

Automated Diagnosis and Cause Analysis of Cesarean Section Using Machine Learning Techniques

Ayesha Sana, Saad Razzaq, and Javed Ferzund

Abstract—Machine learning techniques provide learning mechanism that can be used to induce knowledge from data. A few studies exist on the use of machine learning techniques for medical diagnosis, prediction and treatment. In this study we evaluate different machine learning techniques for birth classification (cesarean or normal). Data on cesarean section is collected and different medical factors are identified that result in cesarean births. A birth classification model is built using decision tree and artificial neural networks. It can classify the births into normal and cesarean with an average accuracy, precision and recall of 80%, 85% and 84% respectively. Association rule mining is used to extract disease patterns from the collected data. It highlights the important medical factors that are associated with cesarean births.

Index Terms—Birth classification, cesarean section, machine learning, medical diagnosis.

I. INTRODUCTION

Knowledge Engineering (KE) and Machine Learning (ML) make the patterns extraction more convenient. Data can be classified and prediction can be made using variety of ML algorithms. ML facilitates tools, methods and techniques that can provide diagnosis and analytical facilities in the number of medical domains. These are used for clinical parameters and predictive analysis for example disease prediction, extraction of medical knowledge for conclusion study, therapy dealing and help in patient management. A reasonable amount of work exists in the domain of biomedical, biomedicine and diagnosis of diseases using different machine learning techniques. However, there are many areas in the medical domain that are not yet explored using ML/KE techniques.

Due to elective cesareans rate of cesarean delivery has increased dramatically over the past decade. [1]. There is a need to find its causes and control the rate of C-section. According to UK National Health Service, the chance of death in cesarean section is 3 to 4 times higher than the vaginal delivery [2]. According to mother's survey most of the women who have cesarean section in previous cases would have to like vaginal birth but did not have it. The reason for that, the physician of health does not take the risk of vaginal birth after cesarean section. There are a lot of complications seemed to a child and mother after C-Section. Like in women, chance of uterine infection is ten times

higher than vaginal infection. There are anesthesia problems and respiratory problems for the new born. Children delivered by cesarean are more likely to asthma, allergies and diabetes. The risk of asthma is 79% and diabetes 20% significantly more than the vaginal delivery [3]. In order to avoid these problems, there is a need to create awareness among people about its causes. It does not only affect the mother's health but also the new born. This study aims at finding the reasons for an increased rate of cesarean section and developing a prediction model for child birth. The objectives of this study are as follows:

- To identify the significant factors influencing the type of birth.
- To build a Birth Prediction Classification (BPC) model based on the identified factors.
- To extract association rules among the identified factors and find the disease patterns.

Identification of the causal factors for cesarean births will help the doctors and patients to take preventive measures for C-Section. The BPC model will be used to classify the birth into normal and cesarean delivery. Birth type prediction will be obtained for any patient by providing the necessary parameters. So that necessary actions can be taken during pregnancy to avoid the C-Section. Extraction of association rules will help to identify the birth patterns.

The remainder of this paper is organized as follows: In section 2 we discuss the related work. The details of the experiment are presented in section 3. We discuss the results in section 4 and conclude the paper in section 5. Finally we discuss the future directions in section 6.

II. RELATED WORK

In the recent years researchers of machine learning techniques have shown interest in the medical field. Research in biomedicine, genetics and medical diagnostics involves a variety of machine learning techniques. These techniques have been used for comparison, prediction and predictive analysis. We discuss a few studies here that are relevant to our work. Eken C et al. [4] presented a model on kidney pains using artificial intelligence techniques. The authors prepared a datasheet and used genetic algorithms (GA) and ANN for predictive analysis. They concluded that data mining techniques like ANN and GA can be used for predictive analysis in emergency of suspicious kidney stone pains.

Gholami B et al. [5] used vector machine learning techniques for the evaluation of pain in patients who could not converse orally. This was beneficial to the medical professionals to check the pain of patient that could not tell verbally.

Manuscript received May 25, 2012; revised August 30, 2012.

The authors are with the Computer Science Department, University of Sargodha, Sargodha, Pakistan (e-mail: asana@uos.edu.pk; saadrzzaq@uos.edu.pk; jferzund@uos.edu.pk).

Gholipour et al. [6] collected the retrospective data of laparoscopic cholecystectomy (gallbladder) (LC) to open surgery from patients and applied ANN and linear discriminate analysis to check future ability of conversion. The authors found that LC to open surgery was better predictable with preoperative health characteristics of patients using ANN.

Guan P et al. [7] improved the accuracy of cancer detection. The authors used a database consisting of lungs cancer data sets and cancer related gene. Support vector machine, K nearest neighbor classifier and ANN were used in this study. The authors claimed an accuracy of 98% to 100% for the training set.

