CLASSIFICATION OF FOETAL DISTRESS AND HYPOXIA USING MACHINE LEARNING

ROUNAQ ABDALRAHMAN ABBAS

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Master of Philosophy

2018

ACKNOWLEDGEMENTS

Firstly, I am grateful to Allah for the good health and wellbeing that were necessary to complete this dissertation.

I would like to express my sincere gratitude to my supervisor Dr.Abir Hussain for the continuous support of my M.PHIL study and related research, for her patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my M.PHIL study.

I would like to take the opportunity to express my sincere appreciations to Prof.Dhiya Al-jumeily, for his continuous support and motivating words, which helped me through this entire journey.

My great gratitude to my family and everyone who supported me and believed in me especially my mother and my husband, I don't believe I can finish this dissertation without their support.

ABSTRACT

Foetal distress and hypoxia (oxygen deprivation) is considered a serious condition and one of the main factors for caesarean section in the obstetrics and gynaecology department. It is considered to be the third most common cause of death in new-born babies. Foetal distress occurs in about 1 in 20 pregnancies. Many foetuses that experience some sort of hypoxic effects can have series risks such as damage to the cells of the central nervous system that may lead to life-long disability (cerebral palsy) or even death. Continuous labour monitoring is essential to observe foetal wellbeing during labour. Many studies have used data from foetal surveillance by monitoring the foetal heart rate with a cardiotocography, which has succeeded traditional methods for foetal monitoring since 1960. Despite the indication of normal results, these results are not reassuring, and a small proportion of these foetuses are actually hypoxic. This study investigates the use of machine learning classifiers for classification of foetal hypoxic cases using a novel method, in which we are not only considering the classification performance only, but also investigating the worth of each participating parameter to the classification as seen by medical literature. The main parameters that are included in this study as indicators of metabolic acidosis are: pH level (which is a measure of the hydrogen ion concentration of blood to specify the acidity or alkalinity), as an indicator of respiratory acidosis; Base Deficit of extra-cellular fluid level and Base Excess (BE) (which is the measure of the total concentration of blood buffer base that indicates metabolic acidosis or compensated respiratory alkalosis). In addition to other parameters such as the PCO2 (partial pressure of carbon dioxide can reflect the hypoxic state of the foetus) and the Apgar scores (which shows the foetal physical activity within a specific time interval after birth).

The provided data was an open-source partum clinical data obtained by Physionet, including both hypoxic cases and normal cases. Six well known machine learning classifier are used for the classification; each model was presented with a set of selected features derived from the clinical data. Classifier evaluation is performed using the receiver operating characteristic curve analysis, area under the curve plots, as well as confusion matrix. The simulation results indicate that machine learning classifiers provide good results in diagnosis of foetal hypoxia, in addition to acceptable results of different combinations of parameters to differentiate the cases.

TABLE OF CONTENTS

Contents

List of Tables

List of Figures

ABBREVIATIONS

ACOG	American college of Obstetricians and gynecologist criteria
AI	Artificial intelligence
ANNs	Artificial neural networks
AMIS	Hospital information system
AS1	Apgar score at 1st
AS5	Apgar score at 5th
ACC	Accuracy
AUC	Area under curve
BDecf	Base deficit
BE	Base excess
BP	Back-propagation
CART	Classification and regression trees
CS	Caesarean section
CTG	Cardiotocography
CO2	Carbone dioxide
EFM	Electronic Foetal monitoring
ECG	Electric-cardiogram
FN	False negative
FP	False positive
FHR	Foetal heart rate
FSS	Foetal scalp blood sampling
GBM	Stochastic Gradient Boosting

GBDTs	Gradient-boosted decision trees
HCO3	Bicarbonate
HIV	Human immunosuppressive virus
HIE	Hypoxic -ischemic encephalopathy
IA	Intermittent auscultation
IUGR	Intrauterine growth retardation
KDD	Knowledge discovery in databases
KNN	K-Nearest Neighbor
LSVM	Linear Support vector machine
ML	Machine learning
MLP	Multilayer Perceptron
NNET	Neural network
pH	Power of hydrogen, expressing the blood acidity
PCO2	Partial pressure of Carbone dioxide
PO2	Partial pressure of oxygen
PDA	Patent dectus arteriosus
PUBS	Percutaneous Umbilical Cord Blood Sampling
PCA	Principle component analysis
PVL	Periventricular leukomalacia
PE	Processing element
RBF	Radial basis function
ROC	Receiver operating characteristic
RF	Random forest
RDS	Respiratory distress syndrome

RCOG	The Royal College of Obstetricians and Gynaecologists
RL	Reinforcement learning
SOGC	The Society of Obstetricians and Gynaecologists of Canada
SSL	Semi supervised learning
SVM	Support vector machine
SVC	Support vector classifier
SMOTE	Synthetic Minority Over-sampling Technique
SE	Sensitivity
SP	Specificity
TPR	True positive rate
TNR	True negative rate
TN	True negative
TP	True positive
tSNE	T-distributed stochastic neighbor embedding
UC	Uterine contraction
UHB	University hospital in Brno

PUPLICATION

2018 International Conference on Intelligent Computing

August 15-18, 2018, Wuhan, China

Authors: Rounaq Abbas, Abir Hussain, Dhiya Al-Jumeily

Paper Title: Classification of Foetal Distress and Hypoxia using Machine Learning Approaches

The paper has been submitted and accepted to present by the 2018 International Conference on Intelligent Computing (ICIC2018).

CHAPTER ONE: INTRODUCTION

1.1 Overview

Foetal distress is a condition of foetal oxygen deprivation and accumulation of carbon dioxide, leading to hypoxia and acidosis during the ante-partum period (before labour) or intra-partum period (during the birth process) [1] as shown in Figure 1.1.

Foetal distress and hypoxia (oxygen deprivation) is considered as a serious conditions and one of the main reasons for caesarean section in the obstetrics and gynaecology department. Foetal distress occurs in about 1 in 20 pregnancies [2]. Foetal distress a severe condition that can lead to foetal death or brain damage. It is also considered to be the third most common cause of new-born death. Many of the foetuses have experienced hypoxia during different stages of the pregnancy period.

Foetal hypoxia can be classified as acute hypoxia or chronic hypoxia according to the stage of the intra-partum foetal life [2]. The former usually occurs during the labour process while the latter occurs during the first, second or the third trimester of the pregnancy. Various methods of intra-partum foetal surveillance have been employed to detect the signs of hypoxia as early as possible to minimize the risk of life-long disability such as cerebral palsy and to reduce the mortality rate among new-borns. Intra-partum monitoring of foetuses during labour has been commonly performed by monitoring the foetal heart with a technique such as the intermittent auscultation (IA), which is considered the most common method of foetal surveillance in labour [1]. However many of these methods have been subject to controversy as existing studies have found no benefit of its use in reducing the rates of cerebral palsy or peri-natal mortality [3].

Continuous labour monitoring is essential to observe the foetal wellbeing. There are many studies indicating foetal heart activity is the prominent source of information about foetal health and especially the detection of foetal hypoxia. However, the physicians have found many challenges in identifying the perfect way of detecting foetal hypoxia by analysing foetal heart rate using the traditional and intermittent monitoring such as (IA) and Doppler machine [1]. Since 1960, cardiotocography (continuous electronic foetal heart rate monitoring) has been developed and replaced all other traditional methods to monitor foetal heart. Normal cardiotocography results in nearly half of all tracings indicating that enough oxygen is delivered to the foetus [4, 5]. However; these results are not encouraging [6], as a number of these foetuses are actually hypoxic, therefore a diagnostic test is necessary and compulsory. During the last decades, new methods have been used by antenatal care and during the labour process such as the cordocentesis (foetal blood sampling by ultrasound guided needle aspiration from the umbilical cord) [7], which can be used to detect hypoxia due to placental development abnormalities. This could cause many threats to the foetal health such as damage to the cells of the central nervous system that lead to life-long disability. Up to three quarters of infants with severe hypoxic-ischemic encephalopathy (HIE) die of multiple organ failure or lung infections caused by irregular breathing. Those who survive are commonly left with gross symptoms such as mental retardation, epilepsy, and cerebral palsy [8, 9]. Saling technique for detecting hypoxia has been the ideal method using a sample of blood from the foetus's scalp during labour and analysing the pH value as an indicator [7]. Westgren et al. [10] used this technique in detecting the perinatal outcome, by using foetal scalp blood sampling to detect the pH value and lactate level, which may indicate hypoxia and identify which one will be the accurate indicator. They selected a "pH" value<7.20 as a cut-off value to recommend intervention, while the abnormal level of lactate in the foetal scalp blood was 3.08 mmol/L. The study findings suggest the measurement of the foetal scalp lactate levels is more useful than pH analysis due to the complexity of pH analysis, demanding of a relatively large amount of blood (30-50 µl), while 5 µl of blood was enough to detect the lactate level. Another reason for the undesirability of using pH as an indicator was the sampling failure rates of 11% that have been reported. The study of Westgren and his colleagues [10]concluded that using lactate level was more successful and accurate in predicting perinatal outcome. James et al. [11]concluded that umbilical cord blood gas analysis can give an indication of preceding foetal hypoxic stress. In this research work, Machine-learning (ML) classifiers have been used for the classification of foetal hypoxia. The relationships between the various parameters and their effect on hypoxic state detection were studied.

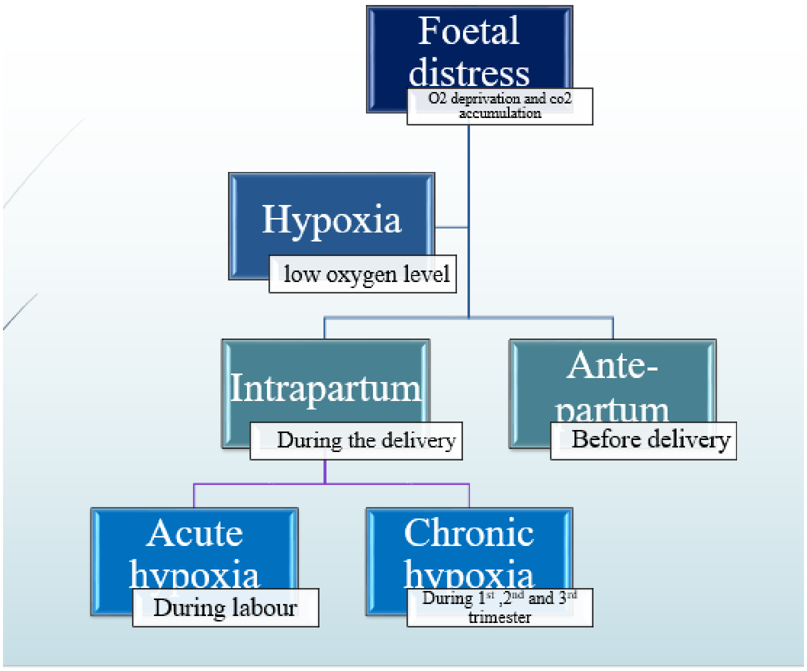

Figure 1-1 foetal distress stages

1.2 Problem statement

The early stages of Embryogenesis, foetal growth, and survival of the perinatal period all depend on optimal maternal health and normal placental development. Maternal exposure to a persistently hypoxic situation could lead to critical injury to vital organs. Failure of the normal placental function could have profound acute and chronic effects on the developing foetus and lead to many complications such as; intrauterine growth restriction (IUGR), asphyxia, multi-organ failure, premature delivery, and perinatal demise. In the United States, IUGR and prematurity complicate about 12% of the deliveries and represent the leading cause of perinatal mortality and morbidity to this day, accounting for up to 75% of perinatal deaths [12]. In addition to long-term disabilities such as cerebral palsy, hearing loss, retinopathies, and chronic lung disease, there are also the associated substantial emotional burdens for affected families and health care costs to the society as well machine learning
[12].

Many procedures have been used in the management and prevention of foetal hypoxia, however most of the diagnostic decisions that affect a foetus's health are conducted after the delivery and dependent widely on the physician's experience. Extensive researches indicate that there are high risks of wrong diagnosis due to poor investigation and irregular antenatal follow up for the mother during the pregnancy especially in the third world countries. The diagnosis of foetal asphyxia can be tricky as there are many abnormal signs and symptoms that will be developed by new-born babies but will not necessarily cause foetal complication or serious outcome.

Diagnosis of foetal hypoxia is still considered to be difficult because the consequences of hypoxia/acidosis are very different, depending on whether this is acute or chronic. The normal human foetus is adapted to survive labour and has compensatory mechanisms that allow it to withstand even severe hypoxia and acidosis for short periods of time. Several studies have looked at the neurological outcome of neonates who were severely asphyxiated at delivery and showed that the pH level can be highly sensitive to the results of the final diagnosis. Findings indicated that the predictive value of acidosis at birth for neurological sequelae, especially in term neonates, depending on the pH values alone is poor, while the long morbidity rate was much higher with the chronic type hypoxia [13].

Finding other parameters that can help in diagnosis and classification of neonatal cases using artificial intelligence (AI) methods was another challenge in this research. ML classifiers are used for this purpose, it has been trained using medical data sets of infants with both normal and hypoxic states, labelled by using the pH Level and BDecf (which are the measure of the total concentration of blood buffer base that indicates the metabolic acidosis or compensated respiratory alkalosis) as a threshold for both the respiratory and metabolic acidosis which are the main types of foetal hypoxia [14].

An alternative pathway to diagnose and manage foetuses delivered with hypoxic state is necessary to improve the neonatal care as well as to conquer the challenges facing the medical staff.

1.3 Aims and objectives

The aim of this study is to provide a robust and effective diagnostic support classifier to improve the diagnostic accuracy of foetal hypoxic infants using ML methods. Determine the parameters that affect the diagnostic decisions using computer science approaches. This could assist physicians in the clinical management and follows up of hypoxic neonates after delivery and prevents many serious medical complications.

The objectives of this thesis are as follows.

1. Review and comprehend foetal hypoxia (asphyxia), in accordance with the clinical guidelines.
2. Review and evaluate various research studies that are aimed at improving the classification or diagnosis of foetal hypoxia.
3. Develop and evaluate various diagnostic procedure or classification methods using a machine learning classifiers trained with data records of cases with foetal hypoxia.
4. Examine various medical parameters using data science approaches to find relations between those parameters for the diagnosis of foetal hypoxia.

1.4 Research scope

This study focuses on providing a simple yet powerful method to enable the monitoring and follow-up of new-borns with foetal hypoxia as early as possible, even before real signs and symptoms appear, this can be achieved by monitoring the infant's clinical tests at delivery and finding a final classification of the cases through an ML-based diagnostic classifier. This study also focuses on defining the exact parameters of the new-born clinical tests that can provide a better way of following-up the foetal state and give a better decision whether the infant has a sign of respiratory acidosis or any other symptoms of metabolic acidosis.

The following is a list of the main types of foetal hypoxia that this study will investigate:

Respiratory acidosis: occurs when the arterial partial pressure of carbon dioxide (PCO2) is elevated above the normal range (>44 mm Hg) leading to a blood pH lower than 7.35 [15]. By definition, the diagnosis of respiratory acidosis requires measurement of PCO2 and pH.

Metabolic acidosis, defined as the accumulation of non-carbonic acid equivalents, arises from excessive production or inadequate excretion of hydrogen ions or from an increased loss of bicarbonate. In practice metabolic acidosis may result from birth asphyxia, cold stress, hypovolaemia, sepsis, congenital heart disease (particularly hypo plastic left heart syndrome, coarctation and interruption of the aortic arch) and renal disease. The majority of conditions in which foetal acidosis is present are associated with foetal hypoxemia and the accumulation of lactic acid in the foetal tissues and blood. When the cells are not receiving adequate oxygen, they revert to anaerobic metabolism as a compensatory mechanism, which produces acidic by-products, such as lactic acid; when too many acid by-products are in the blood, acidosis occurs. BDecf levels and the base excess can be used as the best threshold in detecting the metabolic acidosis. While respiratory acidosis means the acidosis is due to impaired gas exchange (elevated carbon dioxide), metabolic acidosis is acidosis caused by metabolic reasons, such as a low HCO_3^- or the occurrence of anaerobic metabolism. In simplest terms, HCO_3^- represents the metabolic component and PCO_2 represents the respiratory component of acid base status [20].

Mixed acidaemia: Mixed acidaemia is metabolic acidosis that develops when respiratory acidosis is prolonged. The baby is not getting rid of enough carbon dioxide, which causes acidosis (respiratory acidosis). Then, the prolonged oxygen deprivation causes anaerobic metabolism, which produces a metabolic acidosis. This is the most common pattern seen after prolonged end-stage bradycardia [20].

1.5 Structure of the thesis

The remainder of this thesis is organised as follows:

Chapter Two will discuss the main concept of foetal distress, including the definition, types and the main causes of foetal hypoxia. The next section will give a detailed explanation of the pathophysiological changes of the foetus during and after a hypoxic state, with the main adaptive mechanism of the foetus that helps to overcome the early stages of oxygen deprivation. All the diagnostic procedures and different methods of diagnosis are also included in this section, in addition to the main criteria of acute intrapartum hypoxia, which will be used as a parameter during the classification process. The parameters that have been used in this study as an evidence of foetal hypoxia will be according to the guidelines of both the Society of Obstetricians and Gynaecologists of Canada (SOGC) and The American College of Obstetricians and Gynaecologists (ACOG).

Chapter Three will discuss machine learning technology as an AI discipline, resulting in huge development of human knowledge. Different types of ML problems, as well as types of classifiers, will be provided with a brief description of each type. The importance of the machine learning concept and how it influences human knowledge in different aspects has also been included in this chapter. In addition to the main ML classifiers that will be used for data classification and detecting the hypoxic cases will be discussed at the end of Chapter Three.

Chapter Four will discuss the project's methodology, depending on related research and studies, which give us a better idea of how we can deal with the medical data and the classification process in general. The following section of this chapter will include the main steps of the primary data analysis. Understanding the data we use is considered an important step in analysis and will give us a better idea about how we can deal with it. Preparing the dataset by eliminating missing data, observing the classes' distribution and features selection will be included in this chapter as important steps for the classification process. The last section of Chapter Four will discuss the relationships between different parameters that will be used for the classification. Although we depended on the SOGC and ACOG guidelines in detecting hypoxic state, in this study we will examine different combinations of these parameters and observe any differences in the final decision of the classification.

Chapter Five will explain the classifiers implementation on the preprocessed data as well as the main evaluation methods that will be used for each classifier. This chapter will have an explanation of the classification as one of the ML problems, in addition to the general approach of three different experiments of the classification process using a different combination of parameters.

Chapter Six will include the simulation results of the three classification experiments using six different classifiers in each experiment, then will discuss and analyse these results to determine the performance of each classifier in detecting the hypoxic cases.

Chapter Seven will have the final conclusion of our study, providing the best classifier/s and will determine if different combination of parameters will be useful to detect the hypoxic cases. Finally, the main ideas of possible future work will be discussed in the last section of Chapter Seven.

CHAPTER TWO: FOETAL DISTRESS CLASSIFICATION

2.1 Introduction

This chapter will introduce foetal hypoxia as a serious condition that could be considered the third most common cause of new-born death. In addition to its serious effect on the mother, it is also responsible for about1 in 20 of pregnancy terminations by caesarean section to save the foetus before developing serious complication such as life-long disabilities, or even death. The main causes of each type of this disorder will be discussed and how it can be developed during the antenatal and postnatal period. The main diagnostic tests are discussed with the main clinical criteria and variable threshold values of foetal hypoxic cases according to different national committee guidelines such as The American College of Obstetricians and Gynaecologists criteria (ACOG) [16], the Society of Obstetricians and Gynaecologists of Canada (SOGC) [17] and the Royal College of Obstetricians and Gynaecologists (RCOG) [18].The main management lines that could prevent or minimise the risk of serious complication will be discussed.

2.2 Foetal Distress

Foetal distress is a condition of foetal oxygen deprivation and accumulation of carbon dioxide, leading to hypoxia and acidosis during the antepartum period or intrapartum period. Foetal distress and hypoxia (oxygen deprivation) is considered a serious condition and one of the main reasons for caesarean section in the obstetrics and gynaecology department. Foetal distress occurs in about 1 in 20 pregnancies. Many of the foetuses have experienced hypoxia during different stages of the pregnancy period. It is common among physicians to use the term foetal distress to explain foetal hypoxia, or birth asphyxia, however each term has a different meaning and explains a distinct process of foetal health.

According to the SOGC task force report on cerebral palsy and asphyxia, there are different definitions of each hypoxic state of the newly born infant. The following definitions explain the accurate meaning of each term:

Hypoxemia: decreased oxygen content in blood

Hypoxia: decreased oxygen content in tissues

Acidaemia: increased H+ content in blood

Acidosis: increased H+ content in tissues

Asphyxia: hypoxia with metabolic acidosis

2.3 Foetal Hypoxia Causes

Foetal hypoxia can be classified according to the stage of the intra-partum foetal life, which can be acute or chronic hypoxia. The former usually occurs during the labour process while the latter occurs during the first, second or third trimester of the pregnancy [2].

Each type has many factors contributing to a reduction in normal oxygen levels as shown in Table 2.1.

Table 2.1 Foetal types and the related causes

Acute Hypoxia	Chronic Hypoxia
Maternal causes: • Hypotension or hypovolemia* • Hemorrhage • Vaso-vagal attack • Epidural anaesthesia • Uterine contractions can also interrupt the uterine blood flow by increasing the pressure on the fetus and blood supply	**Maternal causes:** • Severe respiratory or cardiac disease (reduced oxygenation of maternal blood) • Connective tissue diseases; systemic lupus erythematosus, pre-eclampsia. (Reduced blood flow to the placenta) • Significant anaemia (iron deficiency anaemia, hemo-globinopathies). • Antiphosphoslipid syndrome
Placental causes: • Abruption can disrupt the utero-placental circulation by separating and so tearing the uterine spiral arteries from the placenta. • Hyper stimulation secondary to oxytocin, prostaglandins E2	**Placental causes:** • Inadequate trophoblast invasion of the myometrium in early pregnancy* • Utero-placental dysfunction • Abnormal placenta implantation (Placenta previa)
Foetal causes: • Umbilical cord compression • Oligohydramnios • Cord prolapse	**Foetal causes:** • Anaemia from rhesus disease • Parvovirus infection • α-thalassaemia • Foetal-maternal haemorrhage • Serious cardiac structural abnormalities.

*Any causes of haemorrhage or hypotension and hypovolaemia can reduce the maternal blood supply and so oxygen delivery to the uterus.

2.4 Foetal distress physiology

Foetal oxygen concentration (PO2) is lower than maternal (40mm Hg in umbilical vein compared to 95 mmHg in maternal artery). Oxygen saturation in the umbilical vein is almost the same as the maternal arterial blood and this could be explained by the fact that there is a higher haemoglobin concentration with higher affinity for oxygen in the foetal blood. This considers the first compensatory mechanism helping the foetus to release more oxygen to the tissues during the intra-partum foetal life.

Another compensatory mechanism for low oxygen concentration (PO2) provided by the maternal blood is redistribution of the blood to the most vital organs secondary to high cardiac output and increasing the extraction of the tissue oxygen [19] [25]. Figure 2.1 shows the patho-physiological process of foetal hypoxic disorder.

