cardiotocography in clinical practice, the predictive capacity of these methods remains controversial and still inaccurate. FHR patterns are observed manually by obstetricians during the process of CTG analyses. For the last three decades, great interest has been paid to the fetal heart rate baseline and its frequency analysis, Fetal Heart Rate (FHR) monitoring remains as a widely used method for detecting changes in fetal oxygenation that can occur during labor.

Yet deaths and long-term disablement from intrapartum hypoxia remain an important cause of suffering for parents and families, even in industrialized countries. Confidential inquiries have highlighted that as much as 50% of these deaths could have been avoided because they were caused by non-recognition of abnormal FHR patterns, poor communication between staff, or delay in taking appropriate action. Computation and other data mining techniques can be used to analyze and classify the CTG data to avoid human mistakes and to assist doctors to take a decision.

3. Related Terms

3.1 Cardiotocography

Cardiotocography is a medical test conducted during pregnancy that records FHR and UC. Tests may be conducted by either internal or external methods. In internal testing, a catheter is placed in the uterus after a specific amount of dilation has taken place. With external tests, a pair of sensory nodes is affixed to the mother's stomach. CTG trace generally shows two lines. The upper line is a record of the fetal heart rate in beats per minute. The lower line is a recording of uterine contractions from the TOCO. Uterine contractions, Four fetal heart rate features-Baseline heart rate, Variability, Accelerations, Decelerations. Uterine contractions are quantified as the number of contractions present in a 10 min period and averaged over 30 min. Normal: ≤ 5 contractions in 10 min and High: ≥ 5 contractions in 10 min represent uterine tachysystole. Baseline heart rate is the average baseline fetal heart rate. Reassuring feature: 110 – 160 beat per minute (bpm), Non-reassuring feature: 100 – 109 bpm OR 161 – 180 bpm and Abnormal feature: < 100 bpm OR > 180 bpm. Variability is the fluctuations in the fetal heart rate this causes the tracing to appear as a jagged, rather than a smooth, line. Variability is indicative of a mature fetal neurologic system and is seen as a measure of fetal reserve. Reassuring feature: ≥ 5 bpm, Non reassuring feature: < 5 bpm for ≥ 40 minutes but < 90 minutes and Abnormal feature: < 5 bpm for > 90 minutes. Decelerations are decreases in fetal heart rate from the baseline by at least 15 beats per minute, lasting for at least 15 seconds. There are three types of decelerations, depending on their relationship with uterine contraction. Early deceleration begins at the start of uterine contraction and ends with the conclusion of contraction. It is due to increased vagal tone due to fetal head compression. Variable deceleration occurs at any time irrespective of uterine contractions. It is due to the umbilical cord compression. Late deceleration begins at or after the peak of a contraction and ends long after it, hence “late” when compared to early decelerations. Reassuring feature: No deceleration, Non reassuring feature: Early deceleration, variable deceleration or single Prolonged deceleration up to 3 minutes and Abnormal feature: Atypical variable decelerations, late deceleration or single prolonged deceleration greater than 3 minutes. Three categories of CTG traces are as follows:

Normal trace: Tracings with all four features:

Baseline rate 110-160 bpm, Normal variability, Absence of decelerations, and Accelerations (may or may not be present). Suspicious trace: Tracing with ONE non reassuring feature and the other three are reassuring. Pathological trace: Tracing with TWO or more non reassuring features or ONE or more abnormal feature.

3.2 Data Mining

With the internet age the data and information explosion have resulted in the huge amount of data. Fortunately to gather knowledge from such abundant data, there exist data mining techniques. The data mining is - Extraction of interesting, non-trivial, implicit, previously unknown and potentially useful patterns or knowledge from huge amount of data. Data mining is the process of discovering patterns in large data sets. The overall is to extract information from a data set and transform it into an understandable structure for further use. Data mining has been used in various areas like Health care, business intelligence, financial trade analysis, network intrusion detection etc. General process of knowledge discovery from data involves data cleaning, data integration, data selection, data mining, pattern evaluation and knowledge presentation. Data cleaning, data integration constitutes data preprocessing. Here data is processed so that it becomes appropriate for the data mining process. Data mining forms the core part of the knowledge discovery process. There exist various data mining techniques viz Classification, Clustering, Association rule mining etc.