Thomas S et al. [8] developed a prediction model for antimicrobial peptides (AMPs). The model was developed using the free online database of AMP. The authors validated the data and used it for prediction of AMPs. They used machine learning algorithms like SVM, Random Forest and Discriminate analysis or predictions.

Wang J et al. [9] used ANN and decision tree for classification of cancer patients. The authors collected data from 2005 to 2007 of 1008 gastric cancer patients from one hospital of china. They used this data to train artificial neural network with one hidden layer on three hidden nodes and decision tree models. The efficiency of ANN and Decision Tree was compared for patient classification. The authors found that ANN performed better in predicting hospital gastric cancer patient than the decision tree. We also use ANN and decision tree in our study but we apply these techniques on a different data set

III. THE APPROACH

Predictive analysis can be made when historic data is available. Data can be integrated from multiple sources. It should be cleaned and transformed to make it suitable for the application of ML techniques. The architecture of our BPC model is depicted in Fig. 1. In the first step data is obtained by interviewing patients, doctors and the hospital sources. Data from the three sources is combined into a single data pool. This data pool is cleaned and transformed to obtain the

refined data set. Then ML techniques are applied on this data to obtain the classification and the association rules. We have collected data from 15 hospitals of Sargodha, Pakistan between 2nd February 2010 to 28th February 2010, by interviewing the medical experts in this area and the patients. We have identified about 50 factors that can influence the type of birth. These factors include Pre pregnancy factors like maternal age, body weight, education, drinking routine, diabetes, hypertension and various other factors given in Table I. Some factors are identified during pregnancy like HIV, uterine rupture, blood sugar, abnormal presentation and many more mentioned in Table I. There are also some social factors like low education, dieting, fear of pain etc. We have applied the chi square method of independence on the dataset created in Table I to find out the dependence relationship between the birth type and various attributes. This test will help to reduce the number of variables to be used in the BPC model. It will filter out the attributes that have no influence on birth type.

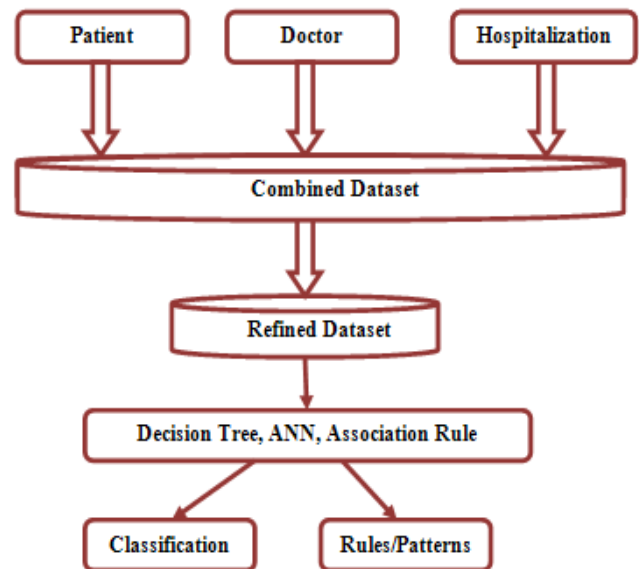


Fig. 1. Architecture of Birth Prediction Model

TABLE I: FACTORS FOR CESAREAN SECTION

Sr. No.	Pre-Pregnancy	During-Pregnancy	Medical	Social
1.	Maternal age	HIV	Repeat C-section	Low priority
2.	Weight	Multiple births	Cesarean hysterectomy	Low Education
3.	Education	Abnormal presentation	Complications in pregnancy	Limited awareness
4.	Drinking	Overly large baby	Failure of labor	Patient insurance
5.	Diabetes	Umbilical cord abnormalities	Prolonged Labor	Physician practice
6.	Hypertension	Hypertension(At time of delivery)	Umbilical cord prolapsed	Dieting to look smart
7.	Asthma	Amount of protein in urine	Uterine rupture	Fear of pains
8.	Operated Organ	Weak woman	Precious (high risk fetus)	Psychologically pressure
9.	Abortion	Less Healthy food	Blood sugar	Socio-economic Status
10.	Hepatitis	Before expected date	Exercise	Over fattening food for healthy baby
11.	Smoking	Breech	Multi Vitamin	Infection
12.	Age	After Expected Date	Iron tablets	Poverty

BPC model is applied on the collected dataset. BPC includes the decision tree and Multilayer Perceptron is used for the classification. We use these two algorithms as they are the best classifying models [10]. Association rules are extracted by using the Apriori algorithm. These association rules help us in determining the major factors that causes Cesarean birth.