At the early stage of the foetal distress or during short-term episodes of acute hypoxia (lasting a few minutes) caused by uterine contractions and/or compressions of the umbilical cord [20], the foetus may tolerate the oxygen flow reduction with no signs of hypoxic complications. The foetus may be able to adapt the hypoxic effects by causing the blood flow to be redistributed and increasing the blood supply containing the nutrients and oxygen to the most vital organs such as the brain, myocardium, and upper body and reducing the perfusion of the kidneys, gastrointestinal tract, and lower extremities [21].

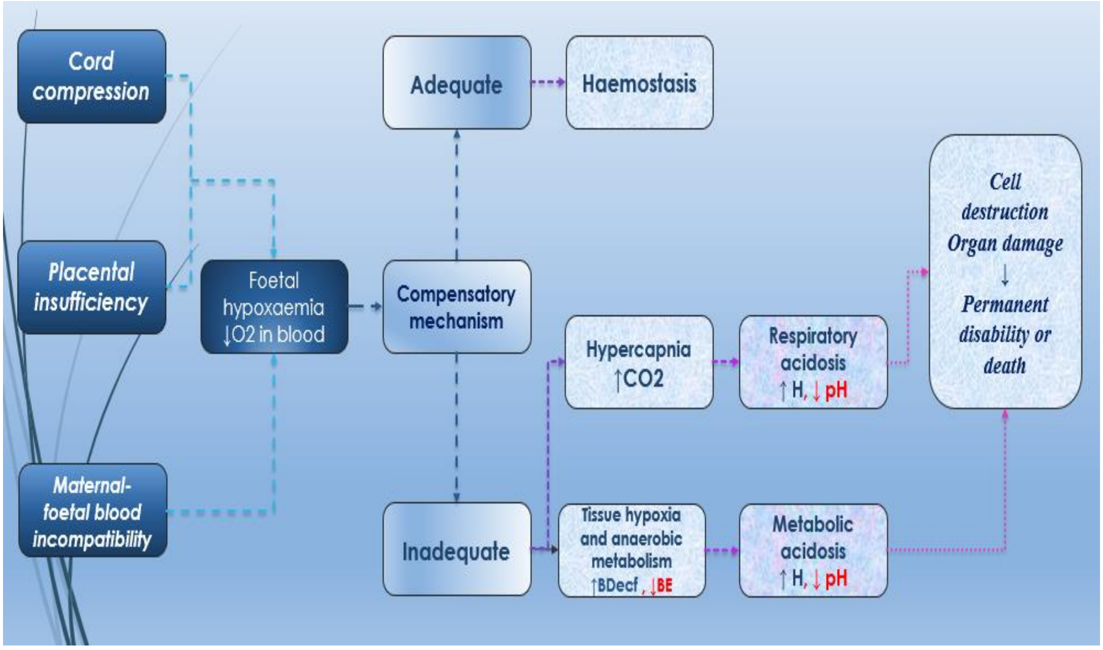

Figure 2-1 Foetal distress pathophysiology

In the case of chronic hypoxia, the long-time redistribution of oxygen has adverse consequences for the foetal heart. Further reduction of the foetal oxygenation will be overcome by the deterioration in cardiac output and the development of arterial hypertension. Barker et al. [22] indicated the “foetal origins of adult disease hypothesis”, “Surviving babies seem to be susceptible to the development of serious postnatal complication such as cardiovascular disease later in life”. Barker’s theory states “physiologic adaptation enables the foetus to survive a period of intrauterine deprivation resulting in permanent reprogramming of the development of key organs that may have pathological consequences in postnatal life”. Furthermore, reduction in oxygen perfusion will cause carbon dioxide accumulations in the foetal blood causing an increase in the partial pressure of carbon dioxide (PCO 2) and a concomitant decrease in pH producing respiratory acidosis which is similar to the adult respiratory acidosis [22].

Continuous hypoxia deprives the foetus of the required oxygen to perform the aerobic reactions, resulting in accumulation of organic acids with the accumulation of pyruvic and lactic acids as a result of anaerobic metabolism and subsequently developing metabolic acidosis [23].

Compensatory foetal response to prolonged or profound reductions of oxygen can be summarised as foetal tachycardia (FHR >160 bpm), decreased movement, tone and breathing due to the decrease in oxygen consumption and finally decreasing the blood flow to some organs such as Kidney, Lungs, Gut, Liver and the Peripheral tissue (increase the anaerobic metabolism) compared to the increased blood flow to the Brain, Heart and Adrenals [24].

Thus, any damage that has happened to the foetal brain is correlated to the damage of other organs. Severe respiratory or metabolic acidosis indicates damage to several organs such as Lungs, Heart and Kidneys. Accordingly, evidence of multi organ failure, as well as metabolic acidosis, is required to diagnose foetal birth asphyxia [25].

2.5 Evidence of Multi-organ Failure

The following is a list of evidence for Multi-organ failure [25]:

- Apgar score <3 at 5 minutes, means the neonatal vital signs have weakened with or without neonatal neurological squeals (hypotonia, seizure, irritability).
- Kidney-oliguria, anuria.
- Lung-respiratory distress syndrome (RDS), pulmonary hypertension.
- Heart-cardiomyopathy, patent dectus arteriosus (PDA).

- Liver–hypoglycaemia (low blood sugar) elevated liver enzymes.
- Biochemical profile (pH <7.0, Base deficit 8-16).

2.6 Foetal hypoxia Diagnosis:

Various methods of intra-partum foetal surveillance have been employed to detect the signs of hypoxia as early as possible to minimize the risk of life-long disability such as cerebral palsy and to reduce the mortality rate among the new-borns.

During hypoxia, numbers of cardio-respiratory system modifications we be made as hypoxia progresses in foetal life. These responses have been used as a compensatory mechanism to preserve the oxygenation of vital organs such as the brain and heart. As the foetal hypoxia progresses, Foetal heart pattern can be changed as a response to hypoxia. These pattern changes can be considered a useful way to detect the correlation between the foetal distress and the FHR changes [26]. Foetal heart rate monitoring, typically by use of the electronic assessment methods such as cardiotocography (CTG) and intermittent auscultation, are accepted methods of antenatal screening. However, the precise relationship between hypoxia and foetal heart rate changes is not well understood. This situation is likely to explain some of the inaccuracy inherent in diagnoses of asphyxia based on foetal heart rate observations [26].

Another way of foetal monitoring was summarised by Manning et al. [27]. Manning assessed the general foetal activity by evaluating five biophysical variables (foetal breathing movements, foetal movements, foetal tone, qualitative amniotic fluid volume, and the non-stress test) to form a biophysical profile of the foetus. Although Manning et al. had determined the foetal distress condition and perinatal mortality rate by studying each single variable, however the false negative rate was low and was similar between tests, while the false positive rate was high (>50%) and varied significantly between tests. On the other hand, studying all variables together shows significant changes in both the false negative and false positive rates as compared to any single test. According to Thacker et al.'s study [28], it has been shown that foetal condition is diagnosed more accurately using an ensample of measurements of all the biophysical variables profile instead of studying each variable

Many other methods have been used through the antenatal care and during the labour process such as the cordocentesis. Cordocentesis is a foetal blood sampling by ultrasound guided needle aspiration from the umbilical cord, usually used to detect hypoxia due to placental development abnormalities [29]. The main challenge to the physicians is still the detection of foetal hypoxia during labour/birth process because foetuses usually are subjected to maximum stress due to the increased duration and frequency of the uterine contraction [29].

This may cause many threats to the foetal health such as damage to the cells of the central nervous system that leads to life-long disability. Up to three quarters of infants with severe hypoxic-ischaemic encephalopathy (HIE) die of multiple organ failure or lung infections caused by irregular breathing. Those who survive are commonly left with gross symptoms such as mental retardation, epilepsy, and cerebral palsy [30]. Continuous labour monitoring is essential to observe the foetal well-being. There are many studies indicating that foetal heart activity is the prominent source of information about foetal health and especially for detection of foetal hypoxia [31].

2. 6.1 Electronic foetal monitoring

Intra-partum monitoring of foetuses during labour has been commonly performed by monitoring the foetal heart such as the intermittent auscultation, which is main method of foetal surveillance in labour [32].

Significant changes can be detected using IA, if the time period was immediately after the contraction. Repetitive decelerations of FHR detected by IA may give the first clue of the Asphyxia presence and may be further clarified by the EFM changes. It was a challenging attempt to use the correlation between the FHR pattern and the foetal health state as an outcome predictor. However, the ability of the EFM to predict the outcome was poor, particularly in low-risk pregnancies. An evaluation study comparing the EFM and IA role in foetal hypoxia detection showed that, although the operative delivery rate by caesarean section was significantly increased, however, the foetal outcome in preventing serious complication was not improved. Many of these electronic methods have been the subject of recent controversy, as existing studies have found no benefit from using such methods in reducing rates of cerebral palsy or peri-natal mortality. Therefore, physicians have found many challenges to identify the best way of detecting foetal hypoxia by analysing the foetal heart rate [3]. In 1960, cardiotocography (continuous electronic foetal heart rate monitoring) was introduced to replace all the other traditional methods in monitoring the foetal heart. Since then, this method has been routinely used by obstetricians to assess the foetal heart rate and uterine contraction (UC) [33].

Examining Cardiotocography trace patterns by healthcare professionals to interpret the intra-partum FHR patterns and provide suitable management decisions is a big challenge [34]. CTG records need high proficiency of the medical staff to identify the suspicious and pathological changes of the FHR that may correlate with the maternal uterine contraction. According to ACOG, SOGC, RCOG and FIGO (International Federation of Gynaecology and Obstetrics) guidelines, any deceleration in the FHR $\geq$ 15 bpm for $\geq$ 15 sec or isolated prolonged deceleration $\geq$ 3 min will indicate the foetal distress condition [35].

Their decision will identify the appropriate course of action (such as performing a Caesarean section). However, there is great variability among physicians in terms of how they perform this task. Furthermore, because significant hypoxia is usually rare, false alarms are common, leading physicians to disregard truly abnormal signals. Approximately 50 percent of birth-related brain injuries are deemed preventable, with incorrect CTG interpretation leading the list of causes [36, 37]. On the other hand, it is possible that some physicians miss some abnormalities of the FHR which are considered life-threatening changes that lead to life-long handicaps in the babies as well as the emotional distress of the parents and the financial compensation for the families who suffer from a medical catastrophe during labour.

Many techniques and methods have been used ante-natally to assess the foetal well-being and help understanding some of the life events before birth such as *'ultrasound imaging'* which, provide some ideas about the foetal size, skeleton abnormalities and vital organs development and Doppler test, which studies the foetal circulation and detects the vessels (arteries, veins) abnormalities before the birth [38]. Although, these methods can provide the physicians with a good assessment of the foetal wellbeing, unfortunately, none of the discussed methods can give an accurate diagnosis of the foetal hypoxia or asphyxia.

2.6.2 Foetal blood sampling

Further diagnostic tests are important in case of a typical (non-reassuring) foetal heart rate FHR signs. Foetal scalp sampling or umbilical cord sampling has been considered an essential test to confirm the foetal asphyxia diagnosis. In early 1980s, cordocentesis technique was used to identify the acid base state of the foetus and identify the early sign of respiratory acidosis (when the umbilical cord pH< 7.0) and metabolic acidosis (BDecf 8-16). Cordocentesis or Percutaneous Umbilical Cord Blood Sampling (PUBS) is a diagnostic test firstly performed by Daffos under an ultrasound guidance using a fixed needle guide attached to the base of the ultrasound transducer to study the biochemical profile of the umbilical blood sample [39]. According to Okamura et al. [40],Cordocentesis has considered “the precise evaluation of foetal condition to determine the timing of the delivery and to prevent the neurological sequelae caused by hypoxia” [40].

2.6.2.1 Foetal scalp blood sampling (FSS)

Foetal scalp blood sampling (FSS) should be performed when there is a typical (non-reassuring) FHR pattern detected by either IA or the EFM. Scalp blood sampling can be used in detecting fetal hypoxia by providing the pH level as an indicator. Table 2.2 presents the thresholds of the pH that relate to the clinical health state of the foetus and required management by the health staff.

Table 2.2 Thresholds for foetal health state

Threshold	Period
pH >7.2	FSS should be repeated if the FHR abnormality still exist
pH 7.21- 7.24	Repeat within 30 minutes or delivery should be considered
pH < 7.2	Enhanced delivery should be considered

Foetal blood sampling can be considering an ideal method in detecting foetal hypoxia. Many researchers depended on this technique in detecting and predicting foetal hypoxia depending on the pH level, such as Saling et al. [7] and Westgren et al. [10]. However, this technique has some limitations as it gives instant, not continuous results making repeated sampling mandatory to follow the foetal status. In addition, there are difficult technical restrictions such as the operator skills, risk of sample contamination with the amniotic fluid, risk of foetal infection that may be transported from an active maternal infection (HIV, Genital herpes). Other limitations are the large amount of the blood that is required for the analysis (at least 30-50 ml) and the requirement of the cervix dilation as it should be at least *2cm* dilated for the sampling to be done. Furthermore, FSS can give a false negative result in the case of metabolic acidosis, as it takes a long time for the H+ ions to cross the peripheral tissue in to the blood stream [10].

2.6.2.2 Umbilical cord blood sampling

According to the SOGC, it is important to do the cord blood gas sampling after all deliveries and it has been recommended to be done routinely for all neonates. Many international organizations have different recommendations about the use of umbilical artery or the vein blood sample. SOGC recommends measuring both umbilical artery and vein samples assessment. However, if only one is possible, it should be the arterial sample. On the other hand, ACOG recommends sampling the umbilical artery only for selective measurement of the cord gases (ACOG, 2006) [16]. While the Royal College of Obstetricians and Gynecologist recommends selective measurements of acid base status in the umbilical artery as a minimum (RCOG, 2001) [18].

Physicians usually use Umbilical cord blood sampling to detect five important thresholds to diagnose foetal asphyxia (pH, PCO2, PO2, BDecf and BE) [41]. In addition to the Apgar score, which is an expression of the infant's physiological condition at one point in time, which includes subjective components (colour, heart rate, reflex irritability, muscle tone and respiration) [42].

PCO2 level is usually increased in the case of respiratory acidosis, which can be developed rapidly following the first neonatal breaths as a normal response to the umbilical cord clamping after delivery. The cord clamping or in the case of cord compression will cause interruption of the blood flow from the foetus to the placenta, which will cause accumulation of CO2 in the foetal blood vessels causing a decrease in the pH level as a result of the reaction between the CO2 and water producing bicarbonate and hydrogen ions, which will accumulate in the blood vessels causing the decrease in the pH level.

On the other hand, PCO2 will not change in the case of metabolic acidosis as in this case the pathogenesis will be different from the respiratory acidosis. Metabolic acidosis takes longer to develop as a result of decreasing the PO2 level that causes the foetus to shift to anaerobic metabolism to maintain the energy balance. As a result of the anaerobic metabolism, lactic acid will accumulate in the tissue and dissociate into lactate and hydrogen ions causing a decrease in the pH level when some of the hydrogen ions move from the tissue to the blood vessels [24].

Using pH in detecting the foetal hypoxia is considered an important indicator to quantify perinatal asphyxia [43]. However, pH alone will not differentiate between the respiratory and metabolic acidosis as it increases in both conditions as shown in Figure 2.1. The last threshold of the umbilical cord blood sampling is the base deficit/base excess which can provide a better differentiation between the respiratory and metabolic acidosis as they are both normal in the respiratory acidosis while in case of metabolic acidosis the base deficit will increase and the base excess will decrease [44]. In the case of a normally delivered infant with no sign of hypoxia or any other health state that could be related to oxygen level deprivation, the umbilical cord blood sampling should have the following thresholds:

Normal values in an umbilical arterial sample in a term newborn [44]:

- pH: 7.18 – 7.38
- PCO2: 32 – 66 (mmHg)
- HCO3: 17 – 27 (mmol/L)
- PO2: 6 – 31 (mmHg)
- Base excess: -8 – 0 (mmol/L)
- Base deficit: 0 – 12

In this case the infant will not need any further intervention by the medical staff.

However any changing of these levels could be an early sign of developing foetal hypoxia disorder. Although the hypoxic state can be varied in its severity, we will depend on two of the main guidelines that are followed by many physicians; the first guideline for acute intra-partum hypoxia is from the society of Obstetricians and Gynaecologists of Canada

Criteria of acute intra-partum hypoxia according to the Society of Obstetricians and Gynaecologists of Canada (SOGC):

- Apgar score 0 – 3 for >15 minutes
- Neonatal neurological signs (hypotonia, coma, seizure)
- Evidence of multi-organ failure
- Umbilical cord pH < 7.0
- Umbilical cord arterial base deficit < 16 mmol/L

It cannot conclude the existence of the hypoxic acidaemia without any of these criteria. All these conditions should be present for definite diagnosis of foetal acidaemia.

The American College of Obstetricians and Gynaecologists (ACOG) have another guideline to classify the health state of the foetus according to the stage of the foetal life. The ante-partum birth asphyxia could happen when the foetus is still inside the womb, while the intra-partum birth asphyxia could happen during or after delivery. Therefore, there are two different sets of criteria to diagnose the foetal hypoxia.

A. The American College of Obstetricians and Gynaecologists criteria of intra-partum birth asphyxia:

- Sudden or sustained bradycardia or the absence of FHR variability in the presence of persistent, late or variable decelerations.
- Apgar score 0-3 beyond 5 minutes
- Onset of multi organ failure within 72 hours of birth.
- Early imaging study showing evidence of acute non focal abnormality

B. Ante-partum criteria of Asphyxia:

- Evidence of metabolic acidosis (Umbilical cord pH< 7.0, base deficit 8-12)
- Early onset of sever or moderate neonatal encephalopathy in infants born at 34 or more weeks of gestation
- Cerebral palsy of the spastic quadriplegic or dyskinetic type.
- Exclusion of other identifiable etiologies such as trauma, coagulation disorders, infectious conditions, or genetic disorders.

2.7 Foetal non- reassuring status management

The initial treatment used for non-reassuring foetal status is the intrauterine resuscitation. This will prevent any unnecessary intervention.

The main intrauterine resuscitation techniques that can be done by the physicians during the intra-partum stage are:

- Maintain continuous oxygen supply to the foetus is the first priority during foetal distress management. Physicians usually start by changing the mother's position to decrease the unnecessary pressure on the uterus or the placenta.

- Ensuring the mother is well hydrated and has adequate oxygen.
- Amnio-infusion: refers to the instillation of fluid into the amniotic cavity. This procedure is typically performed during labour through an intrauterine pressure catheter introduced trans-cervically after rupture of the foetal membranes. Alternatively, fluid can be infused through a needle trans-abdominally (the reverse process of amniocentesis). The rationale for amnio-infusion is that augmenting amniotic fluid volume may decrease or eliminate problems associated with a severe reduction or absence of amniotic fluid [45].
- Tocolysis (using some medication or techniques to suppress the uterine contraction temporarily and delay the preterm delivery).

Nonetheless, there are some conditions in which emergency cesarean section is mandatory. However, due to the over-diagnosis of foetal distress and potential misinterpretation of the foetal heart rate, it is recommended to confirm a potential foetal distress diagnosis with a foetal blood acid base study. Overall, this condition points to the importance of prenatal care and proper monitoring of the mother and foetus throughout pregnancy.

2.8 Summary of the chapter

In this chapter, foetal asphyxia disorders are discussed. Main types of foetal hypoxia classified according to the main reason of oxygen level deprivation have been shown. Foetal hypoxia general causes can be summarized according to the time of hypoxic state development during foetal life into acute and chronic hypoxia. Discussion on the foetal hypoxia path physiological process to understand the main relationships that may affect the diagnostic variable combination was provided. Other information has been included in this chapter on the diagnostic procedures of the foetal hypoxia including, the electronic foetal monitoring and foetal blood sampling. Each of them is presented with the main clinical features and diagnostic criteria based on latest clinical guidelines and references. In general, this deep investigation of the different types of foetal hypoxia and the variable ways of diagnosis showed that foetal blood sampling during the ante partum stage of the foetal life as a diagnostic tool, can be depended upon by the physicians to start the management process according to the ACOG guideline. There is no big deference between the American College (ACOG) and the SOGC in identifying the hypoxic foetus criteria apart from the base deficit threshold and the timing of the asphyxia. The next chapter will discuss machine learning classifiers as well as the types that could help in analysing foetal data for classification purposes.

CHAPTER THREE: MACHINE LEARNING TYPES AND APPLICATION

3.1 introduction

In this chapter, Machine learning classifiers will be identified as an artificial intelligence discipline. ML classifiers allow the computers to imitate the way by which humans can deal with large data automatically via various methods including analysis, self-training, observation and experience. Different classifiers that show advantage in data analysis especially medical data and various applications of real world problems, such as the decision tree classifiers, neural network and the support vector classifiers will be discussed.

3.2 Machine Learning Development

Machine learning is an artificial intelligence discipline geared toward the technological development of human knowledge. It is considered the fundamental sub-area of artificial intelligence and one of the fastest growing fields in computer science [46]. As digital technology increasingly infiltrates our daily life, more data is continuously getting bigger, either generated or collected. Machine learning technology plays a significant role in data analysis in variant fields such as astronomy and biology. The collection of datasets is getting larger every year and also it is represented in different ways, not just numbers or character strings anymore but images, video, audio, documents web pages, graphs and more.

Machine learning has the ability to improve its own performance through the use of software and algorithms that use artificial intelligence techniques to imitate the ways humans seem to learn, such as repetition and experience through continuous exposure to new scenarios, testing and adaptation. Machine learning classifiers are allows computers to perform new learning situations automatically without human intervention or assistance, which can be done via analysis, self-training, observation and experience.

Jaime et al. [47] showed that the computer systems do have the ability to perform any task through examples or previously solved tasks; in addition computers cannot improve the task performance on the basis of past mistakes or obtain new abilities by observing and imitating experts.

Machine learning research has been extended to identify the possibility of instructing the computer system in a new way to match the human abilities. Alpaydin et al. [48] stated that

"*With advances in computer technology, we currently have the ability to store and process a large amount of data, as well as access it from physically distant locations over a computer network. Most data acquisition devices are digital now and record reliable data*".

Programming the computer systems through machine learning methods can optimise the standard performance by using example data or past experience. Machine learning usually uses statistical theories in building a mathematical classifier, and then refines up this classifier to some parameters using the training data or past experience. The developed classifier may have predictive characteristics to make future predictions, or descriptive features to gain knowledge from data, or both.

The learning process can be achieved by training the classifier through an efficient algorithm to solve the optimised problem. The representation of the learned classifier and algorithmic solution for interpretation needs to be efficient as well.

The major advantage of machine learning is the fact that it can extract patterns from massive amounts of data which humans cannot do because humans cannot retain everything in memory and they cannot perform obvious/redundant computations for hours and days to come up with interesting patterns. Machine learning has found major applications in finance, healthcare, entertainment, robotics, and many more. Once a Machine Learning classifier with good generalization capabilities is learned, it can handle previously unseen scenarios and take decisions accordingly.

Some tasks cannot be defined well except by example; that is it is possible to specify input and output pairs but not a concise relationship between inputs and desired outputs. Machines are able to adjust their internal structure to produce correct outputs for a large number of sample inputs.

Machine learning methods can often be used to extract any relationships or correlations that are hidden among large piles of data [50]. Machines can adapt to the changing environment over time, which would reduce the need for constant redesign.

3.4 Machine Learning: Algorithms Types

Machine learning algorithms are organized in specific assembling, based on the desired outcome of the algorithm. Common algorithm types include, Supervised learning, unsupervised learning, Semi-supervised learning, Reinforcement learning [52]. We will use the supervised learning as the main concept for data classification in this research.