3.3 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[5]. These parameters guide the discrimination process. The unsupervised classification is also known as classification without the teacher. This classification uses, in most cases, the methods of cluster analysis. The device that performs the function of classification is called classifier. The classifier is the system containing several inputs that are transported with signals carrying information about the objects. The system generates information about the competence of objects into a particular class on the output. In a supervised learning scenario, a training data set of records is accompanied by class labels. New data can be classified based on the training set by generating descriptions of the classes. In addition to the training set, there is also a test data set which is used to determine the effectiveness of a classification. In principle, the popular classification method can be trained to recognize the data directly. High dimensionality of the training data consumes a lot of time and the accuracy of classification varies with the increase of dimensions in the training data. Data collection streams are broadening. The number of variables of concern to modelers has increased by at least an order of magnitude. Traditional methods were not designed to work with one hundred or more variables. The classification algorithm has been successfully applied in a variety of settings including direct marketing, intelligence and process control.

Precision: which is defined as proportion of instances that are truly of a class divided by the total instances classified as that class.

$$\text{Precision} = \frac{t_p}{t_p + f_p} \quad (1)$$

where, tp= No. of examples predicted positive that are actually positive and fp = No. of examples predicted positive that are actually negative.

Recall: Recall is defined as proportion of instances classified as a given class divided by the actual total in that class. Recall means how complete the results are.

$$\text{Recall} = \frac{t_p}{t_p + f_n} \quad (2)$$

where f_n is No. of examples predicted negative that are actually positive.

F-measures: It is a measure that combine recall and precision which is given as below:

$$\text{F} - \text{measure} = \frac{(2 * \text{Recall} * \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (3)$$

Mean Absolute Error: The MAE measures the average magnitude of the errors in a set of forecasts. It measures accuracy for continuous variables.

$$\text{Mean Absolute Error(MAE)} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i| \quad (4)$$

where, fi is prediction value, $y_i$is true value

Kappa Statistic: It measures the agreement of prediction with the true class, formulated as given below:

$$k = \frac{p_o - p_e}{1 - p_e} \quad (5)$$

where, p_o is the relative observed agreement, p_e is the hypothetical probability of chance agreement

Accuracy: It is a measure to determine the utility of the dataset [17].

$$\text{Accuracy} = \frac{\text{Correctly classified instances}}{\text{Total number of Instance}} \text{x } 100 \quad (6)$$

ROC Curve: Receiver Operating Characteristic (ROC) Curve is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a diagnostic test. The area under the curve is a measure of text accuracy.

4. Literature Survey

Nidhal et. al. [1] proposed a new algorithm for to estimate baseline based on digital CTG using Mathlab programming. The baseline values were detected indicating fetal status and health condition. The results showed slight difference with the experts Opinion. Advance classification techniques were suggested for improved outcome in future.

Costa et.al. [2] evaluated the accuracy of computer analysis of fetal heart rate and ST event signals in prediction of neonatal academia. Based on automated analysis of FHR and ST event signals, FHR tracings were evaluated to identify red alerts provided by the system, and compared with the occurrence of umbilical artery acidemia (pH <= 7.05). It was The authors concluded that the computer analysis of FHR and ST event signals provide higher accuracy in predicting neonatal acidemia.

Huang et.al. [3] evaluated fetal distress with discriminant analysis, decision tree, and artificial neural network. The results show that the accuracies of DA, DT and ANN are 82.1%, 86.36% and 97.78%, respectively. The authors suggested to use the classification techniques to fit in discrete attributes and applying feature selection technique in data preprocessing stage.

Geetaramani et.al. (2012) analysed the performance neural network based classification model with CTG dataset. The performance of the supervised machine learning based classification approach provided significant performance. The ANN based classifier was capable of identifying Normal, Suspicious and Pathologic condition, from the nature of CTG data with very good accuracy. ANN based classifier provided excellent performance in terms of Rand Index, Precision, Recall and F-Score. It was capable of identifying Normal and Pathologic condition with almost equal accuracy. The performance to identify the Suspicious CTG pattern was poor as compared to other two classes.

Chitradevi. et. al.(2012) evaluated the performance of k-mean clustering method with respect to four different Precision, Recall, F-Score and Rand Index. The paper considered class-wise Precision, Recall and F-Score to make the analysis very specific. If we consider only the precision as a metric, then arrived results proves that, even though the traditional clustering methods can distinguish the Normal CTG patterns from the Suspicious and Pathologic patterns with respect to precision, but, they were incapable of distinguishing Suspicious and

Pathologic patterns. The paper suggested to consider machine learning based method to design the CTG data classification system and address hybrid models using statistical and machine learning techniques for improved classification accuracy.

Sunder et.al. (2013) proposed a model based CTG data classification system using a supervised Artificial Neural Network (ANN) which can classify the CTG data based on its training data. The performance of neural network based classification model has been compared with the most commonly used unsupervised clustering methods Fuzzy C-mean and k-mean clustering. The results obtained showed that the performance of the supervised machine learning based classification approach provided significant performance than other compared unsupervised clustering methods. The traditional clustering methods could identify the Normal CTG patterns, but were incapable of finding Suspicious and Pathologic patterns.