We have selected decision tree due to its successful applications in many domains. A decision tree consists of a set of nodes forming a tree, where each internal node represents a test on an attribute and the branches represent the outcome of the test. Classes are represented by leaf nodes. It is suitable for inductive inference and can also be represented as if-then rules [10].

Artificial Neural Networks simulate human brain and consist of a set of nodes called neurons. ANNs provide a learning mechanism to perform a task given the initial data.

Advantage of using ANN are that they are fault tolerant and can make its own representation of information. ANNs have been successfully used in engineering and financial applications [10].

Association rule mining can identify patterns from the given data. Apriori is the commonly used algorithm for extracting association rules. It can transform the patterns that have sufficient support into a rule. Using Apriori we can find associations among different attributes [10].

IV. RESULTS AND DISCUSSION

We applied Chi Square on the processed data in order to know whether the attributes under question have an influence on the type of birth. In other words, we wanted to know whether child birth is independent of the factors mentioned in the previous section. Chi Square of independence can tell us whether the difference between observed and expected outputs is by chance or due to some other factors. We have applied Chi Square test on pairs, with each pair have one attribute of birth and the other is selected from the list of attributes. For each pair of attributes, we have defined two hypotheses that are as follows:

H0: The selected attributes are independent of each other.

H1: The selected attributes are not independent of each other, i.e. the birth attribute is influenced by the other attribute.

There are two types of the birth attribute, normal and cesarean. Our data consists of 40% normal birth and 60% cesarean. We have chosen 0.05 as the level of significance to prove the Null hypothesis. The results of each Chi Square test are presented in Table II. Our Null hypothesis is accepted only in case of working, indicating that working or not-working does not influence the type of birth. For the remaining attributes alternate hypothesis is accepted. It means that birth type is influenced by age, income, condition, blood pressure, pulse surgery, exercise, pre-pregnancy, food and use of multivitamins. Decision tree is used to classify the births into normal and cesarean. We used J48 algorithm that is implemented in WEKA. According to this tree women having exercise during pregnancy with no previous surgery and normal blood pressure (BP) have normal delivery.

TABLE II: CHI SQUARE TEST OF INDEPENDENCE

Attribute	Chi Square	DOF	P Value	Dependency
Working	0.34	1	0.56	No
Age	1.98	4	0.74	Yes
Income	2.02	2	0.365	Yes
Condition	8.96	2	0.012	Yes
Blood pressure	8.28	2	0.016	Yes
Pulse	2.85	1	0.092	Yes
Surgery	2.02	1	0.156	Yes
Exercise	9.63	1	0.002	Yes
Pre-pregnancy	2.8	1	0.95	Yes
Food	5.93	2	0.069	Yes
Multivitamin	6.3	1	0.012	Yes

All of the women having a previous surgery got cesarean section. Most of the women, who did not use multivitamin during pregnancy or did not take exercise, got cesarean section. To evaluate the classification results of J48, we have calculated the precision and recall values for each class i.e. normal and cesarean. In Table III the precision for normal class is 77% and the recall is 72%. The precision for cesarean class is 82% and the recall is 86%. It indicates that this classifier can be successfully used to predict the type of birth before delivery based on the factors. Recall is the probability that a relevant document is retrieved in a search. Precision is the probability that retrieved document is relevant. A receiver operator characteristic (ROC) shows how much the data assures the true positive to get aside the false positive (FP). ROC is 0.82 in calculated data.

TABLE III: ACCURACY BY DECISION TREE

Class	Precision	Recall	F-Measure	ROC Area
Normal	0.778	0.718	0.747	0.816
Cesarean	0.817	0.86	0.838	0.816
Weighted Average	0.801	0.802	0.801	0.816

The second classifier is Multilayer Perceptron algorithm. Ten fold cross validation is used to train and test the model. As shown in Table IV the precision and recall for the normal class is 76% and 82% respectively. For the cesarean class, the precision and recall is 87% and 82% respectively. The precision and recall values indicate that ANN can be successfully used to predict the type of birth before delivery. F-measure is 0.85 in the given data. ROC area is 0.90 in calculated data.

TABLE IV: ACCURACY BY MULTILAYER PERCEPTRON

Class	Precision	Recall	F-Measure	ROC Area
Normal	0.76	0.821	0.79	0.90
Cesarean	0.87	0.825	0.847	0.90
Weighted Average	0.826	0.823	0.824	0.90

Association Rule mining is applied to extract the rules from the processed data. We used the Apriori algorithm implemented in WEKA. It can find the relationships among the attributes in the form of association rules. We have extracted those rules in which our class variable (birth) is a consequent. Apriori requires different parameters to be set to limit the number of rules extracted. These

parameters include confidence, coverage and lift. We present the top ten rules in Table V that are extracted from the collected data. It is evident from Table V that high blood pressure and pulse rate result in cesarean births. Low income of a family and the lack of education are also associated with cesarean births.