3.4.1 Supervised learning

It is considered the most important methodology in machine learning, which means the ability of systems to infer a function of variant tasks depending on supervised training data. The training data usually consist of a set of training samples. Each instance has a pair consisting of input data (typically a vector) and a desired result as an output value (also called the *supervisory signal*) [53]. Supervised learning involves learning a mapping between a set of *input* variables X and an *output* variable Y and applying this mapping to predict the outputs for unseen data. An optimal consequence will allow for the algorithm to correctly determine the class labels for unseen instances. The main characteristic of the supervised classifiers is the generalization ability, which means having the ability to produce reasonable outputs for inputs not encountered during the training [52].

The standard formulation of the supervised learning task is the classification, when the learner needs to learn (approximate the behavior of) a function that maps a vector into one of several classes by observing several input-output examples of the function [52]. In the building phase, the training procedure continues until the algorithm is able to achieve the best accuracy on the given data.

3.4.2 Unsupervised learning

Unsupervised learning is also one type of machine learning model that can be applied to drive implication from training datasets involving input data without output (labelled responses). In unsupervised learning, the system will learn by a set of particular input patterns that reflect the statistical structure without specific target output or environmental evaluation for each input [54]. Unsupervised learning has more similarity to the human brain compared to supervised learning. For instance there are about 10 million photoreceptors in the human eye with constantly changing activities in visualization of different objects, providing all the information that indicates what the objects are. The structural and physiological properties of neocortex synapses are known to be influenced by the sensory neurons' pattern of activity. However, none of the information is available, making the unsupervised learning more essential in computational models for synaptic variation. Although the unsupervised learning shows no benefit in data analysis for prediction of a response as the supervised learning does, it shows a precise importance in identifying unknown subgroups among the variables.

3.4.3 Semi-supervised learning

Semi-supervised learning is a learning paradigm concerned with the study of how computers and natural systems such as humans learn in the presence of both labelled and unlabelled data. Availability of large amounts of unlabelled data compared to the small amounts of labelled data, in addition to the expensive labelling or data annotating has motivated machine learners to study new methods that can use information of the input distribution.

Semi-supervised learning (SSL) can be considered a midway between supervised and unsupervised learning. The standard semi-supervised learning has data sets divided into two parts, one points with labels while for the other points, the labels are not known [54].

Semi-Supervised Learning in Practice

Semi-supervised learning is useful whenever there are far more unlabelled data than labelled. This type of algorithm is a significant tool in machine learning due to the ability to use unlabelled data to enhance supervised learning tasks and can yield considerable improvement in accuracy especially when the data is expensive and scarce [55]. This is the case in many application areas of machine learning, for example: In speech recognition, it costs almost nothing to record huge amounts of speech, but labelling it requires some human to listen to it and type a transcript. Billions of Webpages are directly available for automated processing, but to classify them reliably, humans have to read them.

Another example is the study of the protein sequences. Protein sequences are nowadays acquired at industrial speed (by genome sequencing, computational gene finding, and automatic translation), to resolve a three-dimensional (3D) structure or to determine the functions of a single protein may require years of scientific work.

3.4.4 Reinforcement learning

In the machine learning domain, Reinforcement learning (RL) is usually identified as techniques whereby an algorithm learns from the regular consequences of its actions instead of being explicitly taught based on previous experiences (exploitation). The machine is exposed to an environment where it trains itself continually using trial and error. This machine learns from past experience and tries to capture the best possible knowledge to make accurate business decisions [56].

The reinforcement machine learning instructs the algorithm about the policy of how to act, given an observation of the world. Every action has some impact in the environment, and the environment provides feedback that guides the learning algorithm [57].This kind of machine learning would not emphasise which action should be taken under consideration, but instead would have to discover which action produced such an excellent reward. RL is usually utilised in different applications to solve a number of complex tasks. For instance, RL has performed in medical diagnosis, speech recognition, bioinformatics, computational vision, spell recognition, and robots Locomotion [52].

3.5 Machine Learning Classifiers

Machine learning algorithms are used to optimize the performance criterion using example data or past experience [58]. .

Machine learning uses the theory of statistics in building mathematical classifiers, because the core task is to make inference from sample data. In training, an efficient classifier is needed to solve the optimization problem, as well as to store and process the massive amount of data generated. Once the classifier is learned, its representation and algorithmic solution for inference needs to be efficient. For the classification and regression problem, there are various choices of Machine Learning classifiers each of them can be viewed as a black box that solves the same problem [59].

In this study, we chose different types of classifiers, based on their ability to classify a real world medical data. For instance, Jezewski et al. 2010 [80] and Spilka et al.2012 [84] used the NNET and SVM to classify their data and the results showed an acceptable accuracy outcomes. In addition to use the cross-validation and test harness to determine which classifier performs best on test data [125]. Studying our data characteristics (real world data, simple row data containing numerical variables and categorical classes) encouraged us to choose the decision tree based classifiers as well. Its methodology considered very easy to understand, as it doesn’t require analytical background to interpret the data. It has been considered one of the quickest ways to classify variables and find the relationship between them to predict the target variable, and it can control both numerical and categorical variables.

3.5.1 Artificial neural networks

ANNs are considered as devices that process information in a similar manner to the real neuron of the human nervous system. Their importance emerges with their ability to solve many complex problems that cannot be solved by traditional computing methods. All neural network classifiers consist of a number of neurons, connections and transfer functions. There are many applications that use the neural network such as prediction, classification, data association, clustering, and optimization and pattern completion. Many application areas that used the ANNs are finance, marketing, manufacturing, operations, information systems, etc.

The main idea of ANNs depends on the basic work of the human brain's neuron which, unlike the other cells of the body, has the ability to adapt the strength of connections with other neurons, of which they can connect with up to the 200,000,in addition to the natural evaluation and learning ability which gives the human mind the great power to solve very complex problems [60].

Artificial Neural Network Concepts

Artificial Neural networks (ANN) are considered an inspirational form of the biologically inspired algorithm, based on the assemblage of connected neurons observed in the biological brain [63]. ANNs are considered as inspired software programs that were developed to simulate the way that the human brain processes by modelling the actual systems depending on the information provided to it. This kind of machine learning involves hundreds of single unit artificial neurons that are capable of representing complex input and output connections. The main motivation behind developing an artificial neural network is the capacity to perform intelligent tasks, which can perform in a similar way to the human brain. Furthermore, the main idea of how the ANN works comes from the network connections of the many neurons that have ability to represent non-linear and linear relationships.

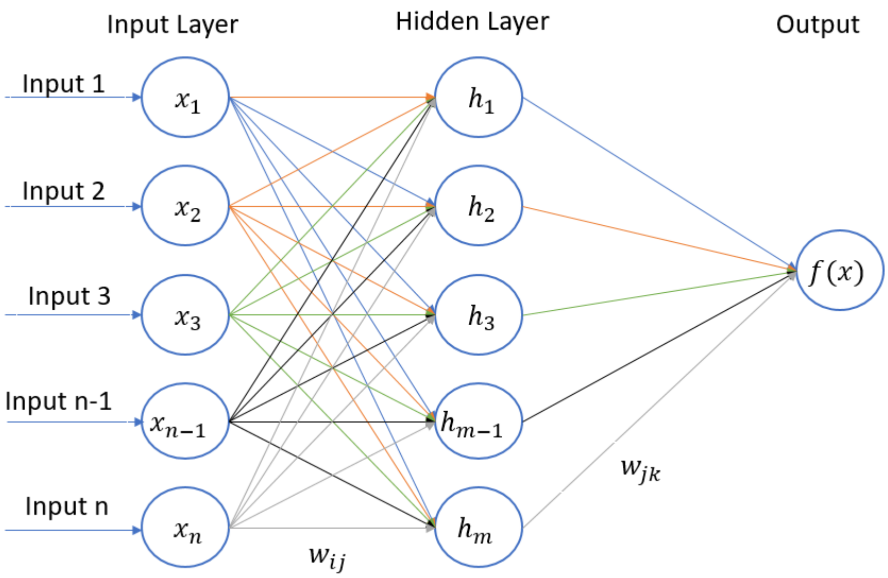

Figure 3.1 Typical ANN Classifier

Computational modelling of the artificial neural networks needs a number of connected neurons to form a network. In this context, generating an output, neurons are organised in layers with processing units that take one or more inputs. In this case, at each neuron, all inputs have to be connected with a weight that modifies the strength of each input. As a result, neurons will simply collect all the inputs together to calculate an output as demonstrated in Figure 3.1. The weights in each ANN are trained using different types of learning algorithm, for example supervised and unsupervised learning. Therefore, this can be achieved through a procedure called a training algorithm. The training set is utilised during learning to stimulate the learning algorithm, such that the desired outputs are produced, given the input values. In order that the network promotes the most important features within the training process, learning algorithms are utilised to update the weights of the ANN using mathematical equations.

The main backbone of using weights is to check if the output is too high, then the weights should be lowered by a certain amount to fit to the output for the entire input instance. On other hand, if the predicated output is too low, then the weights need to be increased by the set amount. The hidden layer learns to provide a representation for the inputs. In each ANN, one or more hidden layers can be applied.

MLPs

The perceptron is a neuron with a transfer function and a weight adaptive mechanism (learning) by comparing the actual and the desired output responses for any input or stimuli. Therefore its considered as the simplest form of neural network [66].

Multilayer Perceptron (MLP) is considered one type of feedforward network that is designed by an assembly of summing parts that are associated to the weights. Its concept depends on the comparison between the actual and the desired output responses for any input or stimulus. MLP has been successfully tested in many applications; among those are signal processing [67] and function approximation [68]

MLP consists of a number of neurons grouped in layers; it has three types of layers, which are:

Input layer, which sends the input signal (data) to the other layers in a forward direction.

Hidden layer/s: which transmit the signal (data) from the input nodes to the output nodes and enables the network to learn complex tasks and solve different problems.

The hidden nodes can perform two tasks: first, sum the weighted inputs to the hidden neuron, and second, pass the products of the sum through the non-linear activation function.

Output layer, this layer provides the product as the actual response of the network

The MLP network operation can be divided in to two phases:

Training phase: using a training algorithm, the MLP can be trained for its specific purpose.

Retrieval phase: using the previously trained MLP network to generate the output

The output of the MLP network is calculated as follows:

$$Y_K = \sigma(\textstyle\sum_{j=1}^{J} W_{kj}\ \sigma\left(\sum_{i=1}^{N} w_{ji} X_i + W_{jo}\right) + W_{ko} \qquad (2)$$

Where σ is a sigmoid transfer function, X_i represents the input value, w_{ji} is the weights from the input layer to the hidden layer, W_{kj} is the weights from the hidden layer to the output layer, W_{jo} is the bias for the hidden node and Y denotes the network output.

MLP has been successfully applied in different applications such as signal processing [67], financial time series prediction [69] and functions approximation [68].

The back-propagation learning algorithm

Backpropagation (BP) is one of the ‘most popular algorithms for ANNs’: it has been estimated by Paul Werbos, who first worked on the algorithm in the 1970’s [70], that between 40% and 90% of the real world ANN applications use the BP algorithm. Werbos applied the algorithm in political forecasting [71]. David Rumelhart, Geoffery Hinton and others applied the BP algorithm in the 1980’s to problems related to supervised learning, particularly pattern recognition [72].

Back propagation learning has this name after using the error back propagation algorithm to solve complex problems using a supervised training, when multilayer networks learn the mapping from a given data set [73].

Learning iteration of a neural network can be summarized into two phases [74]:

Feedforward pass: in this phase, the input vector is applied to the processing units of the network following the direction from input layer to the output layer. Thus, the actual output is produced. Consequently, all the weights of the network are fixed through this phase

Backward propagation: the synaptic weights of the network in this phase are adjusted to reduce the error between the desired output and the actual output. The actual output is subtracted from the target output to produce the error signal; this error is propagated backward from the output layer to the input layer through the network so that this algorithm is called “error propagation”

3.5.2 Support vector machines (SVM)

Support vector machines (support vector networks) are considered as one of the supervised machine learning techniques with associated algorithms that are able to analyse data sets used for regression and classification analysis. (SVM) theory as stated by Gonzalez et al. [86] was originally developed on the basis of a separable binary classification problem, (SVM) performs classification by constructing an N dimensional hyper plane that optimally separates the data into two categories. SVM classifier is closely related to neural networks (classical multilayer perceptron). In fact, an SVM classifier using a sigmoid kernel function is equivalent to a two layers perceptron neural network .The overall idea of SVM was established into the most commonly used form (soft margin classification) by Cortes and Vapnik [87], the method itself invented from the ideas inherent in statistical learning theory [88]. A classification task by SVM usually involves separating data into training and testing sets. Each instance in the training set contains one “target value” (i.e. the class labels) and several “attributes” (i.e. the features or observed variables). The goal of SVM is to construct a classifier (based on the training data), which predicts the target values of the test data depending on the test data attributes only.

The goal of SVM is to choose a Maximum-margin hyper-plane that can minimize the structural risk of misclassification during training, otherwise known as maximum margin classification, in contrast to empirical risk minimization, which is typical in classifiers such as ANNs.

SVM classifies both linear and nonlinear data. LSVM classifier separates between two classes in the training data by the hyper-plane, which represents linear decision surfaces that split the space in to two parts. While the Nonlinear-SVM classifier uses the kernel functions to transform the original training data into a high-dimensional feature space in which the input data become more separable compared to the original input space. Maximum-margin hyper-planes are then created.

The MLP classifier uses a back propagation algorithm to adjust the weights and determine the set of weights and bias values with the goal of minimising the error rate. In contrast, the SVM classifier in this study uses a Gaussian radial basis kernel function (RBF) to map the data into high dimensional space, where it is easier to create linear decision boundary in the features space. The decision boundary, also called hyper-plane, should maximise the margin between the foetal state classes for an optimal diagnosis.

Although SVM and MLP classifiers are behave differently, they are able to handle complex nonlinear relationships between the hypoxia features and the outcome diagnosis when they exist. The hidden nodes within MLP classifier allow the network to model complex nonlinear relationships, while different kernel functions, e.g. polynomial function, can be adopted by SVM classifier to turn a linear classifier into a nonlinear classifier.

The simple task of binary classification in Support Vector Machines (SVMs), successfully applied to many real-world problems involve prediction over two classes, such as face recognition [89] bioinformatics ,hand-written character recognition [90] and detecting spam

The linear SVM classifier can be used for datasets with a nonlinear decision boundary via the kernel trick. Given the extra flexibility that a linear decision surface in feature space actually corresponds to a nonlinear decision surface in the input space. The kernel trick allows us to carry out all operations via the kernel function computed in input space, rather than having to map the points into feature space as shown in Figure 3.2

$$K(x_i, x_j) = \varphi(x_i)^T . \varphi(x_j) \qquad (28)$$

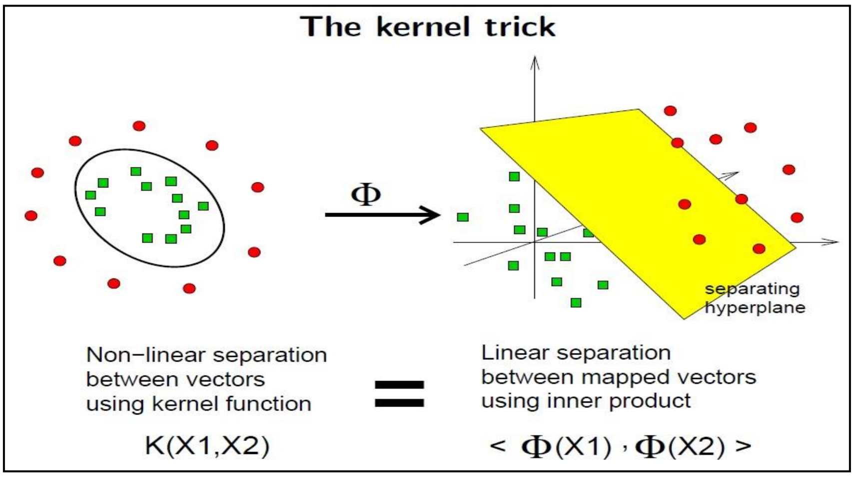

Figure 3.2 kernel trick demonstration

SVM has become a vital classification technique within the flood forecasting in a broad range of research and applications [93]. The standard mechanism of SVM classifier is the utilisation of a hyper-plane, which performances a discriminative boundary of data points in association with classes, where such a hyper-plane is established by maximum margin optimisation throughout the training procedure.

The addition of a kernel to the SVM permits such a separation to be improved within a feature space of higher dimension than that of the original problem representation, thereby allowing non-linear boundaries to be addressed.

In order to use a suitable kernel trick, the classifier can be applied and learnt without explicitly computing $\emptyset(x)$. This technique attempts to apply linear classifiers work into a nonlinear setting.

There are four basic kernels; linear, polynomial, radial basis function (RBF) and sigmoid. Radial basis function (RBF) kernels were used in our research. Equations below shows the RBF kernels function.

Radial basis function (RBF): $K\,(x, y) = \exp(-\frac{\|x-y\|^2}{2\sigma^2})$ (44)

$\|x - x⍰\|^2$ Recognized as the squared Euclidean distance between the two feature vectors.

As mentioned previously, the separation hyper-plane is carried by solving an optimization problem that selects the support vector points on the wrong side of the resulting hyper-plane. The penalty parameter C is the critical tuning parameter for construction of a good classifier that will generalize well. In addition, a number of different kernels are available to better account for transforming complex data spaces into a form that can be more easily separated. The selected kernel and associated parameters have a significant effect on how well the resulting classifier properly classifies the data.

There are two particularly attractive properties of SVMs:

1. A decision boundary, called a maximum margin separator is built so that the distance between points of different instances on either side of it is as large as possible. This helps in generalization.
2. The decision boundary is a linear separating hyper-plane, but SVMs embed the data into a higher dimensional space, with a kernel trick. This allows data that is not separable in the original space, to be more easily separated in a higher-dimensional space.

3.5.3 k-nearest neighbours algorithm (KNN)

K-Nearest neighbour algorithm (KNN) is considered one of the supervised learning classifier that have been used in the field of statistical pattern recognition, data mining, and many others. It's dependent for classifying features on closest training samples in the attributes space as shown in Figure 3.3. The main steps of the KNN algorithm workflow implementation are; Getting Data, Splitting the data into Train & Test Data, Euclidean Distance Calculation, KNN prediction function and Accuracy calculation

To demonstrate a KNN analysis, there is the procedure of classifying a new value (test value) among known samples. Figure3.3, shows the instances with the green signs (class 1) and blue signs (class2) and the target point with a black circle. The aim is to classify the output of the target point depending on a nominated number of its nearest neighbors. More accurately, we need to check the class that the black circle is belonging to, if its class 1 or class 2.

KNN is a classifier that is easy to understand but works exceptionally well during training and testing the algorithm [94]. This classifier is commonly applied for regressing and classification, which is used in pattern recognition and statistical estimation as a non-parametric technique. The main purpose of using kNN classifier is to use a datasetwhere the data points are divided into a number of separate target classess to predict or classify new instances. This type of algorithm is classifying the data sets by taking the majority vote of its neighbours, with the object being assigned to the class most common amongst its K nearest neighbours measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbour [95].

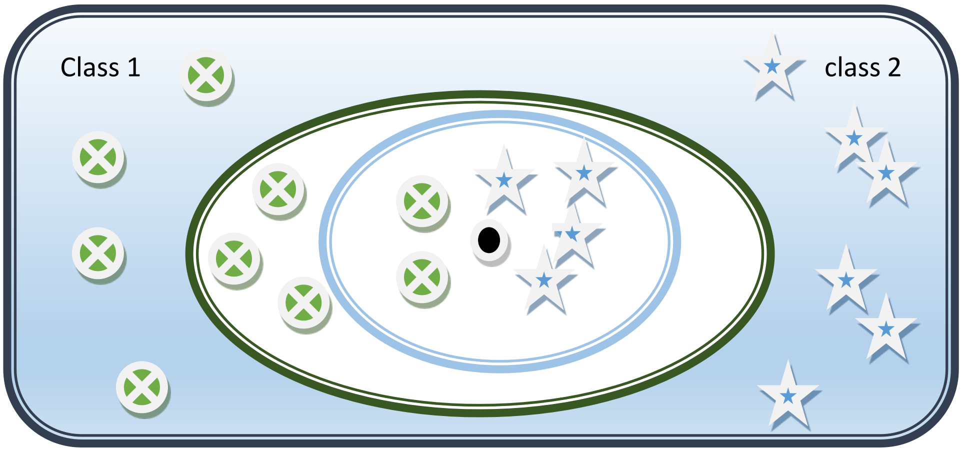

Figure 3.3 K-nearest neighbours algorithm (k-NN) example

The test instance in Figure 3.3 is the black circle, which should be classified to the green circles or to the blue stars. If K = 6 (blue line circle), it is assigned to the blue classes due to the fact that there are 4 blue stars and only 2 green circle inside the red line circle. If K= 9 (green line circle) it is assigned to the green class, which are the green circle (5 green circle and 4 blue stars inside the outer circle. [96].

KNN selects from a set of objects as the majority in the feature space for which the correct classification is obvious. The KNN works as follows. Firstly, determine the parameter K, for instance, the total number of nearest neighbours previously. Then, Distance need to be measured between the target-feature and all the training samples of the datasets is considered utilizing any distance measure classifier. In order to find the measurement of distances for the training sets, the nearest neighbour method that depends on the K-th minimum distance is confirmed. As mentioned earlier, K-NN belongs to supervised learning algorithms, training all datasets which fall under K to sort the feature. Then consequently, the prediction feature is measured throughout the majority of nearest neighbours.

In order to use the classification method, K is an unlabelled vector (a target or test point), and a user-defined constant is classified through assigning target label with the majority vote among the K training instances nearest to that test point. The outcomes of this classifier can be calculated as the class with the highest frequency when use for classification from the K-most similar samples. In addition, each sample in essence votes for their class and the class with the most votes is taken as the prediction [97]. There are two types of metric commonly used in the KNN, the Euclidean and the Manhattan distances. Euclidean distance refers to the distance between two points. These points can be in different dimensional space and are represented by different forms of coordinates. In one-dimensional space, the points are just on a straight number line. In two-dimensional space, the coordinates are given as points on the x- and y-axes, and in three-dimensional space, x-, y- and z-axes are used. Finding the Euclidean distance between points depends on the particular dimensional space in which they are found The purpose of using these type of metrics is to improve the accuracy of K-NN using specialized classifiers, for instance, neighbourhood components analysis or large Margin Nearest Neighbour [98].

Euclidean distance can be calculated by the equation 46 as shown below:

$$\mathbf{d}(\mathbf{x},\mathbf{y}) = \| \mathbf{x}-\mathbf{y} \| = \sqrt{(\mathbf{x}-\mathbf{y})^{\mathbf{T}}(\mathbf{x}-\mathbf{y})} = \sqrt{\sum_{\mathbf{j}}(\mathbf{x_{i-}}\ \mathbf{y_{j)}}^{\mathbf{2}}} \tag{46}$$

When d refer to Euclidean distance, $\| x \|$ and $\| y \|$being the norm of x, y respectively. $x_j and y_j$ is represented the element of x and y.

While Manhattan distance is the distance between two points measured along axes at right angles. Manhattan distance between two vectors (or points) a and b is defined as $\sum_i |a_i - b_i|$over the dimensions of the vectors.