Karabulut et.al. (2014) analyzed CTG data by an ensemble approach of adaptive boosting (AdaBoost) and investigated its effect for perfect determination of fetal distress from CTG data in this study. The study confirmed that ensemble machine learning approaches performed better than single classifiers.

5. Methodology

5.1 Process of Knowledge Discovery

Step involved in the process of knowledge discovery to achieve insight from CTG data is as follows:

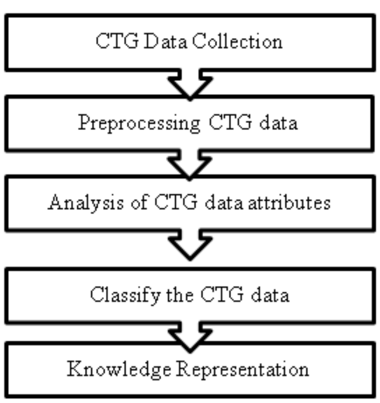

Fig 1 Process of Knowledge Discovery from CTG data

5.2 Data Description

The cardiotocography data set used in this study is publicly available at “The Data Mining Repository of University of California Irvine (UCI)”. 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. Attribute information is given as:

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).

5.3 Classification

It is a three class classification problem. The three classes are:

Normal - A CTG where all four features fall into the reassuring category, Suspicious - A CTG whose features fall into one of the non-reassuring categories and the reassuring category and the remainder of features are reassuring and Pathological - A CTG whose features fall into two or more of the Non-reassuring the reassuring category or two or more abnormal categories.

The classifiers implemented are as follows:
Naive Bayesian Classifier: This classifier is based on Bayes' theorem with independence assumptions between predictors. It is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. It is often does surprisingly well and is widely used because it often outperforms more sophisticated classification methods.

Decision Tree: This classifier is expressed as a recursive partition of the instance space. It consists of nodes that form a rooted tree, meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves. In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values. In WEKA, J48 is the class for generating a pruned or unpruned C4.5 decision tree.

Random Forest: It is a class for constructing a forest of random trees.

JRIP: This class implements a propositional rule learner, Repeated Incremental Pruning to Produce Error Reduction (RIPPER), which was proposed by William W.

Classification Via Regression: It uses regression methods. Class is binarized and one regression model is built for each class value.

Multi Layer Perceptron: It uses backpropagation to classify instances.

Attribute Selected Classifier: Dimensionality of training and test data is reduced by attribute selection before being passed on to a classifier.

5.4 Experimental Setup

WEKA 3.6 was used for implementing and evaluating the framework for the analysis and classification of fetal cardiotocography. WEKA stands for Waikato Environment for Knowledge Analysis[15].

6. Results

Four classification models were evaluated on randomly selected 1298 samples having 963 normal, 253 suspicious and 82 pathologic with respect to MAE, kappa statistics and accuracy. Table 1. shows the performance of various classifiers. Table 2, 3, 4, and 5 shows values for Precision, Recall, F-measure, and ROC Curve.

Table 1: Performance of Classifiers

Algorithm	**MAE**	**Kappa Statistics**	**Accuracy**
J48 Classifier	0.0408	0.8716	94.33 %
JRIP Classifier	0.0595	0.8339	92.74 %
Naive Bayes	0.1167	0.6221	82.31 %
Random Forest	0.0776	0.8423	93.20 %
Classification Via Regression	0.0652	0.8606	94.45%
MLP	0.063	0.82	93.22
Attribute Selected Classifier	0.05	0.85	94.22%

Table 2: Analysis of performance of Classifiers – Precision

Algorithm	**Precision**		
	Suspicious	Normal	Pathologic
J48	0.814	0.981	1.000
JRIP	0.804	0.959	1.000
Naïve Bayes	0.563	0.951	0.667
Random Forest	0.813	0.962	1.000
Classification Via Regression	0.873	0.962	0.944
MLP	0.882	0.954	0.818
Attribute Selected Classifier	0.884	0.954	0.961

Table 3: Analysis of performance of Classifiers – Recall

Algorithm	**Recall**		
	Suspicious	Normal	Pathologic
J48	0.933	0.962	0.800
JRIP	0.876	0.953	0.829
Naïve Bayes	0.798	0.864	0.514
Random Forest	0.876	0.968	0.743
Classification Via Regression	0.842	0.982	0.817
MLP	0.767	0.980	0.878
Attribute Selected Classifier	0.814	0.980	0.890