V. CONCLUSION

In this paper we have presented the successful application of machine learning techniques in the medical domain. Knowledge engineering and machine learning are used to extract disease patterns from the available medical data. The extracted patterns can be used for medical diagnosis, prediction and treatment. This paper has covered three goals: First, it has identified the significant factors that influence the type of birth. Second, it has presented a prediction model for the type of birth that can help doctors and patients. Third,

it has presented the association rules that show cesarean birth patterns. Decision Tree can classify between normal and cesarean births with 80% accuracy. Artificial Neural Network can classify between normal and cesarean births with 82% accuracy. Results of the Apriori association rule mining indicate that proper exercise is associated with normal births. Similarly use of multivitamins is also associated with normal births. High blood pressure and pulse rate are associated with cesarean births. Lack of education is also associated with cesarean births. These rules can be helpful to design a precautionary plan for cesarean cases. In short, this paper has highlighted the usefulness of machine learning in the medical domain. We used the cesarean section as a case study because its rate is dangerously increasing in Pakistan. Similarly machine learning techniques can be applied to study different diseases, their causes and occurrence patterns.

TABLE V: TOP TEN ASSOCIATION RULES

Sr. No.	Rule	Confidence	Lift
1	BP=high 17⇒Pulse=high Birth=cesarean 10	0.59	4.71
2	Working=no income=low 24⇒ Education=0 6 Temperature=no Birth=cesarean 10	0.42	4
3	Education=0 6 13⇒ Working=no income=low Birth=cesarean 10	0.77	4.34
4	Pulse=high 15 ⇒ BP=high Surgery=no Birth=cesarean 10	0.67	4.27
5	BP=high 17⇒ Pulse=high Birth=cesarean 10	0.59	4.71
6	Pulse=high Temperature=no Birth=cesarean 12 ⇒ BP=high 10	0.83	4.71
7	Education=0 6 13 ⇒ Working=no income=low Multivitamin=No 10	0.77	4.1
8	Exercise=no 54 ⇒ Birth=cesarean 39	0.72	1.22
9	Exercise=no 54⇒ age=25 31 Birth=cesarean 25	0.46	1.17
10	age=25 31 Exercise=no 30 ⇒ Birth=cesarean 25	0.83	1.4

VI. FUTURE WORK

We have obtained quite satisfactory results for classification of births. However these results can be further improved by identification of additional factors that influence the type of birth. We will investigate these factors in the future. In this work three machine learning techniques have been used. We want to evaluate other techniques on the medical data. The prediction models will be more useful if available online. We will make these models online so that they can be further trained on the new data available. We want to increase the area surveyed because there may be certain geographical factors that can influence the type of birth.

REFERENCES

- [1] S. R. Miesnik, B. Reale "A Review of Issues Surrounding Medically Elective Cesarean Delivery," *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, vol. 36, Issue 6, pp. 605–615, November/December 2007.
- [2] Romano A. M. Research summaries for normal birth. *Journal of Perinatal Education*. 2007;16(1): pp.37–40.
- [3] Symonds E, and Symonds I. Essential Obstetrics and Gynaecology 4th ed. Oxford: Churchill Livingstone; 2003.
- [4] C. Eken, U. Bilge, M. Kartal and O. Eray, "Artificial neural network, genetic algorithm, and logistic regression applications for predicting renal colic in emergency settings," *Int J Emerg Med*, 2009.
- [5] B. Gholami, W. M. Haddad, A. R. Tannenbaum, "Relevance Vector Machine Learning for Neonate Pain Intensity Assessment Using Digital Imaging," *IEEE Transactions on Biomedical Engineering*. June 2010.
- [6] C. Gholipour, B. A. Fakhree, R. A. Shalchi, and M. Abbasi, "Prediction of conversion of laparoscopic cholecystectomy to open surgery with artificial neural networks," *BMC Surgery*, vol. 9.
- [7] P. Guan, D. Huang, M. He and B. Zhou "Lung cancer gene expression database analysis incorporating prior knowledge with support vector machine-based classification method," *Journal of Experimental & Clinical Cancer Research* 28:103, 2009.
- [8] S Thomas, S Karnik, R Shankar B V. K. Jayaraman, and S Thomas. "CAMP: a useful resource for research on antimicrobial peptides," *Oxford Journals* vol. 38, no. 1 pp. D774–D780.
- [9] J. Wang, M. Li, Y. Hu, and Y. Zhu, "Comparison of hospital charge prediction models for gastric cancer patients: neural network vs. decision tree models," *BMC Health Services Research*, vol. 9.
- [10] Witten I. H, Frank E, Data Mining: Practical Machine Learning Tools and Techniques, 2nd ed. Morgan Kaufmann June 2005.