Both Euclidean distance and Manhattan distance belong to of Minkowski's distances family. The *p*-distance between 2 vectors is clear as shown below:

$$\mathbf{X} = \left[\mathbf{x_1} \dots, \mathbf{x_2}, \dots, \mathbf{x_j}\right]^{\mathbf{T}} \mathbf{and} \mathbf{Y} = [\mathbf{y_1} \dots, \mathbf{y}, \dots, \mathbf{y}]^{\mathbf{T}}$$

$$\boldsymbol{d_p}(\boldsymbol{x}, \boldsymbol{y}) = \| \boldsymbol{x} - \boldsymbol{y} \|_{\boldsymbol{p}} = \left[\sum_{j}^{J} (\boldsymbol{x_{i-}} \boldsymbol{y_j})^{\boldsymbol{p}}\right]^{\frac{1}{p}} \tag{47}$$

KNN does not require using the training sets to apply any generalization. Lack of generalization leads this technique to keep all the training datasets. This means in other words, there is no explicit training set needed, as the training set is fast. To be more precise, the vast majority of the training dataset is required during the testing sets. In order to make contrast with other methods, for instance SVM can remove all non-support vectors without any problem.

On the other hand, KNN is considered as a lazy algorithm, which creates a decision depending on the entire training dataset. Finally, KNN performs poorly in classification due to the fact that all the parameters do not contribute equally by using the Euclidean distance method. The main disadvantage of KNN is the complexity in searching the nearest neighbours for each sample.

3.5.4 Decision Trees based classifiers

Decision tree classifiers are considered as an automatic computing procedure that is based on logical or binary operations .It can learn a task from a series of examples, applying a sequence of logical steps to classify the instances by sorting them based on their feature values. The node of the tree is called a feature of an instance that has to be classified while each branch represents a value that the node assumes [99]. The classification of the instances starts from the root node and sorted based on their feature values. The branches then lead either to other features or end in leaf nodes, which are classes. The root node divides the training data into possible branches [100].

The main advantage of using decision trees is their comprehensibility. The classification of an instance to a particular class is easily understandable and decision tree works better for discrete/categorical features.

The decision tree method is a powerful statistical tool for classification, prediction, interpretation, and data manipulation that has several potential applications in medical research [101].

Decision trees are useful performance assessment tools for exploring the performance of different classifiers .They discover rules and relationships between variables using a greedy approach hence an attribute chosen at the first step can't be used anymore which can give better classification if used in later steps [100].

Types of Decision Trees:

According to the target variable we can identify that there are two types of Decision Trees:

1. Categorical Variable Decision Tree: it is a type of decision tree where the target variable is based on "YES" or "No" answer.
2. Continuous Variable Decision Tree: it is a type of decision tree that has a continuous target variable.

Advantages of the Decision Tree: [3]

- Easy to understand: this methodology is very easy to understand and it doesn’t require analytical background to interpret the data. It has a graphical representation that makes it easier to be read by the users.
- Useful in data investigation: this method is one of the quickest ways to classify variables and the relationship between them and predict target variable.
- Less data eliminating required: outliers and the absence of certain values do not affect this classifier.
- Data type is not limited: this classifier can control both numerical and categorical variables.
- Non-Parametric Method: this classifier does not require modeller to make assumptions about the distribution population

Disadvantage of the Decision Tree

- The construction of a decision tree may be affected by the irrelevant attributes – E.g. ID numbers
- Small variations in the data can generate various trees
- A sub-tree can be replicated several times
- Error-prone with too many classes
- Not good for predicting the value of a continuous class attribute
- The main disadvantage is the over-fitting and under-fitting, mainly when using a small data set, which can limit the generalization ability and robustness of the resultant classifier

4.5.4.1 Classification and Regression Trees(CART)

a) Classification Trees: in this type of decision tree, the target variable is definite and the tree is used to identify the class within which a target variable would be likely to fall into.

b) Regression Trees: in this type of decision tree, the target is continuous and tree is used to predict its value.

The CART classifier is organized as a sequence of questions, where the answers will determine what the next question is going to be if any at all. The result of these questions is a tree like structure where the ends are terminal nodes at which point there are no more questions [102].

Advantages of CART:

- CART is nonparametric and therefore does not rely on data belonging to a particular type of distribution.
- Outliers in the input variables do not significantly impact CART.
- You can relax stopping rules to "overgrow" decision trees and then prune back the tree to the optimal size. This method reduces the likelihood that important structure in the data set will be disregarded by stopping too soon.
- CART incorporates both testing with a test data set and cross-validation to evaluate the goodness of fit more accurately.
- CART can use the same variables more than once in different parts of the tree. This capability can expose complex interdependencies between sets of variables.

- CART can be used in conjunction with other classification methods to select the input set of variables.

Disadvantage of CART

- Regression-type problems: are those where we try to predict the values of a continuous variable from one or more continuous and/or categorical variables.
- Classification-type problems: are those where we try to predict the values of a categorical dependent variable from one or more continuous and/or categorical variables.
- One of the main limitations of CART is that one is not able to force variables into the classifier.
- This classifier cannot be used for the estimation of average affects.
- This classifier is unstable. Little changes in the data can result in producing a completely different tree.

3.5.4.2 Random Forest

RF is considered as one of the most popular ensemble learning algorithms in the last decades, as it achieves great performance in numerous tasks with low error rate [103] This classifier is an intuitive model that provides a robust probabilistic structure for solving a number of learning tasks. It is clear that, RF efficiently generates partitions of high-dimensional attributes, and in each cell of these partitions, algorithm probability distribution is located. Therefore, they allow the estimation of any densities or arbitrary functions for clustering, regression or classification tasks.

Classification outcomes are subsequently acquired by averaging the results formed throughout all fellows of the forest, permitting the collective knowledge of the tree learners to be operationalized to form a single final decision. Number of decision trees developed based on random selection of data and random selection of variables. The methodology of RFC is described in Equations 48 and 49.

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} f(x, x_i p) \qquad (48)$$

Where x refers to the variable that partial dependence is required, while $x_i p$ is considered the other variable for data.

$$\mathrm{f(x)} = \mathrm{logt_j} - \frac{1}{\mathrm{J}} \sum_{\mathrm{k=1}}^{\mathrm{J}} (\mathrm{logt_k(y)} \qquad (49)$$

Where J refers to the whole number of classes, whereas j refers to class. In addition, $t_k$is belong the proportion of total votes for class j.

Given an M feature set, trees are constructed using m features randomly selected from the feature set at each node of the tree. The best split is calculated using these m features, whichcontinues until the tree is fully grown without pruning. The procedure is repeated for all trees in the forest using different bootstrap samples of the data. Classifying new samples can then be achieved using a majority vote.

There are significant improvements in terms of predication and classification accuracy which have resulted from increasing an ensemble of a number of trees as well as allowing them vote for the most common class. In order to grow these ensembles, RF vectors often produce growth in the number of trees in the ensemble [104]. In Figure 3.4 the RF takes this notion through combining trees with an ensemble. The tree ensembles combine predictor variables of a number of different trees to provide an aggregated prediction. Hence, in ensemble phases, the RF is a strong learner and the trees are weak learners. This classifier is efficient and effective to train and test datasets, and has integrated mechanisms for confidence in each classificationmade as well as for estimating test error.

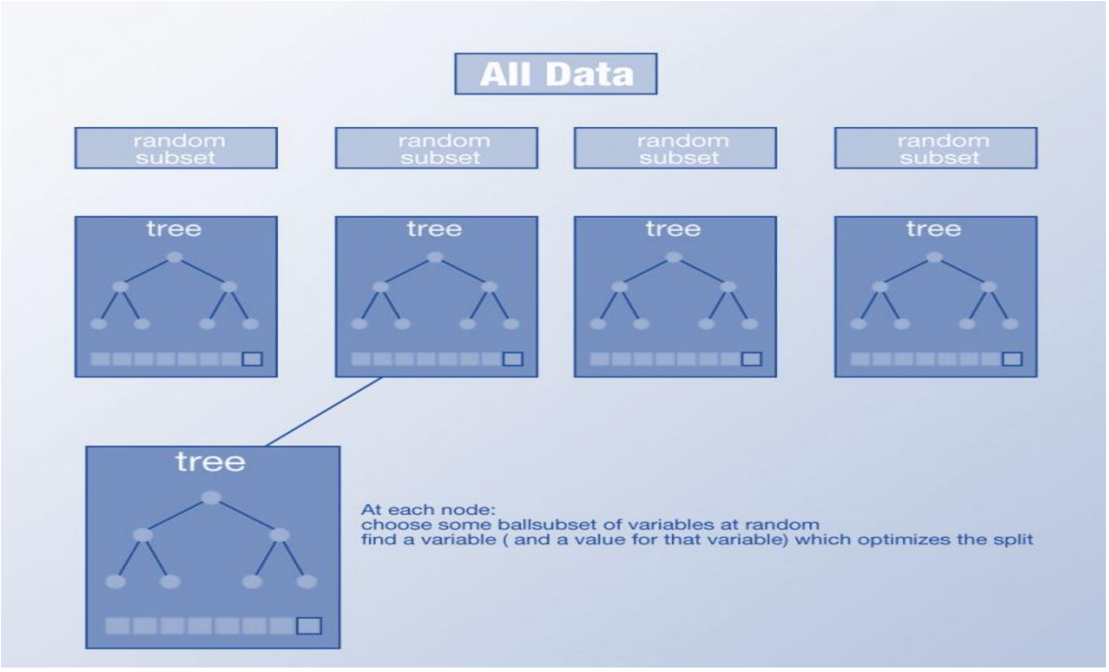

Figure 3.4 Learning process of BNNP

Advantages of Random Forest

- This method can solve both type of problems, Classification and regression.
- Random forest can handle large data sets with higher dimensionality.
- It has an effective method for estimating missing data and maintains accuracy when a large amount of data is missing.
- It has techniques for balancing errors in data sets.
- It involves sampling of the input data with replacement called bootstrap sampling.

Disadvantage of Random Forest

- It doesn't do a good job for the regression problem as it does not give precise continuous nature predictions.
- In random forest, you have very little control of what the classifier does.

3.5.4.3 Gradient Boosting Machines (GBM)

Gradient Boosting Machines are a family of powerful-learning techniques that have shown considerable success in a wide range of practical applications. They are very easy to change according to particular needs of the application. GBM provides excellent results in terms of accuracy and generalization. In addition, the GBMs offer additional visions into the resulting classifier design allowing for more investigation and analysis of the modelled effects.

This method can easily and effectively capture complex non-linear function dependencies. However, this method still has numerous drawbacks.

- Memory-consumption: the cost of storing a predictive model depends on the number of boosting iterations used for learning. For some applications, the desired number of iterations can easily range to tens of thousands. Such massive classifiers require lots of storage, though, this problem is common to all ensemble methods.

Evaluation-speed: in order to use GBM classifier to get predictions, one has to examine all the base-learners in the ensemble. Even though the base-learners are simple, when the ensemble is large, getting predictions at a fast speed can become time-consuming.

However, when the GBM ensemble is learnt, one can take full advantage of parallelization to obtain predictions.

- Parallelization: the learning procedure is sequential and has problems with parallelization. Therefore, GBMs are slower to learn. However, a different approach to parallelization of the GBMs would be to parallelize each of the boosting iterations, which can improve the evaluation speed.
- GBMs do not support fast implementation of smooth continuous base-learner that capture interactions. However, only decision trees can effectively capture non-trivial interactions between variables in reasonable computation time.

3.6 Summary of the Chapter

This chapter discussed the use of different machine learning classifiers for classification. It provided a general idea about the main concept of the machine learning and its rapid development in addition to the huge impact of using the ML through the daily human application. The main machine learning types have been also reviewed in this study with high consideration of providing a better overview of the main types, supervised ML considered the most important methodology in machine learning, which represents the ability of systems to infer a function of variant tasks depending on previously provided training data then testing the classifiers with new unseen data. Various ML algorithms have been discussed in this chapter including decision tree classifiers, neural network (backpropagation), the support vector machine and the KNN with their specific characteristics and application.

CHAPTER FOUR: METHODOLOGY

4.1 Introduction

This chapter presents a detailed discussion of the six stages of the methodology process. First stage is Collecting, defining and identifying the data, secondly, preparation and pre-processing the data. Third stage is extracting the important inputs from the data. Forth stage is structuring and coding of the main algorithms that will be used then the fifth stage will be interpreting / testing the output results and evaluate the algorithm performance and finally reporting and storing the final results. We have dealt with the health status prediction according to the data mining process depending on methods like classification and prediction utilizing different ML techniques. Our Research problem is a type of classification problem as it's based on 6 extracted features of clinical data like (PH,BDecf,BE,PCO2,AS1 AND AS5) of both pathological and normal records that are needed to predict the foetal hypoxia after birth. The assessment of tools and techniques will be the next step to be used in this research. Open access data by Physionet repository is the source of our clinical data. Initially the data preparation task was carried out. Excluding the main and target attribute, finding relationships and statistical analysis is the main elements of initial data pre-processing. The feature selection process has an important role in a successful Machine learning approach, which can lead to a classifier with better predictive ability and improved performance on unknown test data. The implementation part includes a detailed process of feature engineering. Interesting data results can be explored with appropriate bar charts and histograms to observe the correlation between attributes. The next step of exploration includes improving data quality via missing data imputing and rectifying errors. Data pre-processing provides prepared data for utilising the machine learning classifier for major analysis. We are using different machine learning classifiers like Gradient Boosting classifier (GBM); K-Nearest Neighbors (KNN) ; Support Vector Machines (SVM) with radial basis function, kernel support; CART ; Random Forest (RF) and Neural Network.10-fold cross validation have been used according to [106]as the robust method to test the classifier. Accuracy, Sensitivity (Recall), Specificity, ROC Curve area and Kappa Score are taken into consideration for the prediction performance matrix in the context of classifier evaluation. All classifiers have been evaluated and compared to find the best performing classifier. Accuracy of the classifier is more focused on in this study for the scope of this project. The Database in this study has been provided by an open source Physionet, consisting of data, which were originally collected between 27th of April 2010, and 6th of August 2012 at the obstetrics ward of the university hospital in Brno (UHB), Czech Republic. The data consisted of signal and clinical data stored in the hospital information system (AMIS) [98].

4.2 Related research

Various machine-learning techniques will be used for foetal clinical data classification, the parameters used for the diagnosis and detection of the intra-partum foetal hypoxia will be shown. Perinatal asphyxia is a major cause of neonatal and childhood morbidity and mortality in addition to the life-long disability that may happen to the foetus due to the long-time deprivation of oxygen which can cause cerebral palsy [107].

Many studies conclude that perinatal asphyxia and the unfavourable foetal outcome can be predicted, if the foetal acidosis has been proven. Foetal acidosis can be determined by specific parameters such as blood gases PO2, PCO2, umbilical cord pH at birth, base excess, base deficit of extracellular fluid and Apgar score. Furthermore foetal CTG and maternal ECG are important factors in the latest studies in detecting foetal hypoxia.

Several solutions have been proposed to diagnose foetal hypoxia depending on pH value [108]that can be provided by either the arterial blood sample or the venous blood sample [14, 109-112].

Existing research work illustrates strong association between the Low arterial cord pH and the clinical neonatal outcome [113, 114]. The challenge of this MPhil study is to find the main parameters in addition to pH value, as well as the threshold of these values that cause adverse neurological outcome in foetuses, some studies reported that to have a successful predictive outcome, the mean umbilical cord artery blood should be 7.20 Sykes et al.[115]. While in a similar study by Steer et al. [116],the mean pH was 7.26 and in a similar Dutch study by Berg et al., the mean pH was 7.30 [117]. The American College of Obstetricians and Gynaecologists recommends that the arterial cord blood's pH should be less than 7.0 as an indicator of birth asphyxia [16] Multiple studies confirm that birth asphyxia can cause brain damage even when umbilical cord arterial pH is greater than 7.0 as illustrated by Yeh et al [118], The authors studied data of 51,519 singleton, term neonates and found the ideal cord arterial pH to be between 7.26 – 7.30 for all outcomes, the study results showed 10 – 15% of the subjects had adverse neurological outcomes at a pH between 7.0 and 7.11; even for subjects with seizures and encephalopathy within the first 24 hours of life, at least half had a pH > 7.10, whereas 39% had a pH above 7.20. Another key finding of this study is that most neonates with adverse outcomes, even that of seizures in the first 24 hours, are not born acidemic (normal pH level). In addition, the authors concluded, "it appears as

though the lowest risk of any adverse outcome occurs at 7.2 – 7.3, rather than ‘the higher the better’, for there may be a higher risk at higher levels.”
Other important finding is that many babies with birth asphyxia (who often have low Apgar scores) often has a normal ph. Neonates with low Apgar scores who are acidaemic may do better in the long term than those who are not.

On the other hand, there are many studies, which used the clinical blood test of the infant in addition to other assessment tests such as the FHR information to detect the foetal hypoxia by different ML techniques. Strachan et al. [119]is one of these studies that used both clinical blood tests as well as the FHR information in his study. Strachan and his colleagues used pH value ≤ 7.15 and BDecf =8 to represent a cut-off point for the analysis to identify which individual component of the computerised analysis of the foetal heart rate will predict Acidaemia. They found that only three parameters out of six (bradycardia, deceleration and late deceleration) could predict low pH <7.15, BDecf > 8, to indicate hypoxia as an outcome. Georgieva et al. [43] and Jeżewski et al [120] indicated using Artificial Neural Networks (ANN) with clinical data and foetal heart rate data, and pH values <7.1 and 7.2 respectively as division criteria for classes. The performance quality of these studies represented by sensitivity and specificity showed acceptable results as indicated in Table 4.1. Another study by Keith et al [121] concentrated on the use of ANN to provide support to physicians and medical professionals. The objective of Keith et al.’s study is using an intelligent computer system in labour management and compares the performance with experts dealing with foetal monitoring data, patient information, and foetal blood sampling. The result was indistinguishable between the system’s performance and the experts in the 50 cases examined. In addition, Magenes et al [122] examined three neural classifiers to discriminate normal and pathological cases depending on the foetal heart rate only. The three classifiers show very promising performance towards the prediction of foetal outcomes on the set of collected foetal heart rate FHR signals [122]. Table 4.1 shows various studies using ANN and support vector machine to classify foetal health state using clinical blood sampling (pH, BDecf) as a division criterion for classes.

Table 4.1 Overview of various studies that presented classification results

Reference	***Class***	***Criteria for classes***	***Classifier***	***SE %***	***SP%***	***Others***	***Total cases***
Georgieva et al., 2013b) [79]	*2*	*pH< 7.1*	*ANN*	*61*	*68*	*AUC 0.64*	*7568*
Jezewski et al., 2010)[80]	*2*	*pH<7.20 or birth weight <10th perc.*	*ANN*	*67*	*68*	*ACC 67%*	*749*
Warrick et al., 2010) [83]	*2*	*BDecf<8;BDecf≥ 12*	*SVM*	*70*	*75*		*213*
Spilka et al., 2012)[84]	*2*	*pH< 7.15*	*SVM*	*73*	*76*	*AUC 0.78*	*217*
Georgoulas et al., 2006) [85]	*2*	*pH>7.20;pH<7.10*	*SVM*	*70*	*85*	*AUC 0.75*	*80*

In 2010, Warrick, P.A., and his colleagues who analysed data from multiple studies concluded that low arterial pH in umbilical cord blood sample was strongly associated with long-term adverse outcomes. Some of these outcomes included HIE (hypoxic ischemic encephalopathy), periventricular leukomalacia, intracranial haemorrhages, cerebral palsy and death. The study did not provide clear information about the pH level that constituted the clinically significant acidemia [123].

The American College of Obstetricians and Gynaecologists suggests that the arterial cord blood's pH should be less than 7.0 if used as a factor establishing a link between birth asphyxia and neurological injury [16]. However, this view is not supported by the literature. Multiple studies confirm that birth asphyxia can cause brain damage even when umbilical cord arterial pH is greater than 7.0 as indicated by Yeh et al.[118].

4.3 Data collection and analysis

The Database in this study has been provided by an open source Physionet. Data which were originally collected between 27th of April 2010 and 6th of August 2012 at the obstetrics ward of the university hospital in Brno (UHB), Czech Republic. The data consisted of signal and clinical data stored in the hospital information system (AMIS) [124].

4.4 primary analysis of full Dataset

- Database description: classification of hypoxic condition in foetus from medical recorded data
- Dimensions: 552 instances, 7 attributes as shown in Table 4.3
- Type: Binary classification
- Inputs: Numeric
- Output: Categorical, 2 class labels.

4.5 Data description

In this research, clinical data of 552 foetuses has been used for the classification of hypoxic cases. The clinical data includes: delivery descriptions, neonatal outcome, neonatal descriptions and information about mother and the possible risk factors. Table 4.2 includes the main parameters that can used in detecting foetal hypoxia according to ACOG and SOGC guidelines, (pH, BDecf, BE, PCO2, and the Apgar score after 1 and 5 min). All these parameters have been extracted from the clinical data and used for the classification. Using one of these parameters or a combination of different variables in the classification will be the big challenge of our study, finding the perfect classifier for the classification as well as the main parameters will give a better sight for the physicians in diagnosis of the foetal hypoxia. Profound Studying of the data is a mandatory step in the analytical process, as there are many missing data that might affect the analysis, in addition to the main goal of our research, which is finding the important clinical parameter that detects foetal hypoxia, as well as finding the variant combination of these variables to prove the hypoxic diagnosis. Our study concentrates on the main Classification of the foetuses' cases. Defining the foetuses' status as two main classes, either normal or pathological was the first step in data analysis depending on two parameters, pH value ≤ 7.15 and BDecf ≥ 8 as cut-off thresholds for the primary classification based on Strachan B.K. and Warrick, P.A. studies [119, 123] in addition to the ACOG and SOGC guidelines [16].

The following parameters are considered for the selection of the samples.

- Umbilical artery pH– is the most commonly used outcome measure, sign of respiratory hypoxia. Records with missing pH were excluded.
- Base excess (BE): the amount of strong acid (amount of H^{+}) which would need to be added or extracted from a substance in order to achieve normal blood pH level (7.2-7.4) under standard condition. It is indicate an increase in the amount of the buffer base and often used in the clinical setting as a sign for metabolic hypoxia
- Base deficit in extracellular fluid (BDecf): the amount of strong base, which would need to be added or extracted from a substance in order to return to the normal blood pH level (7.2-7.4). Moreover, the number of mEq of bicarbonate needed to restore the serum bicarbonate to 25 mEq/L at a PCO2 of 40 mmHg. It is considered a better measure of metabolic hypoxia than BE, hence in this research, records with missing BDecf values were excluded.

Both base excess and base deficit are terms applied to an analytical method for determination of the appropriateness of responses to disorders of acid-base metabolism.

Table 4.2 summarized all parameters that can have an effect on the pathological cases detection.

Table 4.2 Patients and labour outcome statistics for the CT-UHB database

Parameters	Mean (Median)	Min	Max	Comment
pH	7.23	6.85	7.47	
BE	-6.36	-26.8	-0.2	
BDecf (moll/l)	4.60	-3.40	26.11	
PCO2	7.07	0.7	12.3	
Apgar 1min	8.26 (8)	1	10	AS1 <3: 18
Apgar 5min	9.06 (10)	4	10	AS5 <7: 50

While the Table 4.3 shows the final labelled dataset with the extracted parameters that are used to train the classifiers and test their performance in detecting foetal hypoxic (pathological) cases.