Table 4: Analysis of performance of Classifiers – F-measure

Algorithm	F-Measure		
	Suspicious	Normal	Pathologic
J48	0.869	0.971	0.889
JRIP	0.839	0.956	0.906
Naïve Bayes	0.660	0.906	0.581
Random Forest	0.843	0.965	0.852
Classification Via Regression	0.857	0.972	0.944
MLP	0.820	0.967	0.847
Attribute Selected Classifier	0.848	0.967	0.924

Table 5: Analysis of performance of Classifiers – ROC Curve

Algorithm	ROC Area		
	Suspicious	Normal	Pathologic
J48	0.929	0.954	0.962
JRIP	0.916	0.935	0.959
Naïve Bayes	0.892	0.951	0.954
Random Forest	0.967	0.980	0.996
Classification Via Regression	0.971	0.988	0.985
MLP	0.945	0.970	0.988
Attribute Selected Classifier	0.929	0.955	0.963

The classifiers were evaluated by 10 fold cross validation. It is clear from the values for area under ROC Curve, that the accuracy in all cases except one lies above 0.90 which is considered excellent as per the traditional academic point system to classify the accuracy of a diagnostic test. The attributes selection method selected the six attributes: AC, DP, ASTV, MSTV, ALTV and Min.

7. Conclusion

Earlier, the research works analyzed the same data and observed that maximum accuracy is achieved from ANN as 92.42%. The results obtained from classification of CTG in this paper indicates that the most promising results are received from decision tree based algorithm (J48) with 0.0408 as MAE, 0.8716 as kappa statistics and 94.33% as accuracy with the highest value for precision metric. Random forest and Classification via Regression were close to J48. Area under the ROC curve indicated the maximum accuracy of the three classifiers J48, Random Forest and Classification via Regression.

In future other data mining techniques can be applied for getting more accurate results especially applying on attribute selected data. Future works may also address hybrid models for improved classification accuracy.

References

[1] Shahad Nidhal, M. A. Mohd. Ali and Hind Najah. 2010. A novel cardiotocography fetal heart rate baseline estimation algorithm. Scientific Research and Essays Vol. 5(24), pp. 4002-4010

[2] Antonia Costa, Diogo Ayres-de-Campos, Fernanda Costa, Cristina Santos, Joao Bernardes. 2009. Prediction of neonatal acidemia by computer analysis of fetal heart rate and ST event signals. American Journal of Obstetrics and Gynecology

[3] Huang, M. and Hsu, Y. (2012) Fetal Distress Prediction Using Discriminant Analysis, Decision Tree, and ArtificialNeural Network. Journal of Biomedical Science & Engineering, **5**, 526-533.

[4] Sunder, C., Chitradevi, M. and Geetharamani, G. 2012 Classification of Cardiotocogram Data Using Neural NetworkBased Machine Learning Technique. International Journal of Computer Applications, 47, 19-25.

[5] Sunder, C., Chitradevi, M. and Geetharamani, G. 2012 An Analysis on the Performanace of K-Means Clustering Algorithm for Cardiotocogram Data Clustering. International Journal on Computational Sciences & Applications (IJCSA), 2(5), 11-20.

[6] Sundar, C., Chitradevi, M. and Geetharamani, G. 2013 An Overview of Research Challenges for Classification of Cardiotocogram Data. Journal of Computer Science, 9, 198-206

[7] Yılmaz, E. and Kılıkçıer, Ç. 2013 Determination of Fetal State from Cardiotocogram Using LS-SVM with ParticleSwarm Optimization and Binary Decision Tree. Computational and Mathematical Methods in Medicine, 2013.

[8] Newman, D.J., Heittech, S., Blake, C.L. and Merz, C.J. 1998. UCI Repository of Machine Learning Databases. University California Irvine, Department of Information and Computer Science.

[9] Kuncheva L. 2004 Combining Pattern Classifiers Methods and Algorithms. Wiley-Interscience, 360.

[10] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. The WEKA Data Mining Software: AnUpdate. SIGKDD Explorations.

[11] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy, "Advances in Knowledge Discovery and Data Mining", AAAI Press/ The MIT Press, 1996.

[12] Julie M. David and Kannan Balakrishnan, “Significance of Classification Techniques in Prediction of Learning Disabilities”,

[13] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[14] Kamel E., www.dremeilkamel.com.au /patient-information/obstetrics/cardiotocography-ctg.
[15] http://www.cs.waikato.ac.nz/ml/weka

Divya Bhatnagar is working as Professor in the department of Computer Science and Engineering in School of Engineering, Sir Padampat Singhania University, Udaipur, India. She holds 18 years of teaching and research experience. Her specialization areas include data mining and Neural Networks.

Piyush Maheshwari has completed B. Tech. and M. Tech. in Computer Science and Engineering from Sir Padampat Singhania University, Udaipur, India. Udaipur, Rajasthan. His area of interest is data mining.