CHAPTER 1

INTRODUCTION

1.1 DATA MINING

Data Mining and the technology of Knowledge Discovery from Data (KDD) have brought many new developments, methods, and technologies in the recent decade. Also the improvements in the integration of these techniques and the application of data mining have contributed in handling 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.

One of the major challenges in medical domain is the extraction of comprehensible knowledge from medical diagnosis data such as Cardiotocography (CTG) data. In this information era, the use of machine learning tools in medical diagnosis is increasing gradually. This is mainly because the effectiveness in classification and recognition systems have improved to a great extend to help medical experts in diagnosing diseases.

1.2 CARDIOTOCOGRAPHY

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 datamining techniques can be used to analyze and classify the CTG data to avoid human mistakes and to assist doctors to take a decision.

1.3 CLUSTERING AND CLASSIFICATION

Clustering is a technique that has been widely studied and applied to many real-life applications. Many efficient algorithms, including the well known and widely applied k-means algorithm, have been devised to solve the clustering problem efficiently. Traditionally, clustering algorithms deal with a set of objects whose positions are accurately known. The goal is to find a way to divide objects into clusters. Classification process may be applied in different areas of research and practice, e.g. farms, military, medicine, remote sensing, etc. The classical classification techniques use statistical approach, which typically assumes the normal multidimensional distribution of probability in the experimental data set. Data classification may be supervised or unsupervised (Klimesova et al 2010).
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, which are values vector, covariance matrix), which are computed from the training set. These parameters guide the discrimination process. The Bayesian classifiers are typical representatives (Bayes classifier, Fisher, Wald sequential) (Klimesova et al 2010).

The unsupervised classification is also known as classification without the monitoring of a teacher. In most of the cases, this classification uses the methods of cluster analysis. The device that performs the function of classification is called classifier. The classifier is a system containing several inputs that are transported with signals carrying information about the objects. The system generates information about the competence of objects into particular class on the output.

1.4 ABOUT THIS WORK

Cardiotocography (CTG), consisting of Fetal Heart Rate (FHR) and Tocographic (TOCO) measurements, is used to evaluate fetal well being during the delivery. Since 1970 many researchers have employed different methods to help the doctors, to interpret the CTG trace pattern from the field of signal processing and computer programming (Shahad Nidhal et al 2010). They have supported doctors with interpretations in order to reach a satisfactory level of reliability so as to act as a decision support system in obstetrics. Up to now, none of them has been adopted worldwide for everyday practice (Van Geijnt 1996). There is currently no consensus on the best methodology for baseline estimation in computer analysis of cardiotocographs (Shahad Nidhal et al 2010). More than 30 years after the introduction of antepartum cardiotocography into clinical practice, the predictive capacity of the method remains controversial. In a review of lot of articles published on
this subject, it was found that the reported sensitivity varies between 2 and 100%, and its specificity between 37 and 100% (Diogo et al 2005). So, in this work, some of the statistical, machine learning and datamining techniques for the classification of CTG data are evaluated.

1.5 CONTRIBUTIONS OF THIS RESEARCH

In this research, an elaborate literature survey and background study was made on existing clustering and classification algorithms, outlier deductions and CTG data classification methods.

The Problems in Identifying Suspicious Record in CTG Data using Machine Learning Techniques was made and the results have been presented.

An Evaluation on Automated Classification Methods for the Classification of Cardiotocogram Data was made and the results have been presented.

An Outlier Based Bi-level Neural Network (BL-NN) Classification System for Improved Classification of Cardiotocogram Data has been presented.

An Outlier Based Bi-Model Neural Network (BM-NN) Classification System for Improved Classification of Cardiotocogram Data has been presented.

A PCA based Improved Bi-Model Neural Network (IBM-NN) Classification System for Improved Classification of Cardiotocogram Data has been presented.
1.6 ORGANIZATION OF THE THESIS

Chapter 1 presents the problem and the outline of the contributions of the research. Chapter 2 deals with the basic datamining, Cardiotocography, clustering, classification and outlier deduction. In Chapter 3 an elaborate survey on Cardiotocography data classification has been presented. Chapter 4 presents The Problems in Identifying Suspicious Record in CTG Data using Machine Learning Techniques was made and the results have been presented. Chapter 5 presents an Evaluation on “Automated Classification Methods for the Classification of Cardiotocogram Data” was made and the results have been presented. Chapter 6 presents an Outlier Based Bi-level Neural Network (BL-NN) Classification System for Improved Classification of Cardiotocogram Data was made and the results have been presented. In Chapter 7, proposed an Outlier Based Bi-Model Neural Network (BM-NN) Classification System for Improved Classification of Cardiotocogram Data has been presented. In Chapter 8, proposed a PCA based Improved Bi-Model Neural Network (IBM-NN) Classification System for Improved Classification of Cardiotocogram Data has been presented and the last chapter 9 concludes with conclusion and scope for future research.