Table 4.3 labelled dataset including the main parameters.

1	pH	BDecf	pCO2	BE	Apgar1	Apgar5	'Class '
2	7.14	8.14	7.7	-10.5	6	8	Yes
3	7	7.92	12	-12	8	8	No
4	7.2	3.03	8.3	-5.6	7	9	No
5	7.3	5.19	5.5	-6.4	8	9	No
6	7.3	4.52	5.7	-5.8	9	10	No
7	7.23	1.29	8.2	-3.8	8	9	No
8	7.16	4.35	8.8	-6.8	9	10	No
9	7.36	3.88	4.9	-4.6	8	9	No
10	7.18	7.6	7	-9.6	8	9	No

4.7 Data cleaning

Understanding the data was the first priority before starting the analysis, it was noted that there are some missing data in some attributes. The missing or corrupted data was amputated. The utilized clinical data did not consider time series data and hence the imputation was done by only erasing the missing data or the corrupted data without the need to have the mean or replacing the missing values. Classification of medical data depends on multiple attributes, such as the pH and base deficit to train the modules, so removing the whole instance that had any corrupted or missing data was considered. In this case, 11 instances with missing some values were removed making the final cleaned data to include541 instances.

4.8 Class Distribution and visualization

In any classification problem, it is important to know the proportion of instances that belong to each class label. This is important because it may highlight an imbalance in the dataset that may affect the whole analysis process, as well as the classifier performance, which increases the need to be addressed with rebalancing techniques. Class imbalanced datasets occur in many real-world applications where the class distributions of data are highly imbalanced. For the two-class case, without loss of generality, usually we assume that the minority or rare class is the positive class, and the majority class is the negative class. Often the minority class is very infrequent. Applying most traditional (cost-insensitive) classifiers on the dataset can easily predict everything as negative (the majority class). This was often regarded as a problem in learning from highly imbalanced datasets [125].

The pre-processingstep of the full dataset that was used in this study using different explorers such as (weka explorer) shows the number of instances that belong to each class as well as the percentage weight of each class. Table 4.4

Table 4.4 Output class distribution

Class type	*Label*	*Count*	*Weight*
Pathological	*Yes*	*57*	*57.0*
Normal	*No*	*484*	*484.0*

Data visualization is the fastest and most useful way to summarize and learn more about the data [126].

Plots of the relationships between attributes can give an idea of attributes that might be redundant, resampling methods that may be needed and ultimately how difficult a classification problem might be [127].

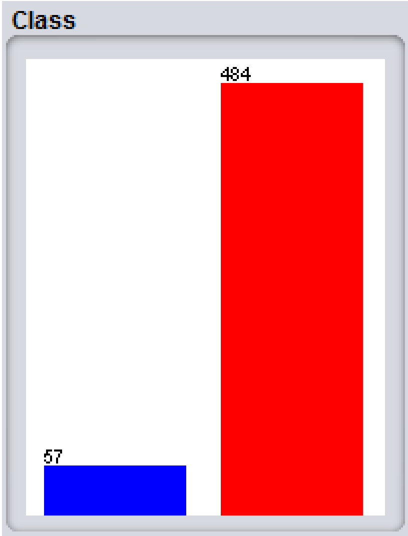

Figure 4.1 Class distributions of the utilised data

The data used in this reserach is a real-world data set which is predominately composed of normal examples with only a small percentage of abnormal examples.Studying the data results of class distribution that were made by statistical and graphical analysis demonstrate a vast imbalance between the two class labels , 484 cases were normal compared to only 57 pathological cases as shown in Figure 4-1.

The Machine learning community has addressed the issue of class imbalance in a specific way which is the re-sampling of the original dataset, either by oversampling the minority class and/or under-sampling the majority class [128-130].

Resampling methods are commonly used for dealing with the class-imbalance problem. Although such classifiers can be very simple to implement, tuning them most effectively is not an easy task. In particular, it is unclear whether oversampling is more effective than undersampling and which oversampling or undersampling rate should be used. According to Estabrooks et al. study, [131] combining different expressions of the resampling approach is an effective solution to the tuning problem Therfore, SMOTE technique (Synthetic Minority Over-sampling Technique) was first utilised in this studyfor over sampling the data to acquire a balanced dataset. Then resampling the data was done by randomly selecting samples and making a new balanced dataset to train and test the classifiers .

4.9 Feature selection

Feature selection is the process of selecting the main subset of features in order to improve the machine learning algorithms' performance and to minimise the cost of the classifier construction and the time consumed by repeating of the classification process to find relevant results.It is known that different features have different quantities of information. Thus to maintain the high performance of classifiers, the most relevant features according to the foetal hypoxic pysiology are used whilst discarding irrelevant, redundant, or noisy ones.

The aim of this section is to select a subset of full term infants' clinical features that will provide more information to describe the data which will help in diagnosing foetal hypoxia.

The following criteria were taken into account during the preparation ofthe data:

• **Umbilical artery pH** (PH) – is the most commonly used outcome measure, as a sign of respiratory hypoxia. Records with missing pH were excluded. Since the main intention was to use a database that could give beneficial results in comparison of different automated classifiers to detect the foetal hypoxia, the samples that included umbilical artery pH in addition to the other umbilical vessels blood tests are used.

• **Base excess (BE)** – Base excess (BE) is an expression of the metabolic component of the acid base disorder. It considers the amount of strong acid that must be added to each litre of fully oxygenatedblood to return the pH to the normal value at a temperature of 37°C [24].

• Base deficit in extracellular fluid (BDecf) –the amount of strong base, which would need to be added or extracted from a substance in order to return to the normal, blood pH level (7.2-7.4). Moreover, the number of mEq of bicarbonate needed to restore the serum bicarbonate to 25 mEq/L at a PCO2 of 40 mmHg is considered a better measure of metabolic hypoxia than BE, hence in this research, records with missing BDecf values were excluded. Rosén et al. [44] indicate that BDecf is a better measure of metabolic hypoxia than BE.

• **Apgar score 1and Apgar score 5**

Apgar scoring is a rapid assessment tool of the newborn health status, first developed by *Dr. Virginia Apgar* [134]. Low Apgar score can be related to different foetal-maternal conditions, such as foetal malformation, immaturity, infection or meconium aspiration. The Apgar score as shown in Figure 4.2 represents an expression of the infant's physiological condition at one point in time, which includes subjective components(colour, heart rate, reflex irritability, muscle tone and respiration), each component has a score between 0 and 2where 2 is the strongest rating. Thesum of all five components will give the final apgar score in specific time.

Apgar score at 1 min of infant's life helps the practitioner to decide whether the baby needs immediate resuscitation. While the five-minute Apgar score helps the practitioner in determining how the infant is progressing during the first 5 minutes of the baby's life and whether the newbornhas responded to any initial medical intervention.

According to the Neonatal Resuscitation Program guidelines,a score of 7 to 10 is reassuring and still considered normal at this point while a score of 4–6 is reflecting a moderate abnormality and a score of 0–3 in the term infant after 5 minutes may indicate hypoxic state especially if it is combined with other foetal hypoxic signs such as altered level of the umbilical or occipital blood pH value and/ or foetal heart rate abnormality.

	0 Points	1 point	2 points	Points totalled
Activity (muscle tone)	Absent	Arms and legs flexed	Active movement	
Pulse	Absent	Below 100 bpm	Over 100 bpm	
Grimace(reflex irritability)	Flaccid	Some flexion of extremities	Active motion(sneeze,coug h, pull away	
Appearance	Blue, pale	Pink body, blue extremities	Completely pink	
Respiration	Absent	Slow, irregular	Vigorous cry	

Sever distress 0-3
Moderate depressed 4-6
Excellent condition 7-10

Figure 4.2 Apgar scoring system

4.10 Relationship between the various parameters and its effect on foetal hypoxia detection

According to Z. Henderson and his colleague J. L. Ecker, serious adverse sequelae in the neonatal period are rare after birth with umbilical cord pH greater than 7.0 or base excess less acidotic than -12 mmol/l [135].

Follow up of infants with cord pH above 7.0 suggests no adverse effect of acidosis on cognitive outcome [96]. Even at pH below 7.0 most infants will still have good progress without any remarkable illness. In this respect cord pH or base excess alone are poor predictors of foetal hypoxic outcome [136].

Most infants with evidence of intra-partum hypoxia do not develop serious long life damage. In a series of around 14.000 new-born infants with routine cord blood gas analysis, King et al. [98] identified pH <7.0 in 58 (0.4%) infants who were born at 35 weeks' gestation or more, 5-mim Apgar score ≥7. 37 of these 58 infants were triaged after birth to the routine postnatal nursery. They were followed closely and none developed clinical manifestations of hypoxic-ischaemic injury. This suggests that infants who are in good clinical condition at birth and are free of cardiopulmonary disturbance do not require neonatal unit admission or detailed investigation purely on the basis of low cord pH.

However, the combination of low pH at birth with other abnormal clinical feature becomes strongly predictive of adverse sequelae. Perlman and Risser et al. 1996 showed that a combination of cord pH <7.0, 5-min Apgar score of ≤5 and necessity for intubation had an 80% positive predictive value for the development of seizures [137].

Portman et al. [138] developed and validated a scoring system for predicting multi-organ impairment following perinatal asphyxia. They found that a score combining a measure of cardiotocographic abnormality, umbilical arterial base excess, and low 5-min Apgar score are more strongly associated with morbidity than any individual factor. In a separate study the score showed a positive predictive value of 73% and negative predictive value of 99% for predicting impairment of three or more organ systems [139].

On the other hand, Goldaber et al. [9] studied the relathionship between low PH level of umbilical artery and adverse neurological damage among 3506 term, singleton infants with cord arterial pH<7.20. Neonatal death was much more likely at pH <7.00. The cut-off at which seizures became more likely was pH <7.05, and for unexplained seizures was pH <7.00. Therefore, they recommended that a realistic value for significant pathological acidaemia was pH <7.00.

Moreover, Williams et al. [140] found that a threshold of pH <7.00 was the best predictor of neonatal seizures when compared with other indices.

Another study by Low and his colleagues [13] shows an association between metabolic acidosis and multi-organ impairment. Using a scoring system for renal, central nervous system, respiratory and cardiovascular morbidity that confirm the metabolic disturbance as a good predictor of multiorgan for both term and preterm infants, they found that the threshold of metabolic acidosis associated with increased risk of newborn complications in term infants was a base excess of minus 12 mmol/l.

The relation between the hypoxia and ischaemic encephalopathy had an advanced concentration by Goodwin et al. [141] who found that 12% of infants with cord pH <7.0, have hypoxic- ischaemic enchephalopathy , while 33% of the infants with cord pH <6.9 , 60% with cord pH <6.8, and 80% with cord pH <6.7.35.

Several other studiesshowincreasing morbidity rate once severe acidosis is developed [142-144]. A static measurement of acid–base status provided by cord blood analysis during and after birth has an important role in dignosing and following up infants as shown by Casey et al's study who found that prenatal infant acidosis with (pH <7.20) and persistance of 2 h beyond delivery had a poorer outcome than those who have recovered from the acidosis after birth [145].

Recent data suggest that persisting lactic acidosis is associated with severe encephalopathy and may be a reflection of the presence and severity of seizures [146, 147].

4.11 Experimental classifiers of foetal data classification

In machine learning techniques, there are two types of classifiers which can be classified into linear and non-linear clarifiers. The classifier can be called linear when the decision boundary is separated by a hyper-plane on the feature space. Let's consider, our dataset involves two classes (xx and ++) distributed as demonstrated in Figure 4.3 (a). These two classes are discriminated into the two classes, one drawn with 'x' in the left side of the line and ++'s on the right side of the line. From the Figure 4.3 below, it is obvious to be called linearly-separable due to each class being on one side. Some examples of linear classifiers are: Naive Bayes, SVM (with linear kernel), and Linear Discriminant Classifier, and Logistic Regression. Perhaps there a number of advantages of using this method. It is unlikely to get overfitting compaired with the non-linear, particularly when utilised in high dimensions task. Linear classifiers are an effective and useful procedure that is always able see the relationship between variables, with all the benefit that is brought by this method, it is easy to implement, inexpensive to train, and has good generalizability.

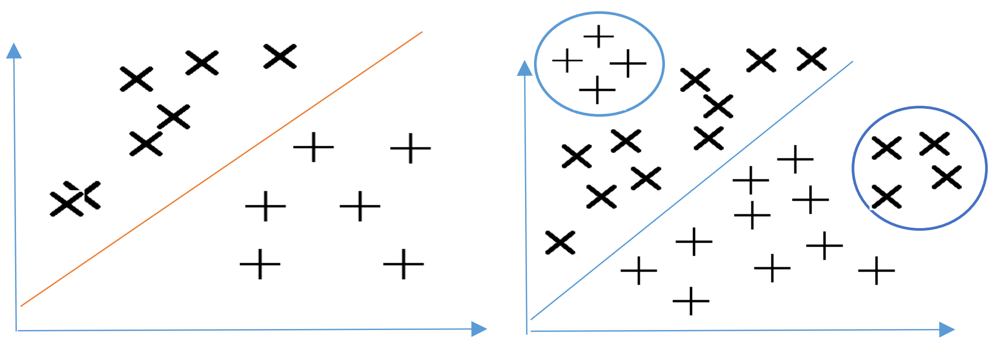

Figure 4.3 Linear (a) verse Non-linear (b) classifier classification problems

On the other hand, nonlinear classifiers are more efficient and accurate than linear classifiers to deal with the clinical data. This is due to the fact that most of the medical data sets are clustered with unseparated classes. This technique is a set of classes that is impossible to be approximated well with linear hyper-planes. Another example is shown in Figure 4.3(b) where the vast majority of class ‘x’ and class ‘+’ can be discriminated through drawing a straight line in the exact location, with different features selection method and their charictaristic. However, there are a few points from both classes which are present in between the data points(shown inside a circle). It is obvious that drawing one line cannot separate all the points of both classes correctly. In our experiment, we tried to experiment linear and non linear classifiers to distinguish which can give the best and most accurate classificationof the classes

The most popular classifiers for the medical data classification are Random Forest classifier (RFC), Artificial Neural Network (ANN), support vector machine (SVM). Each classifier applies a learning algorithm to examine the relationship between features and class label of the input datasets. However, the main objective behind the learning algorithm is to build a classifier that is able to predict the target value that was previously unknown.

Table 4.5 shows six well-known classifiers, which have been used to classify the data including both linear and non-linear methods. These classifiers are the Gradient Boosting classifier (GBM); K-Nearest Neighbors (KNN) [112] ; Support Vector Machines (SVM) with radial basis function, kernel support; Random Forest (RF) [104, 148] and Neural Network.

Table 4.5 Experimental classifiers of foetal data classification

Methods	Classifier	Category
GNM	Stochastic Gradient Boosting	Non linear
CART	Classification and regression trees	Linear
KNN	K-Nearest Neighbor	Non linear
SVM	Radial basis function Kernal support	Non linear
RF	Random Forest	Non linear
NN	Backpropagation Neural Network	Non linear

4.12 General Approach of Classification process

Data classification is a two-step process, consisting of a learning step (where a classification model is assembled) and a classification step (where the classifier is used to predict class labels for given data)as shown in (Figure 5.1) [149]. The main concept of the classification process is learning a target function f that maps each attribute set X to one of the predefined class labels y .Defining the specific function or the mapping that separate the data classes is considered the core of any classification process. The classification process can be viewed as the learning of a mapping or function. $y = f(X)$, that can predict the associated class label y of a given sample X.

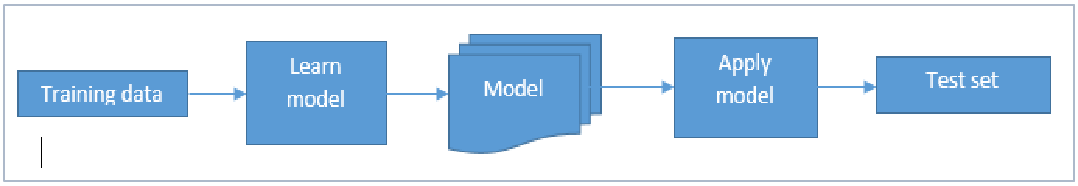

Figure 5.1 classification process map

In the first step, learning step (or training phase), the classification algorithm builds a classifier by analysing the given collection of samples (**training set**); a training set is made up of database samples and their associated class labels. Each sample, is characterized by a tuple (X, Y), where X is the attribute set and Y is a special attribute, designated as the class label [99]. A tuple, X, is represented by an n- dimensional attribute vector,$X = (X_1, X_2 \ldots X_n)$ where the n measurements made on the tuple from n database attributes, respectively, $(A_1, A_2, \ldots., A_n)$. Each tuple, X, belongs to a pre-determined class defined by another database attribute called the **class labelled attribute.** Each value of the class labelled attribute can be considered a distinct-value that serves as a categorical (nominal) class.

For class labelling, we've used two attributes, the pH level as an indicator of the respiratory hypoxia in addition to the BDecf value to detect metabolic hypoxia and cover all the types of the oxygen de-saturation that may affect foetuses during labour.

In our data analysis, the mapping or the function that separates the class label of a given tuple is represented in the form of classification rules that identify the hypoxic state of the foetus as being either hypoxic (YES) or normal (NO). The classifier firstly trained with training set including all the six variables.

On the other hand, in the second step of the classification process, test data are used to estimate the accuracy of the classification rules. The accuracy of the classifier can be evaluated by applying the learned classifier on the new dataset to overcome the over fitting problem that may happen during the learning. Therefore, a test set is used (made up of test instances and the associated class labels) which considers independent data sets. The rules can be applied for new data classification, if the accuracy of the test data classification is acceptable. Evaluation of the classifier performance is based on the counts of test records correctly and incorrectly predicted by the classifier.

A rule predicting the first row in the training set may be represented as following:

$$IF(pH \leq 7.15\ ANDBDecf > 8\ )THENHYPOXICSTATE = YES$$

4.13 Summary of the chapter

This chapter discussed the main steps of the project methodology. The main stages of the pre-processing procedure and data preparation for the analysis have also been included, such as: cleaning the missing or corrupted data, visualization, balancing the dataset and a brief description of each feature, extracted from the provided medical data. Many studies and researches relating to perinatal asphyxia have been also reviewed. Studying the results of all these studies helped in determining the most important parameters that we used for the classification. The followed section shows the relationship between different biomedical features, such as the pH level, BE and the Abgar score, as well as reviewing some results of different studies to detect the thresholds of these features that help in diagnosing foetal hypoxia and eventually classify the cases we have. In the last section we discussed the difference between the linear and non-linear classifiers, to find which type would have more advantages in medical data analysis. However, in this study we will examine the performance of both linear and non-linear classifiers and compare the results to find which type of learning classifier will be the best for real medical data sets.

CHAPTER FIVE: IMPLEMINTATION AND ALGORITHM EVALUATION

5.1 Introduction

A brief definition of the main evaluation methods will be discussed in the first section of this chapter including the confusion matrix and classifiers performance metrics as well as other statistical methods. This chapter will include the definition of classification as the main concept of ML problem that will be used in our data analysis. Three rounds of classification procedures will be performed with various datasets by the same six ML classifiers.

5.2 Performance metrics

To evaluate the performance of the machine learning techniques for the classification of clinical hypoxia data, the receiver-operating characteristic (ROC) curve [150]and the confusion matrix are utilised to characterize the capability of the classifiers. The ROC plot is a graphical method used to describe classifier performance by mapping the sensitivity and one minus the specificity for each value of the classification cut-off threshold. The overall performance and capability of predictive classifiers can be measured using a range of statistical metrics including sensitivity (the true positive rate), specificity (true negative rate) and classification accuracy. These metrics are calculated based on the terms listed in the confusion matrix (as shown in Table 5.1) to identify the individual and relative performance of classifiers in the medical data analysis.

Confusion matrix is an explicit way to display the classification outcomes; it plots the true class of interest (i.e. gold standard) in a binary class classification against the predicted class [151]. These terms are represented as true positive (TP), false positive (FP), true negative (TN) and false negative (FN).

Table 5.1 Confusion Matrix

		Predicted classes	
		Positives	*Negatives*
Actual Classes	*Positives*	*TP*	*FN*
	Negatives	*FP*	*TN*

Sensitivity, also called the true positive rate (TPR) as shown in Table 5.2, is the classifier's ability to identify the class of interest correctly, while the specificity (also called true negative rate TNR) refers to the classifier's ability in excluding the other class correctly. Classification accuracy is the overall correctness of the predictive ability of the classifier, which is the sum of correct predictions (both true positives and true negatives), divided by the total number of predictions made [152]. Classification accuracy is commonly the first step in evaluating the quality of the classifier. However, it could be misleading in some cases especially with a large class imbalance situation [153].

Table5.2 Performance Metrics

Metrics	*Abbreviation*	*Computation*	*Scope*
Sensitivity	*TPR*	*TP/ (TP+FN)*	*[0, 1]*
Specificity	*TNR*	*TN/ (TN+FP)*	*[0, 1]*
Accuracy	*ACC*	*(TP+TN)/ (TP+TN+FP+FN)*	*[0, 1]*

Other scalar values resulting from statistical methods are considered to provide a summary of classifier performance, such as the area under the ROC curve (AUC) and Kappa values, which measure the agreement between predictions and the actual classification in the data set. Higher AUC and Kappa values indicate better classifier performance [154].

The ROC analysis is a standard technique that is designed to summarise the predictive performance of binary classification models. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) measurements at diverse decision thresholds in two-dimensional ROC space [149].

An ideal classifier would have a point in the upper North West corner of the ROC space, which means that the classifier has accurately classified all the positive and negative classes. In contrast, a classifier with random prediction performance will fall along the diagonal line of the ROC curve, in which TPR and FPR are equal over all different decision thresholds. The ROC curve analysis is widely accepted in the medical field, where it provides perfect details of the classifier's predictive performance particularly with imbalanced data. From this graphical representation, the optimal decision boundary can be selected.

5.3 Data Mining Classification

Classification is a form of data analysis that extracts classifiers describing important data classes [99]. These models, called classifiers, can predict categorical class labels. The classification process can be used as a descriptive procedure where the classifier can work as an explanatory tool consists of predicting a certain outcome based on a given input. In order to predict the outcome, we divided our datasets in to training and test sets so the algorithm processes a **training set,** containing a set of attributes ($PHlevel$, $BDecf, BE, PCO_2, Apgar_1, Apgar_5$)in addition to the respective outcome, usually called goal or prediction attribute. The main target of the algorithm is trying to discover relationships between the attributes that would make it possible to predict the outcome. Next the algorithm is given a data set not seen before, called **prediction or test set**, which contains the same set of attributes, except for the prediction attribute – not yet known. The algorithm analyses the input and produces a prediction. The prediction accuracy defines how "good" the algorithm is. Classification has numerous applications, including fraud detection, target marketing, performance prediction, manufacturing and medical diagnosis.

In a medical database used, the training set would have relevant patient information recorded previously, where the prediction attribute is whether or not the foetus had an oxygen saturation problem. Tables 5.3, 5.4 illustrate the training and prediction sets of the data set. Table 5.3 shows a sample of the datasets used in the classification of foetal health condition into hypoxic and normal states that is represented by (YES, NO), respectively.

The attribute set includes multiple parameters that are currently used by the medical staff to detect the foetal hypoxic state ($pH\ level, BDecf, BE, PCO_2, Apgar_1, Apgar_5$) in addition to the class label attribute, which is presented discretely during the analysis

Table 5.3 Medical databases "Training dataset"

pH	**BDecf**	PCO_2	**BE**	$Apgar_1$	$Apgar_5$	**HYPOXIC STATE**
7.14	8.14	7.7	-10.5	6	8	YES
7.22	5.3	7	-7.3	5	9	NO
7.15	9.75	7	-11.7	9	9	YES
7.2	5.5	7.5	-7.3	9	9	NO

Table5.4 medical bases" predicting dataset"

pH	**BDecf**	PCO_2	**BE**	$Apgar_1$	$Apgar_5$	**HYPOXIC STATE**
7.32	4.73	5.3	-5.9	9	10	?
7	11.1	10.1	-15	6	7	?
7.09	9.58	8.3	-12.1	8	8	?
7.33	3.08	5.6	-4.3	9	9	?

5.5.1 General data classification process:

The process shown in (Figure 5.2 a, b) represents the general classification rule of the medical data of hypoxic foetuses (The data sets are simplified for illustrative purposes). Six classifiers have been trained with the **training dataset** including all the attributes that related to the health state diagnosis and the final outcome of the foetuses' health condition as shown in Figure 5.2 a. The next step will be testing the classifier with unseen datasets (**test dataset)** to predict the final outcome as shown in (Figure 5.2 b).

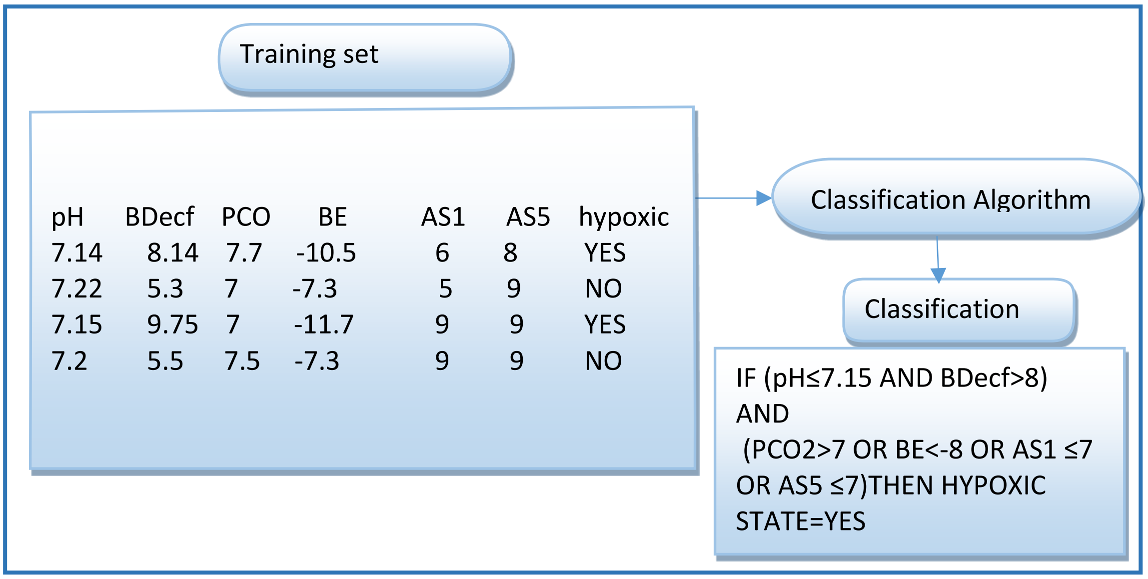

Figure 5.2 (a) Learning stage: training data analysed by a classification algorithm

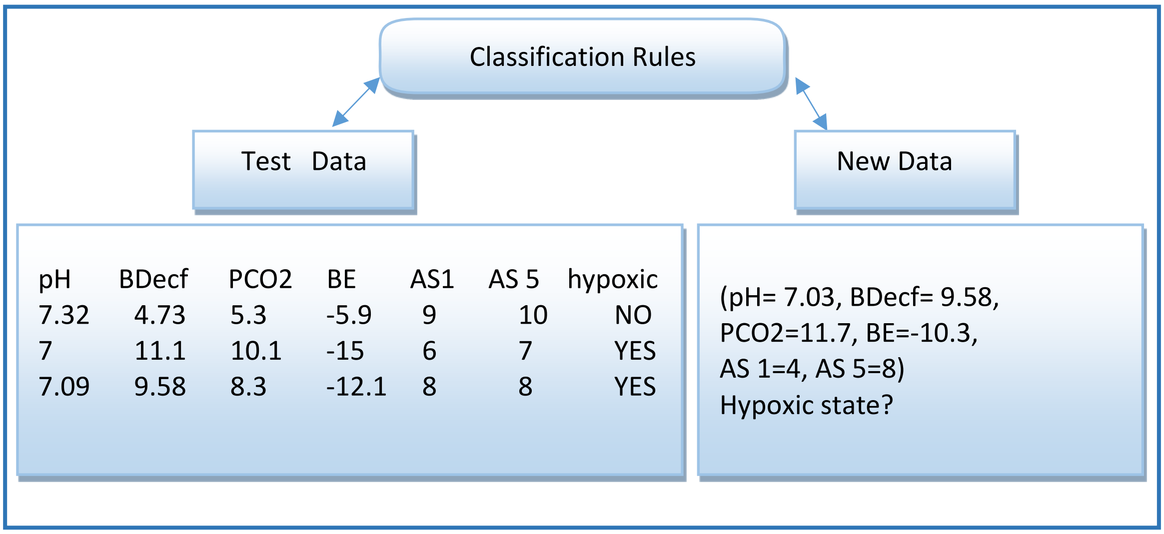

Figure 5.2 (b) Classification stage: test data are used to estimate the accuracy of the classification rule.

5.5.2 First round of classification

The main idea of Classification process in this round has been done using two parameters out of the six parameters to train the classifiers. The two parameters (pH level, BDecf value) that have excluded from the provided clinical data considered the most important parameters in detecting the respiratory and metabolic types of the foetal hypoxia as mention in the previous chapter. Collecting results from using machine learning techniques applied on these data will provide a better idea about how they would help in using real world clinical data and provide an accurate decision that might help the physicians in diagnosing the condition of the new-borns. Using the pH and BDecf values only as the threshold for data labelling for this trial plays a very important role in the classification process as they are known by physicians and all the medical expertise in addition to many references indicating that these two parameters have a very high level of accuracy in detecting foetal hypoxia especially in the early hours of the new-born's life [120]. In this case the results of the machine learning techniques that are provided from these experiments will concentrate on the accuracy level of these classifiers in detecting the labelled hypoxic cases from the normal ones. Using different types of classifiers in data classification will help in providing a better view of how these techniques will classify the clinical data and provide a precise decision of the clinical state of the new-borns. In the first set of experiments, the general idea of the classification rule will be as follows:

$$\mathrm{IF}(\mathrm{pH} \leq 7.15 \text{ AND } \mathrm{BDecf} > 8 \text{ AND } (\mathrm{PCO2} > 7 \text{ AND } \mathrm{BE} < -8 \text{ AND } \mathrm{Apgar\ 1} \leq 7 \text{ AND } \mathrm{Apgar5} \leq 7)) \text{ THEN HYPOXIC STATE} = \mathrm{YES}$$

FIRST ROUND OF CLASSIFICATION

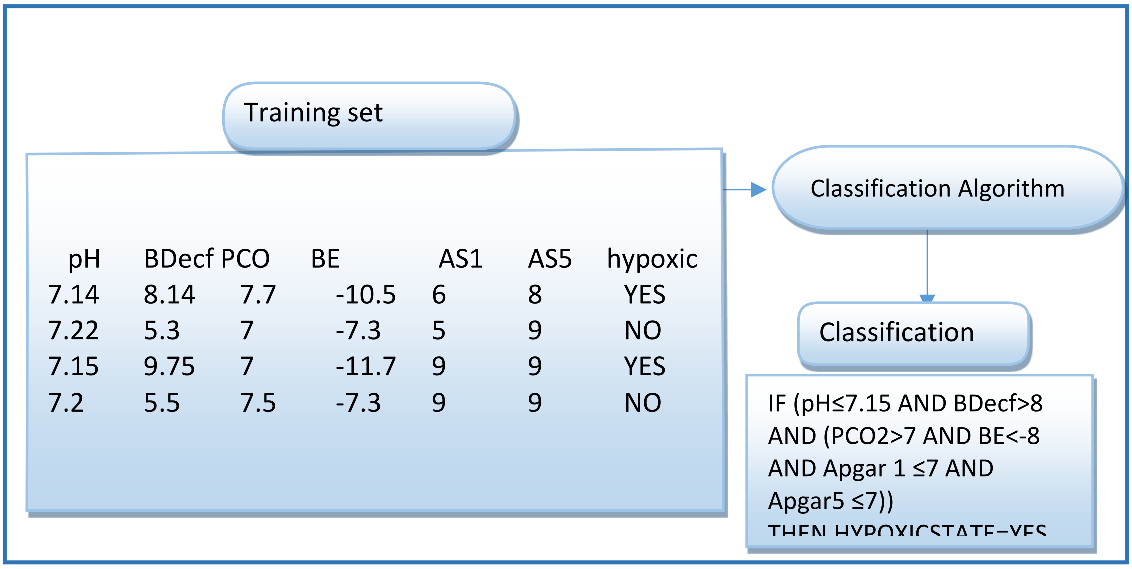

Figure 5.3 (a) Learning stage: training data analysed by a classification algorithm

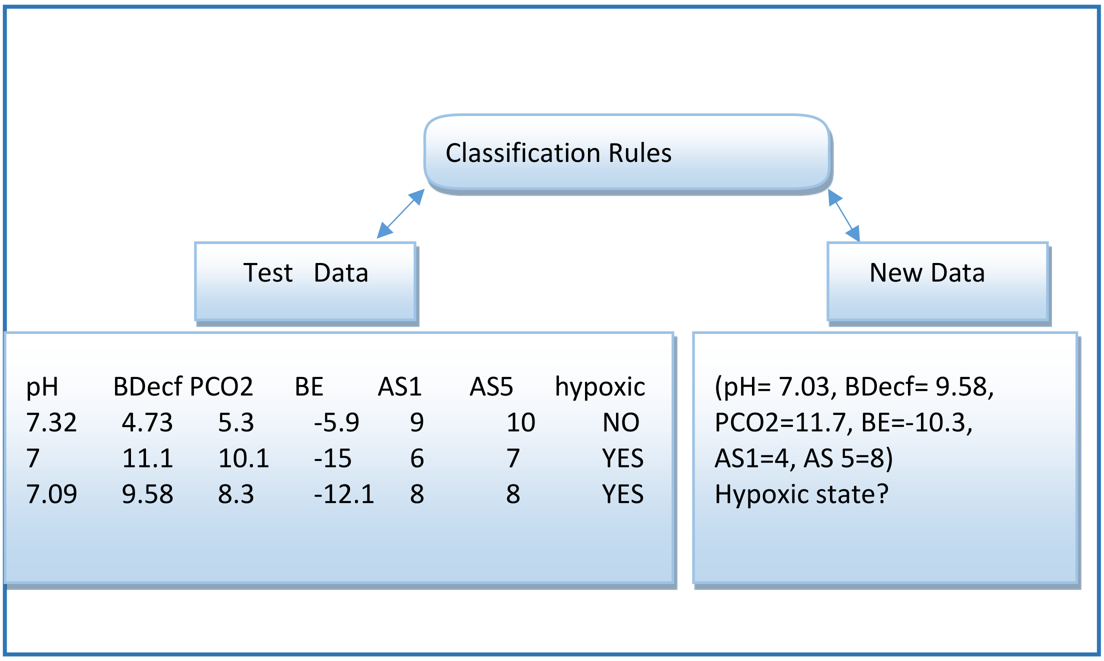

Figure 5.3 (b) Classification stage: test data are used to estimate the accuracy of the classification rules

5.5.3 Second set of classification experiments

The next step in our classification process is testing the same classifiers by the same clinical data but with different parameter sets. The new training and test sets in this experiment are equipped using clinical values, which are also important in detecting the hypoxic state. However, we will depend on the BDecf values in labelling the cases in addition to the other parameters but not the pH values as the first round. Repeating the whole process with the six classifiers again and checking the accuracy levels of each classifier. Then compare the performance of each classifier and how it affects the final decision.

In this trial the classification rule that we used in labelling will be as follows:

$$\mathrm{IF}\ (\mathrm{BDecf} > 8)\ \mathrm{THEN\ THE\ HYPOXIC\ STATE} = \mathrm{YES}$$

The idea of using various set of parameters in this experiment is to find the classification power of the machine learning techniques that were applied in the first set of experiments.

The training set in this set of experiments includes the same clinical data but depending on five parameters only in detecting the final outcome as shown in Figure 5.4. (a). The same six classification algorithm will be trained firstly with the training dataset, then tested again with different and unseen datasets to estimate the accuracy level of each classifier in detecting the foetal hypoxia as a final outcome as shown in Figure 5.4. (b).

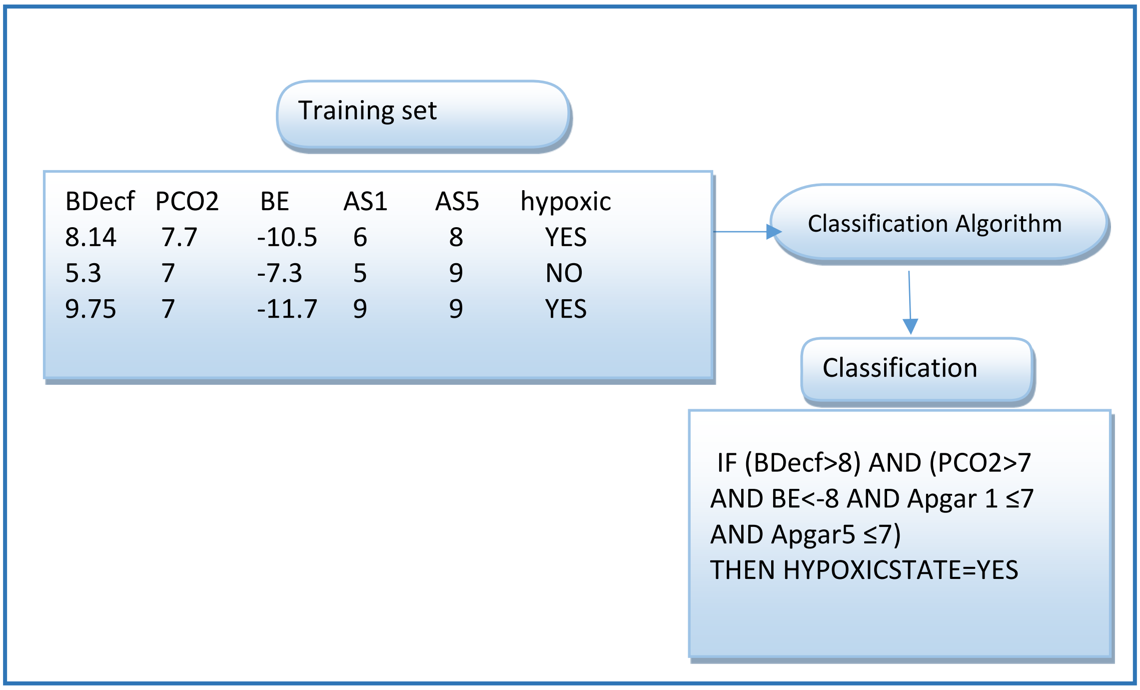

Figure 5.4 (a) Learning stage: training data analysed by a classification algorithm

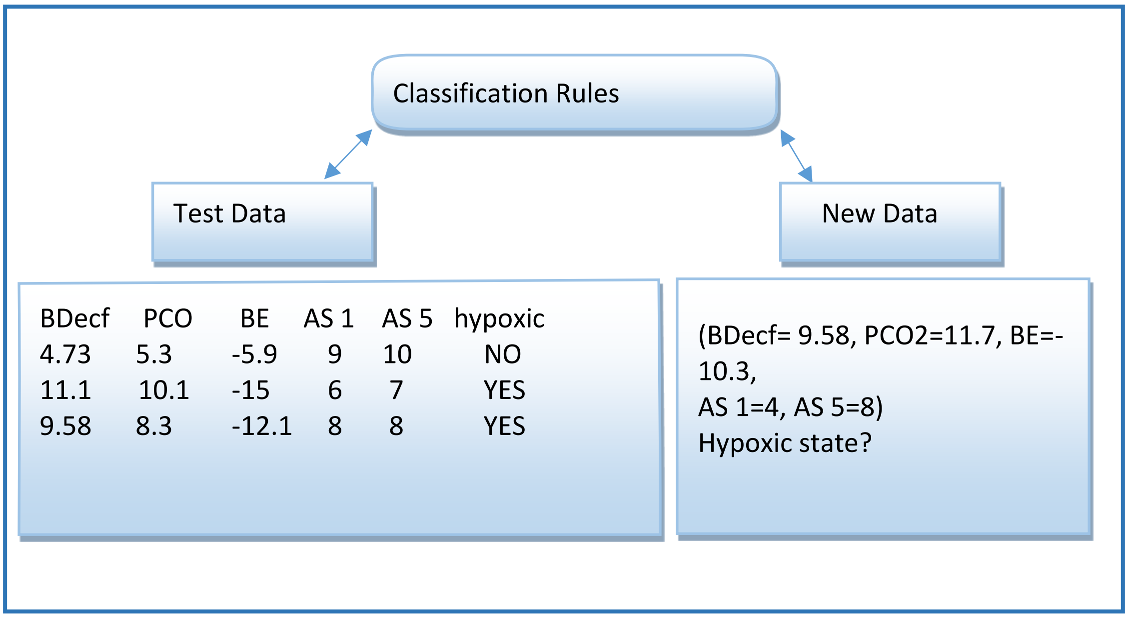

Figure 5.4 (b) Classification stage: test data are used to estimate the accuracy of the classification rules

5.5.4 Third set of classification experiments

The last set of experiments represents the innovative set of our work, which means implementing the same machine learning techniques but with the elimination of the two parameters (pH, BDecf) that were represented as the most important parameters in the previous two trials and let the classifier detect the hypoxic cases from the normal cases depending on the other parameters (BE, PCO2, AS1, AS5) which also showed a strong relation to foetal health condition.

The classification rule of this trial will be "from medical point of view" as follows:

$$\mathrm{IF}(\mathrm{PCO2} > 7 \text{ AND } \mathrm{BE} > -12 \text{ AND Apgar } 1 \ \leq 7 \text{ AND Apgar5} \\ \leq 7)\mathrm{THENHYPOXICSTATE} = \mathrm{YES}$$

This set of experiments can be considered the most important stage of the whole classification process, as it will show various levels of classifier accuracy in detecting foetal hypoxia based on a new set of parameters not used in the labelling process as shown in Figure 5.5 (a). Using the same clinical data but without previous labelling will show how the classifiers can detect hypoxic cases from the normal cases, in addition this experiment will inform physicians if they can depend on other values in detecting hypoxic state apart from the (pH, BDecf) levels as shown in Figure 5.5 (b).

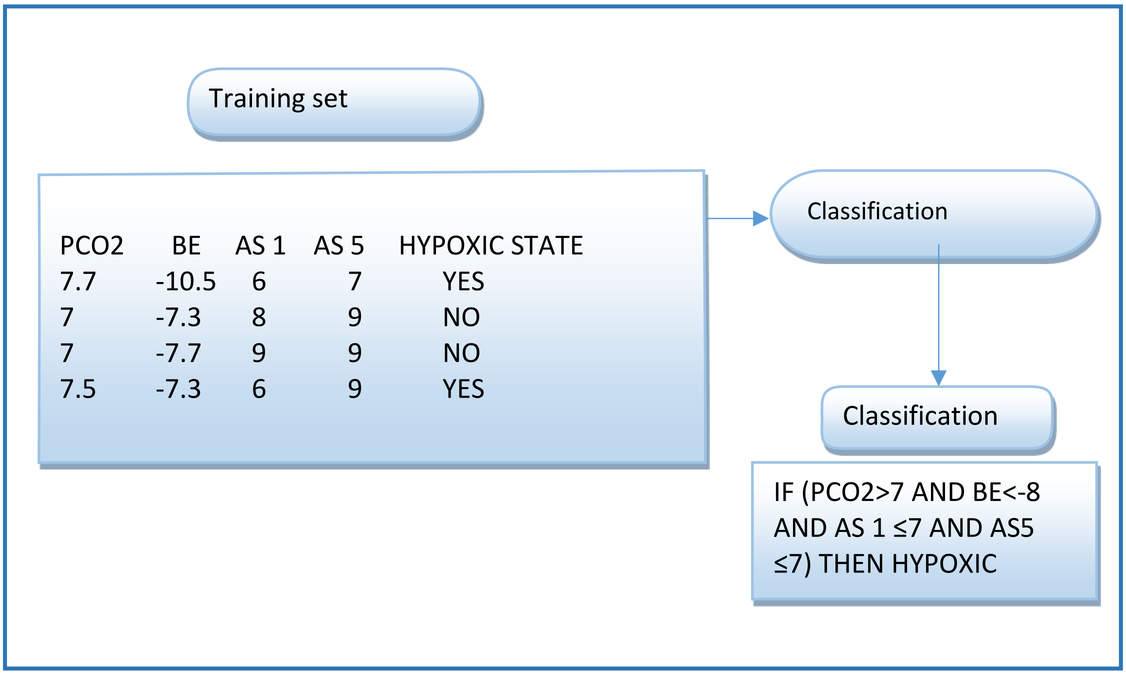

Figure 5.5 (a) Learning stage: training data analysed by a classification algorithm

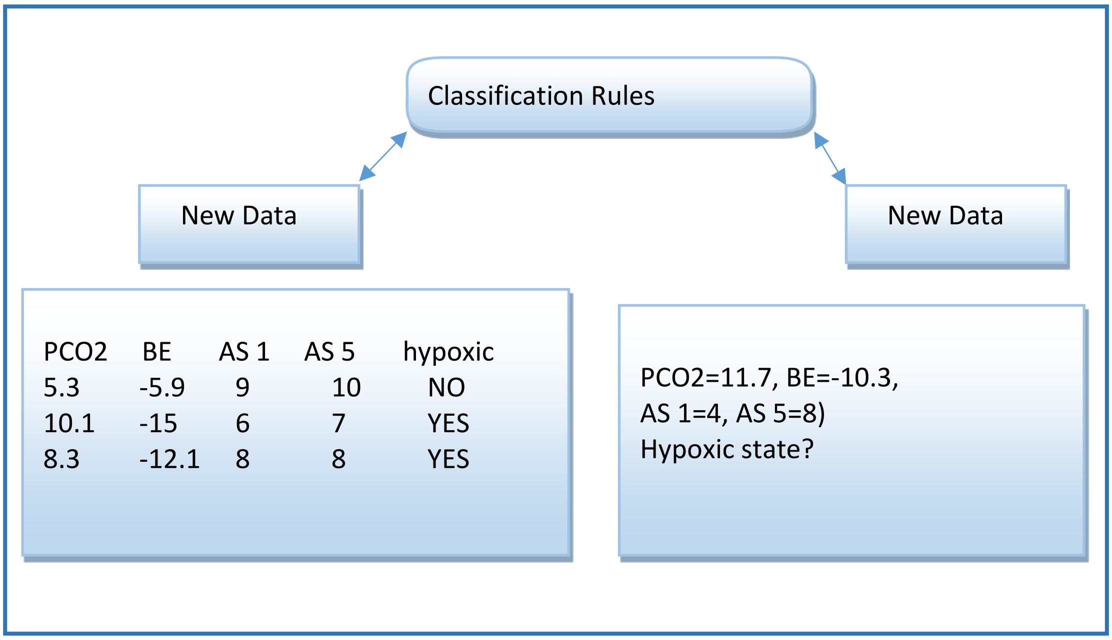

Figure 5.5 (b) Classification stage: test data are used to estimate the accuracy of the classification rules

5.5 Chapter Summary

In this chapter, different measures were shown to evaluate the performance of the classification algorithms, which are used in this work. In addition to the main concept of the classification problem, three different rounds of experiment have been implemented to evaluate the classifiers' accuracy level in detecting the foetal health condition at birth. A novel set of parameters have been used in these different experiments that may affect the final decision of the physicians and help in diagnosing the foetal distress condition with the help of machine learning techniques; simulation results will be presented in the next chapter.

CHAPTER SIX: SIMULATION RESULTS AND ANALYSIS

6.1 Introduction

The intention of this chapter is to present the simulation results of six classifiers used in this work, which include the Gradient Boosting classifier (GBM); K-Nearest Neighbours (KNN); Support Vector Machines (SVM) with radial basis function, kernel support; Random Forest (RF) and Neural Network.

The remainder of this chapter is organized as follows. Section 6.1 represents the simulation results with the previously mentioned classifiers using the clinical data and full set of parameters $(\mathrm{pH\ level}, \mathrm{BDecf}, \mathrm{BE}, \mathrm{PCO_2}, \mathrm{Apgar_1}, \mathrm{Apgar_5})$. This is followed by section 6.2, which illustrates the results of the classification of the clinical data using five parameters$(\mathrm{BDecf}, \mathrm{BE}, \mathrm{PCO_2}, \mathrm{Apgar_1}, \mathrm{Apgar_5})$. The section 6.3 shows the simulation results of using the same classifiers with only four parameters to detect the importance of these parameters in detecting foetal health condition at birth. The analysis of the results is included in section 6.2, 6.3 and 6.3. Section 6.4 shows detailed discussion on the simulation results of this research work. Finally, the chapter is summarized in section 6.5.

6.2 Classification results with full set of parameters

This section represents the simulation results with the previously mentioned classifiers using the clinical data and full set of parameters $(\mathrm{pH\ level}, \mathrm{BDecf}, \mathrm{BE}, \mathrm{PCO_2}, \mathrm{Apgar_1}, \mathrm{Apgar_5})$which signify the full umbilical blood profile samples $(\mathrm{pH\ level}, \mathrm{BDecf}, \mathrm{BE}, \mathrm{PCO_2})$ and the first foetal assessments that are done by the physicians at first and fifth minutes after birth $(\mathrm{Apgar_1}, \mathrm{Apgar_5})$.

Splitting the full dataset into training and test sets will give the ML classifiers the opportunity for full training using all the parameters to find the relationship between each attribute and the final decision of the diagnosis. Table 6.1 shows the overall performance of the machine-learning classifiers with respect to the classification of foetal distress and hypoxia by using the test dataset. KNN classifier achieved the highest sensitivity by detecting the true positive cases with sensitivity value of 93%, followed by RF and GBM that showed similar sensitivity performance, followed by NNET, and CART, which showed again similar values with respect to sensitivity analysis. CART reached higher Kappa and specificity than NNET. Although KNN showed the highest sensitivity, it shows low Kappa measure and specificity analysis. RF classifier obtained the highest classification accuracy and area under the ROC curve. CART shows slightly lower result than RF classifier with respect to accuracy and area under the ROC curve. RF and CART classifiers have both reached the full range of specificity analysis result with 100%. SVM classifier obtained the lowest sensitivity, while KNN obtained the lowest Kappa measure.

As shown in Table 6.1 and Figure 6.1 the experimental results with a full set of parameters identify the Random forest classifier to be the most accurate classifier with 95% for the area under the ROC curve as shown in Figure 6.2 .The second classifier was the backpropagation based neural network with high predictive ability represented by 91% for both ACC, AUC. The simulation results showed that the GBM classifier achieved similar AUC as the RF classifier but less accuracy. Both classifiers had the ability to detect (0.87) the true positive cases (pathological), however, the GBM has failed in recognizing 11% of true negative cases (normal) as its specificity was (0.89) compared to the outstanding performance of random forest in detecting the true negative cases with specificity. On the other hand, SVM showed good prediction ability with AUC of 80%, but indicated less prediction accuracy in comparison to other nonlinear classifiers

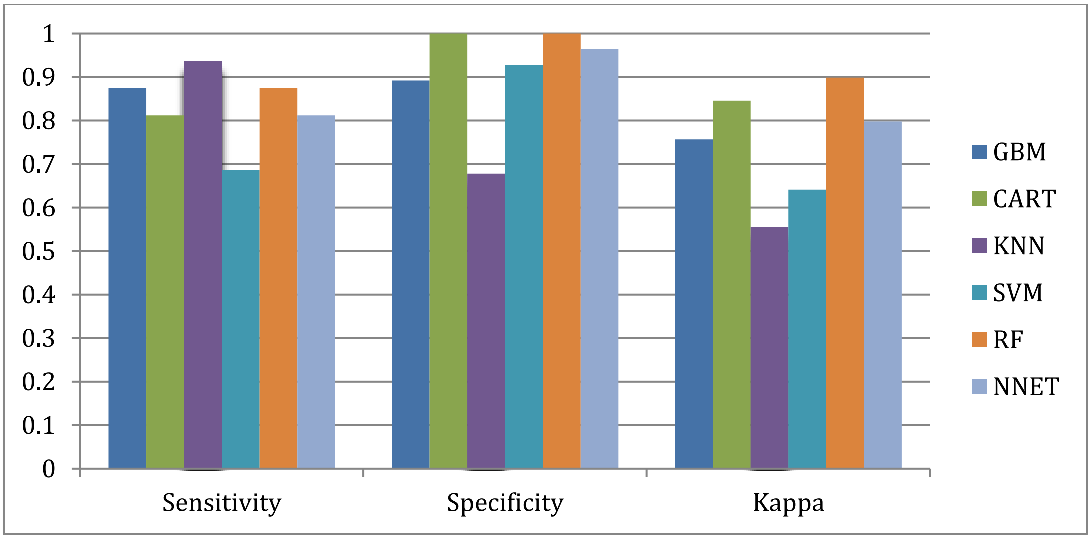

Figure6.1 classifiers performance "using six parameters

Table6.1 performance results of all classifiers using six parameters

Classifiers	Sensitivity	Specificity	Kappa	Accuracy	AUC
GBM	0.875	0.892	0.757	0.886	0.939
CART	0.812	1	0.846	0.931	0.942
KNN	0.937	0.678	0.556	0.772	0.880
SVM	0.687	0.928	0.641	0.840	0.901
RF	0.875	1	0.899	0.954	0.943
NNET	0.812	0.964	0.798	0.909	0.910

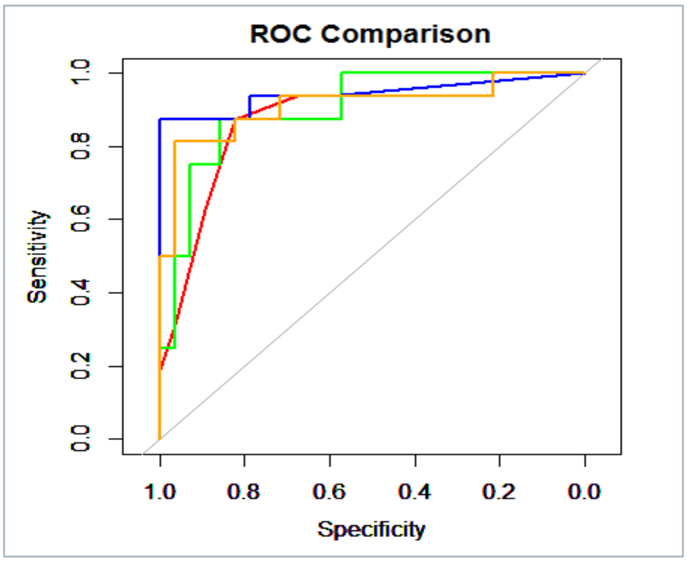

Figure 6.2 Roc plot for all classifiers

Table 6.2 ROC Summary (predictions)

CLASSIFIER	**ROC**
GBM	0.9397321
CART	0.9419643
KNN	0.8805804
SVM	0.9017857
RF	0.9430804
NNET	0.9107143

With respect to the data parameters, we considered both pH and BDecf as predictor variables and a cut-off threshold in labelling the classes prompting a further study of the influence of these variables on the performance of the classifiers. Applying different parameters gives the classifier the opportunity to learn and study the relationship between each variable and the labelled class as well as studying the correlation between all the parameters and its effect on the detection of the pathological cases. As indicated from Figure 6.3, all the classifiers identified pH values as the most important variable for the data classification, with the exception of the SVM and KNN, while the BDecf was the most important variable detected by both SVM and KNN. BDecf was the second most important variable detected by NNET and the third most important for both RF and GBM.

All the classifiers apart from KNN and SVM have been able to detect the hypoxic cases depending on all variables that were provided by the dataset. BE parameter is considered the second most important variable in detecting the hypoxic cases, which can be justified according to the important relationship between the pH, BDecf and BE levels that were explained in section 2.5 "patho-physiological of the foetal hypoxia". Any reduction in oxygen perfusion to the foetus will cause carbon dioxide accumulations in the foetal blood that cause an increase in the partial pressure of carbon dioxide (PCO 2) and a concomitant decrease in pH level producing what we call "respiratory acidosis".

Continuous hypoxia deprives the foetus of the required oxygen to perform the aerobic reactions inside the cells and tissues , resulting in accumulation of organic acids with the accumulation of pyruvic and lactic acids as a result of anaerobic metabolism and subsequently developing “the metabolic acidosis”, which can be detected via the(BDecf and BE levels)

Studying the acid base state of the foetus can identify the early sign of respiratory acidosis and metabolic acidosis.

Physicians usually use Umbilical cord blood sampling to detect five important thresholds to diagnose foetal asphyxia (pH, PCO2, PO2, BDecf and BE) [41]. Our study uses another important variable that affects the diagnosis of foetal hypoxia after birth to improve the final decision. Apgar score, represents an expression of the infant’s physiological condition at one point in time, which includes subjective components(colour, heart rate, reflex irritability, muscle tone and respiration) [42].

Using pH in detecting foetal hypoxia is considered an important indicator to quantify perinatal asphyxia [43]. However, pH alone will not differentiate between the respiratory and metabolic acidosis as it increases in both conditions as shown in Figure 2.1. The last threshold of the umbilical cord blood sampling is the base deficit/base excess which can provide a better differentiation between the respiratory and metabolic acidosis as they are both normal in the respiratory acidosis while in case of metabolic acidosis the base deficit will increase and the base excess will decrease [44].

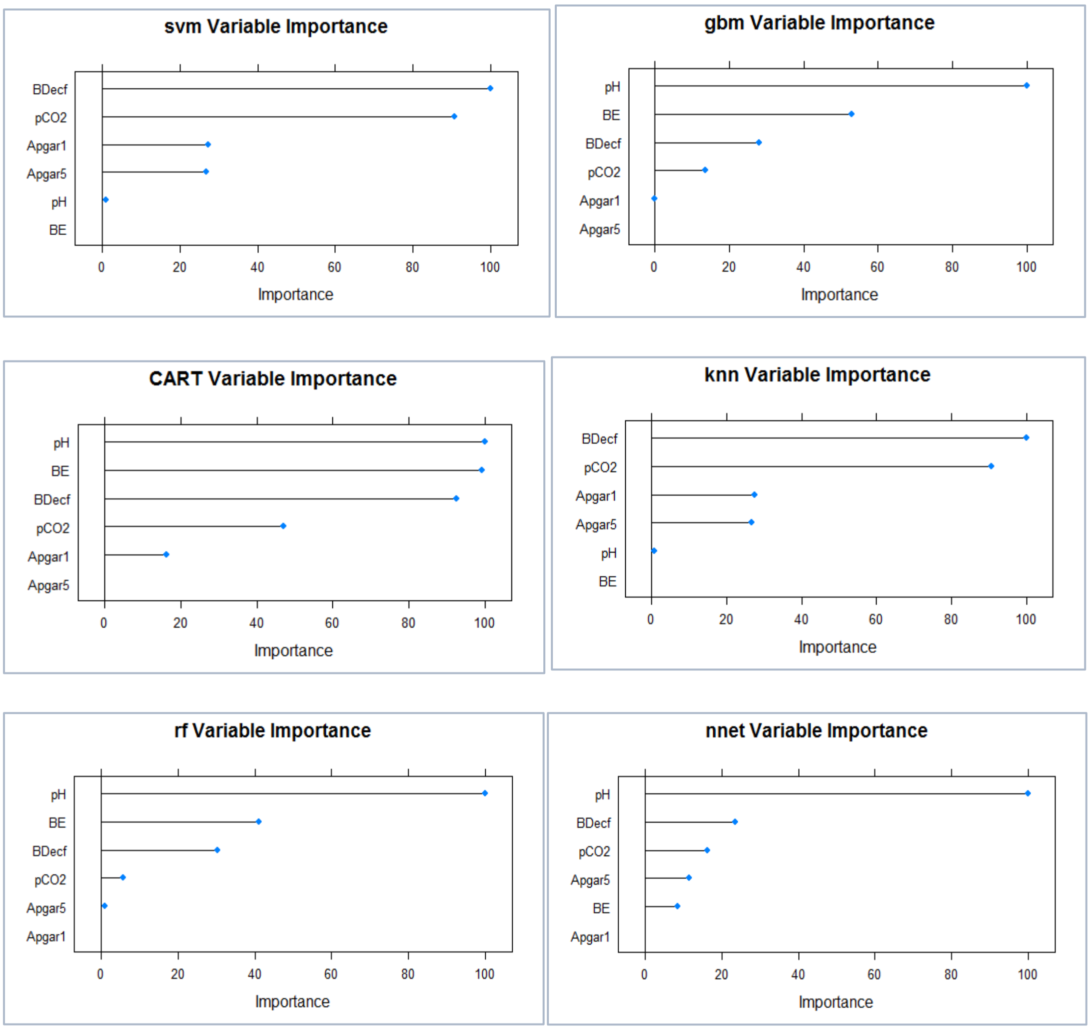

Figure 6.3 variable importance plotted figures

6.3 Classification results using five parameters

In the second type of analysis, we have used all data parameters except pH, in order to give the previously learned classifiers the opportunity to detect the hypoxic cases without the main variable (pH). This set of experiments will test the classifiers with a new data set to examine its efficiency and reliability when used by physicians without pH value. Similar to the first set of experiments, Table 6.4 illustrates the main performance measures of the classifiers when KNN achieved the highest sensitivity (93.7%), while SVM achieved the lowest sensitivity with only 56.2%, which seems no better than a random guess. The second highest sensitivity was shown by GBM CART, RF and NNET obtained equal sensitivity of 75%. CART and RF showed the highest specificity of more than 95%. Followed by NNET and GBM that showed similar results to each other. CART and RF showed the highest classification accuracy in this set of experiments, while RF have shown superior results than the CART using the AUC value as shown in Figure 6.4.

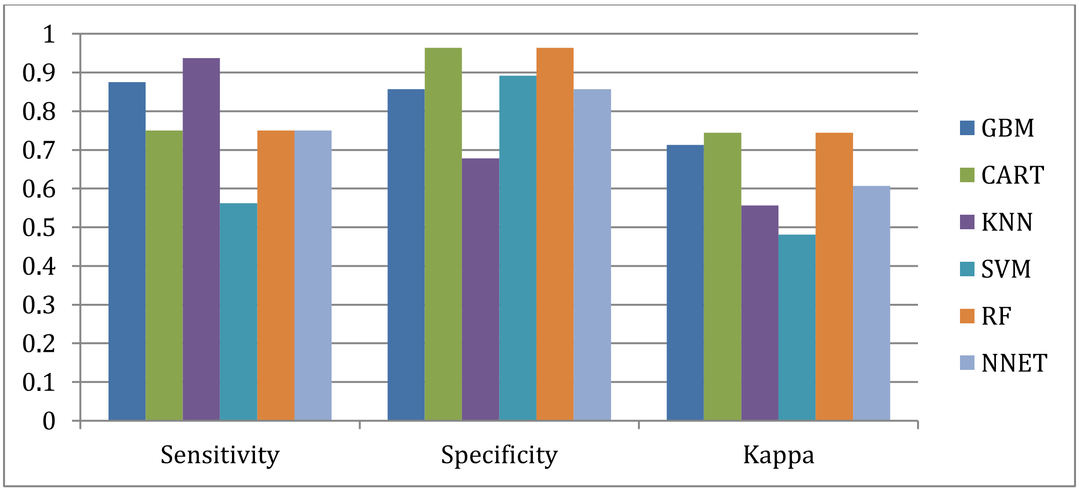

Figure 6.4 classifiers performance using five parameters

Table 6.3 performance metrics for the second trial

Classifiers	*Sensitivity*	*Specificity*	*Kappa*	*Accuracy*	*AUC*
GBM	*0.875*	*0.857*	*0.713*	*0.863*	*0.930*
CART	*0.750*	*0.964*	*0.744*	*0.886*	*0.881*
KNN	*0.937*	*0.678*	*0.556*	*0.772*	*0.885*
SVM	*0.562*	*0.892*	*0.481*	*0.772*	*0.890*
RF	*0.750*	*0.964*	*0.744*	*0.886*	*0.929*
NNET	*0.750*	*0.857*	*0.607*	*0.818*	*0.886*

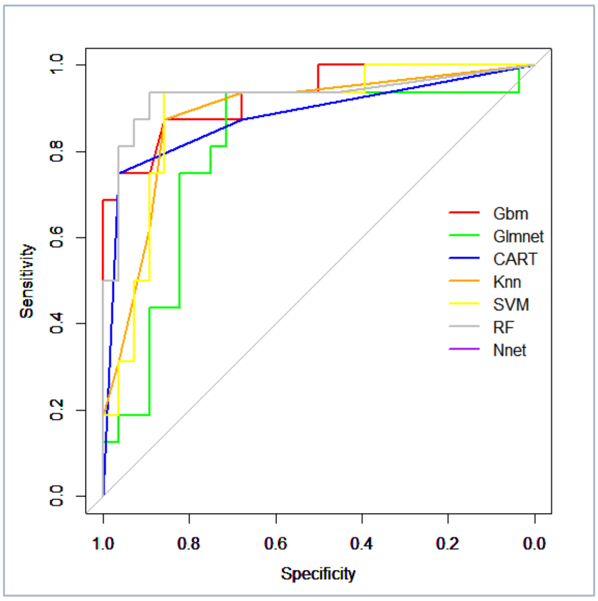

Figure 6.5 ROC plots of all classifiers in the second trial

Table6.4 ROC results for classifier comparison

CLASSIFIERS	ROC
GBM	0.9308036
CART	0.8816964
KNN	0.8850446
SVM	0.8906250
RF	0.9296875
NNET	0.8861607

6.4 Classification using four parameters

The third round of analysis and evaluation of four parameters namely PCO2, BE, Apgar1 and Apgar5 has been utilised. As indicated by Table 6.5, the highest sensitivity was achieved by KNN classifier, followed by GBM classifier that achieved sensitivity above 81%. Sensitivity of 75% was equally obtained by SVM and NNET. CART and RF classifiers achieved the lowest sensitivity with 62.5%. The majority of classifiers improve their performance by achieving highest specificity, except KNN. CART and RF showed identical results with respect to all performance metrics, with the exception of AUC values. Almost all classifiers obtained classification accuracy of more than 80%, with the exception of NNET that showed the lowest classification accuracy.

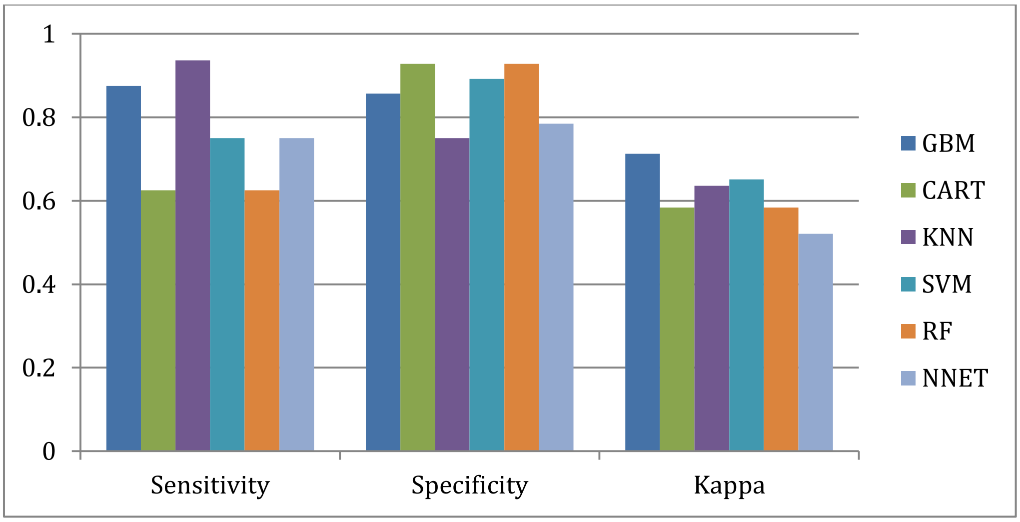

Figure 6.6 classifiers performance using four parameters

Table6.5 performance metrics for the third trial

Classifiers	*Sensitivity*	*Specificity*	*Kappa*	*Accuracy*	*AUC*
GBM	*0.875*	*0.857*	*0.713*	*0.863*	*0.920*
CART	*0.625*	*0.928*	*0.584*	*0.818*	*0.846*
KNN	*0.937*	*0.750*	*0.636*	*0.818*	*0.877*
SVM	*0.750*	*0.892*	*0.651*	*0.840*	*0.859*
RF	*0.625*	*0.928*	*0.584*	*0.818*	*0.918*
NNET	*0.750*	*0.785*	*0.521*	*0.772*	*0.857*

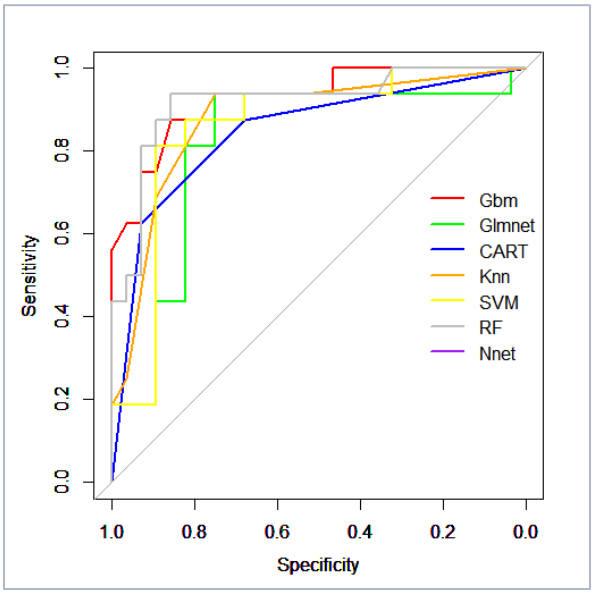

Figure 6.7 Roc plot of all classifiers in the third trial

6.5 Discussion

This study starts with a full set of parameters, where we consider all six data parameters in the classification process. Machine-learning classifiers rank these six parameters according to their importance with respect to the classification of foetal distress and hypoxia. In order to obtain additional insights into each classifier, we used the cross validation re-sampling method that is based on repeatedly drawing samples from a training set of observations and re-fitting a classifier on each sample. From the medical point of view, foetal hypoxia can be detected using two main parameters, namely pH and BDecf [120], therefore our goal is to detect whether we can detect the foetal hypoxia using pH level as well as other parameters such as BE, PCO2, Apgar1 and Apgar5.

The majority of classifiers, except SVM and KNN, have ranked pH parameter as the most important parameter in the classification. All the Parameters (pH, BDecf, PCO2, BE, AS1 and AS5) have been given an identical order by three classifiers namely RF, CART and GBM, regardless of their importance. In this context, pH can be considered as the most relevant and important parameter for the classification. Hence, this might indicate that KNN and SVM have been over-fitted toward (i.e. BDecf and PCO2) parameters as they both ranked these parameters as the most important variables in detecting the hypoxic state while underestimating the pH parameter that has been ranked by them as a useless parameter as shown in the Figure 6.7

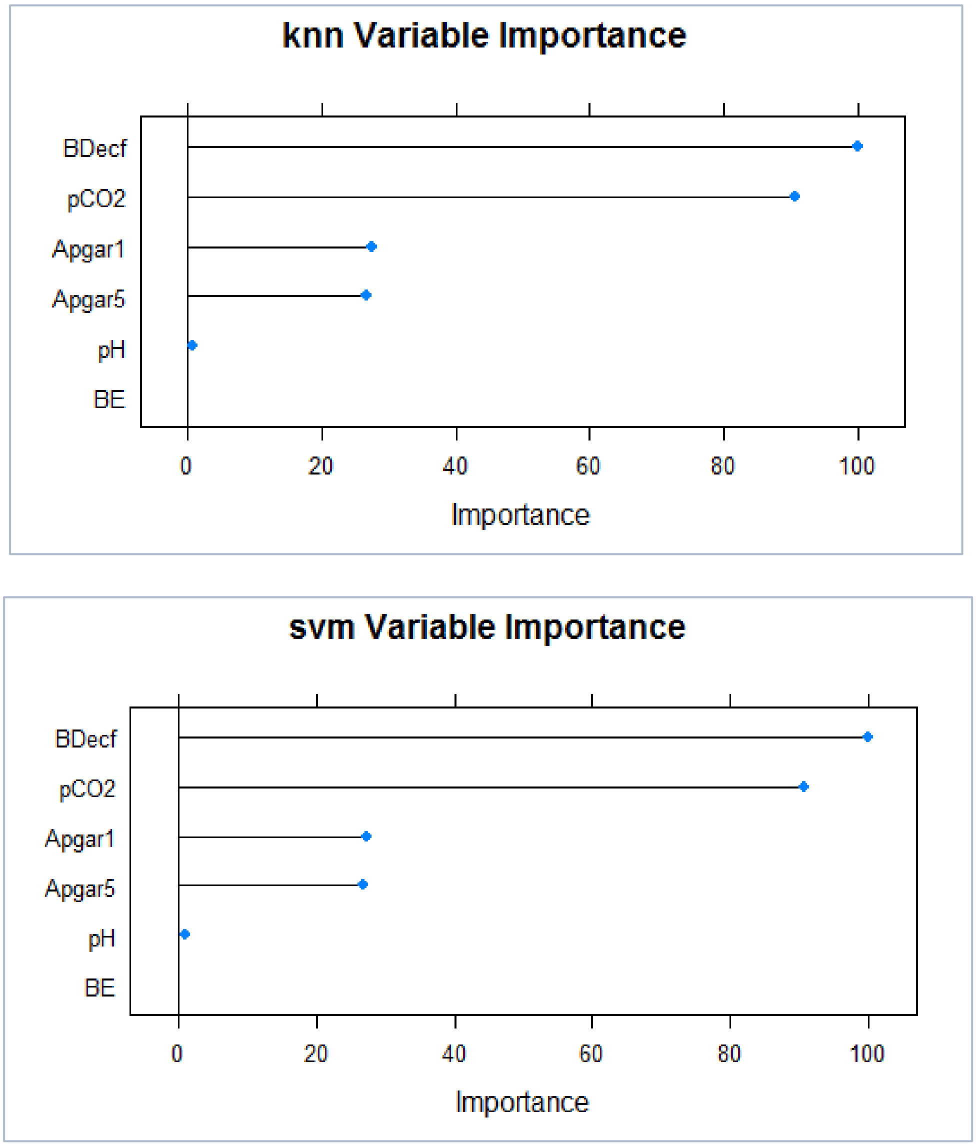

Figure 6.8 (a) SVM variable importance, (b) KNN variable importance

Understanding the full patho-physiological changes that can happen during the hypoxic state gives us a conclusion that it's not reasonable from the medical point of view that changes in both PCO2 and the BDecf can be observed without any change of the pH value, because in the case of hypoxia, elevated PCO2 means that the foetus is producing more carbon dioxide than can be eliminated through circulation. In other words, carbon dioxide is not readily diffusing from the umbilical artery and capillaries into the maternal placenta and maternal circulation. An accumulation of carbon dioxide with typically low PO2 will make the foetal body consider a specific mechanism to manage the oxygen deprivation and shift to a compensatory mechanism that will utilise anaerobic metabolism inside the living cells instead of depending on the oxygen to produce the required energy. This mechanism will cause accumulation of lactic acid and therefore increases the blood acidity (lowering the pH level) in addition to elevation of the BDecf level as a trial to buffer the acidity of the blood [44]. So it's not realistic that the classifier detects the hypoxic cases from a six parameters data set depending on elevated PCO2 level and BDecf changing only, with no detection of any change in the pH level. Furthermore, according to Cook et al. [155] and Peng et al. [156], any changes in the limit of carbon dioxide is less informative because mothers can spontaneously hyperventilate during the labour which may low PCO2. In the absence of compromise to the placental perfusion by mother and foetus there is a linear relationship between maternal and foetal PCO2 [155]. However the changes in foetal pH associated with brief hyperventilation are small. Maternal hyperventilation lowers foetal PO2, which may affect the normal values.

We can notice that tree based classifiers, i.e. CART and RF, have shown similar performance with respect to all quality measures as shown in Table 6.5 . However, RF shows slightly higher AUC results. This is mainly because RF is taking the concept of CART a step further via the generation of dozens of trees and using a range of trees and splitting data parameters and instances into smaller subgroups, and then uses thumbs up of the majority of trees to consider classification. In contrast to CART, which uses all of the parameters along with the whole training dataset to build a classifier, RF can select an arbitrary sample of the data and determine a particular subset of parameters to build each tree individually. In this way, RF will have more accurate results then CART as a single tree.

NNET has shown good results, especially in the first set of experiments using a full set of data parameters. In NNET, there will be no clear representation of how the classification of foetal distress and hypoxia has been performed. Conversely, NNET is widely used for medical data classification and regression. It has been proved to achieve good results particularly when normalising the data in the pre-classification stage.

KNN and SVM over-fitted BDecf and PCO2. These two parameters have been proven, from a medical point of view that, they are not reliable to be an indicator of the hypoxic state with normal pH level

The result of the using the same ML classifiers using input parameters will be presented in this section. The main aims of using different parameters are as follows. From the machine learning point of view, it can provide different set of results for comparison.

From the medical point of view, the aim is to detect foetal hypoxia using other parameters apart from pH and BDecf such as BE PCO2, APGAR 1 and APGAR5 and study their effect in diagnosing the foetal hypoxia.

Table6.3 shows the simulation results using 5 parameters excluding pH value.

The results show that decision tree classifiers and neural network can detect foetal hypoxia better than other classifiers as they show high accuracy levels, which strengthens the previous results when the same ML classifiers used all six parameters. The BDecf parameter is still considered a very important variable in the classification process in addition to the BE level that plays an important role in detecting the hypoxic state (especially the metabolic acidosis) of the foetus by almost all the ML classifiers

With respect to the variable importance study, the NNET classifier, in contrast to the decision tree classifiers indicated that APGAR5 could play a role in detecting the foetal hypoxia. However, research studies indicated that from the medical point of view, this parameter has no prediction power for detecting the hypoxic state. APGAR5 score is considered the last variable in detecting foetal hypoxia and it is usually used to confirm the results that already detect the hypoxic state by other variables such as the APGAR 1 for instance and other umbilical blood profile samples that confirm the biochemical disturbance of the foetal body.

Other factors that may lower the prediction ability of the Apgar score alone in detecting the hypoxic state are the effect of the maternal sedation or anaesthesia, congenital malformations, gestational age and trauma. Elements of the score such as tone, colour, and reflex irritability can be subjective, and partially depend on the physiological maturity of the infant. The healthy preterm infant with no evidence of asphyxia may receive a low score only because of immaturity, which may affect the physician’s final decision.

The final conclusion of this study gives us the valuable achievement of the decision tree and NNET classifiers in the medical detection of foetal hypoxia compared to the KNN and SVM, which failed in detecting any important variable or any relationship between the parameters that could be confirmed medically. KNN and SVM classifiers show the lowest results of both accuracy and Kappa measures compared to other classifiers.

Table 6.6 variable importance of all classifiers using 5 parameters

CLASSIFIER	BE	BDecf	PCO2	APGAR1	APGAR5
GBM	34.5	10.7	8.05	1.713	1.33
CART	33.3	31.06	15.82	4.9	42.9
KNN	0.87	0.8	0.32	0.32	0.11
SVM	0.87	0.8	0.32	0.32	0.11
RF	27.5	8.7	4.57	0.38	0.36
NNET	41.4	19.9	17.3	12.7	8.5

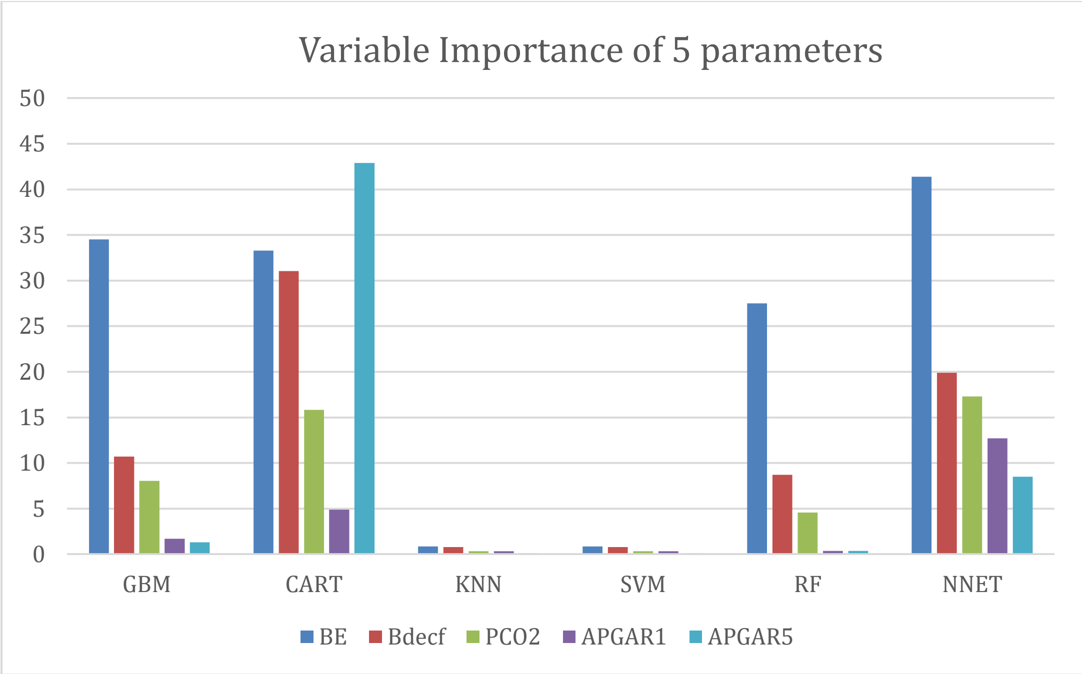

Figure 6.9 variable importance of 5 parameters

In the third set of experiments, only four parameters, namely BE, pCO2, Apgar1, and Apgar5are utilised to test prediction power of ML classifiers.

As shown in Table 6.7, the prediction results of the tree-based classifiers (RF, CART, GBM) and neural network confirm the previous results regarding the BE level in addition to the PCO2 which ranked in the second place compared to the other variables. KNN and SVM failed in defining any of the variables as an important indicator for the diagnosis.

Again APGAR 5 shows a distinguish evaluation by the neural network classifier in determining the hypoxic state which, as we mentioned earlier, can be useful to the physicians when combined with other variables such as pH, BDecf, BE, and PCO2. The simulation results indicated that NNET was the best classifier followed by the CART in detecting foetal hypoxic cases using all the variables regardless of their importance level. RF and GBM detect the cases depending on the BE and PCO2 only which is considered a good clue for the physicians to follow the neonate, as they both reflect the early diagnostic changes in the hypoxic disorder.

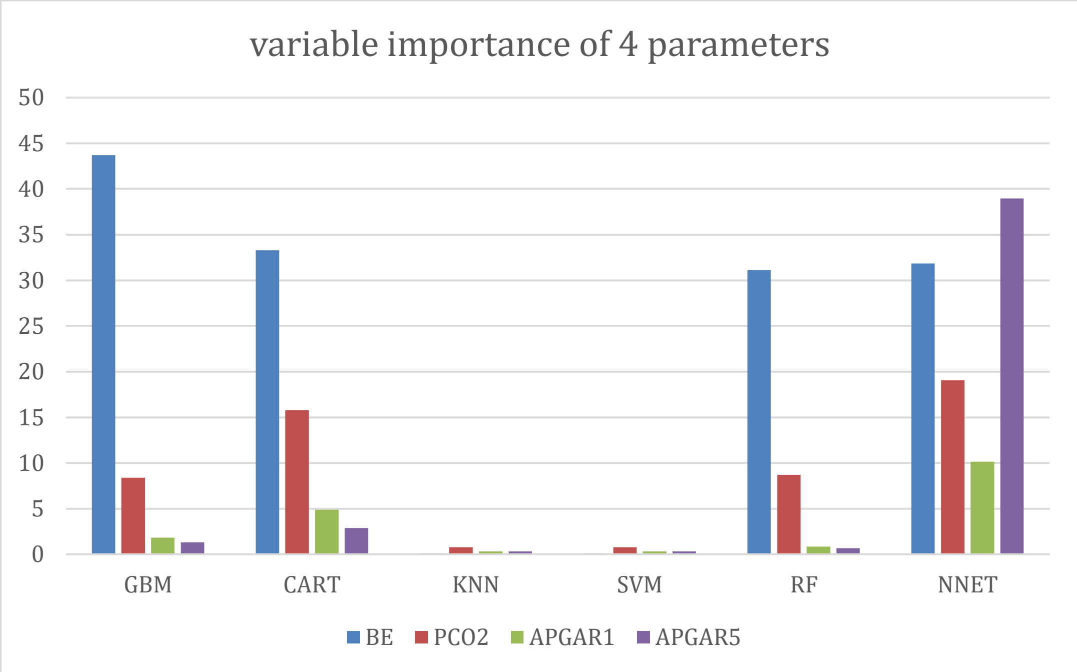

Figure 6.10 variable importance of 4 parameters

Table 6.7 variable importance of all classifiers using 4 parameters

CLASSIFIERS	BE	PCO2	APGAR1	APGAR5
GBM	43.7	8.4	1.85	1.3
CART	33.3	15.8	4.9	2.9
KNN	0.11	0.8	0.32	0.32
SVM	0.11	0.8	0.32	0.32
RF	31.1	8.7	0.86	0.67
NNET	31.84	19.05	10.15	38.96

6.6 Summary of the chapter

This chapter discussed the simulation results of three different rounds of experiment. We tried, in each round, using different datasets for the training and testing different classifiers. This gave us the opportunity to compare the classifiers' performance and detect which one is the best in real medical data analysis. The results show decision tree-based classifiers worked with almost the same performance, compared to each other; however, the RF shows a slightly higher AUC compared to other decision tree-based classifiers (i.e. CART and GBM). The NNET classifier shows a very high level of accuracy in detecting foetal hypoxia. These experiments show valuable achievement of the decision tree-based classifiers and NNET in detecting the most important variables that can be used in the medical detection of foetal hypoxia compared to the KNN and SVM. The later classifiers have failed in both the detection of any reliable variable or any relationship between the parameters to detect foetal hypoxia that could be confirmed medically.

On the other hand, the three experiments gave us the opportunity to detect other variables that might help us in detecting the pathological cases in case of missing pH or the BDecf values, all classifiers, apart from SVM and KNN, show the PCO2 and BE as important features for the classification. SVM and KNN have both failed in detecting any important variables in the third round of the study, when both pH and BDecf were not considered as the most important parameters in classification.

CHAPTER SEVEN: CONCLUSION AND FUTURE WORK

7.1Introduction

In this research, six different ML techniques were presented to classify and study various foetal parameters. The main goal of the study is to find the best parameters for hypoxic state detection. In addition to finding relations between these variables, using ML classifiers, different parameters and research studies that affected the final decision were provided.

The direction of the future work has been identified and illustrated in section 7.2 to find the best and most accurate way to diagnose foetal hypoxia.

7.2 Conclusion

In general, this work proposes the use of intelligent approaches to improve the quality of management and diagnosis provided to physicians regarding the foetal status after delivery. The aim is to improve the accuracy of the diagnostic methods of the foetal hypoxic state via two main ways; the first is to improve the diagnosis or the first detection of the early pathological changes of the foetus using the pH and BDecf levels of different umbilical artery blood profile samples, while the second is to find other combination of parameters that may help in hypoxic status detection using different machine learning classifiers and identify the most accurate classifier/s that can be used in classification problem of real world medical data .

This research is inspired by the urgent need for a new pathway that could reduce the burden on the shoulders of medical staff, and at the same time enhance the quality of the new-born's life as well as their family. The use of machine-learning based diagnostic methods could reduce the need for specialist assessment as they can learn from previously diagnosed new-borns to diagnose new cases. These machine-learning classifiers could also be used to train non-specialist doctors to improve their decision-making procedure.

Foetal hypoxia can be detected using two main parameters, namely pH and BDecf [16], however our goal is to discover whether we can detect foetal hypoxia using other parameters such as BE, PCO2, Apgar1 and Apgar5 using various ML classifiers , as well as identifying the threshold of these values that may cause adverse neurological outcome in foetuses.

To establish intelligent diagnostic methods, an experimental procedure was undertaken using six popular supervised machine-learning classifiers. This stage is usually known as the knowledge acquirement stage, where classifiers learned, identified patterns and gained knowledge from new-borns' records in order to classify new hypoxic cases. Thereafter we have tested the learning of the classifiers and generalisation capabilities using a number of samples that have not been used in the training process through different trials using various parameters in each trial. Using a number of statistical measures, we have assessed the classifiers' sensitivity, specificity and classification accuracy to establish a classifier performance evaluation.

In this context, all the ML classifiers having an acceptable performance results through all trials. All classifiers apart of SVM and KNN improved their classification performance in using different variables by considered the combination of (pH, BE, BDecf, PCO2) is important for the classification in the 1st trial while in the 2nd and 3rd trial, they also detected the AS1, AS5 as an important variables which give good combination between all the variables that can help the physician and other medical staff in the diagnosis of the foetal hypoxia.

CART for instance, shows good classification performance by using all the important variables through all the trials. CART classifier considers a non-linear supervised learning method that is typically used to classify non-linear separable data and can be graphically represented as a binary decision tree. The CART classifier uses the ratio of information gain as a splitting criterion. The best spilt would minimise the impurity of the output data subsets. From the resulting subsets, the splitting process is repeated until a stopping criterion is invoked.

This study also employed two ensemble learning classifiers, i.e. RF and GBM methods. They are considers a collection or ensemble of decision trees (i.e. CART). RF and GBM are taking the concept of CART a step further via the generation of dozens of trees. In contrast to CART, which uses all of the parameters along with the whole training dataset to build a classifier, RF and GBM select an arbitrary sample of the data and determine a particular subset of parameters to build each tree individually. Both CART and RF considered the most accurate classifiers in the first and second trial, while the third trial shows the GBM is the most accurate classifier followed by SVM then RF and CART classifiers.

Away from tree-driven classifiers, this study has also implemented the NNET classifier. NNET showed very good and reliable results with respect to the classification of the foetal hypoxic cases especially in the first and second trials when it was the third most accurate classifier. NNET ranked all the four variables (pH, BDecf, BE, PCO2) as an important variables in detecting the hypoxic cases, although it consider the AS5 is the most important one in the third trial , which may affect the model choice in future works as it consider the last parameter in detecting the hypoxic cases. The output of the NNET classifier could be more difficult to interpret when compared with tree-driven classifiers.

On the other hand, some of the ML classifiers such as KNN and SVM have failed to achieve a reliable classification regarding the hypoxic cases in the first and second experiment. Their classification ability are not reliable, as they are biased toward less important parameters. Although their performance improved in the third trial making the SVM is the most accurate classifier among all the classifiers, but this cannot be dependable as both of them failed in detecting the pH value as an important variable in the first trial. The level of pH increased in both types (respiratory and metabolic acidosis) of hypoxia, which means it is very important for the classification. While in the second and third trial they ranked the metabolic variables (BE, BDecf) as important variables with very low favourability toward the respiratory one (PCO2), which may affect the final medical diagnosis as the hypoxic process includes combination of both respiratory and metabolic changes.

Failure of SVM and KNN in classifying the cases using reliable combination of the variables can be explained according to the way of their working, both of them depending on the distribution of the inputs, as well as the distance between the inputs (i.e. Euclidean distance). Not all the parameters can contribute equally by using the Euclidean distance method. Choosing the wrong kernel for the SVM or any other tuning parameters such as the sigma value or K number for KNN can be consider another factors in the poor performance of these classifiers.

7.3 Future work

Many of the machine learning techniques have been used extensively in foetal hypoxia classification for normal and pathological cases depending on different diagnostic signs such as foetal heart rate abnormalities by the CTG test or by ultrasound or Doppler findings during the pregnancy stages and finally by the cordocentesis that provided the physicians with a full blood sample profile. Many studies worked on diagnosing foetal hypoxia using various approaches such as Warrick et al. [123], who proposed a non-parametric approach that focuses on the dynamic relationship between UC (as an input) and the FHR (as an output). They have used the system identification approach to estimate dynamics in terms of an impulse response function (IRF). The author reports that their approached system can detect almost half of the pathological cases (roughly 1 hour and 40 min before the original time of delivery with acceptable 7.5 % false positive rate.

As shown in our research the decision tress based models as well as the neural network (backpropagation) showed a promising performance in classification of foetal hypoxia, while both SVM and KNN have failed in showing a good and reliable classification. Therefore, we are willing to examine these classifiers with different tuning parameter selection that may affect the classifier performance and compare it to previously gained results from this research.

There are various methods proposed and established to enhance the performance of some classifiers other than changing the tuning parameters of each one. For instance using the different types of variables for the classification such as the signal records of the CTG. Artificial Neural Networks (ANN) have been used effectively in many researches for analysing different types of signal data such as the uterine electromyography (EMG) to classify term and preterm delivery records as in [157-159].

Open access data by Physionet that we used to exclude the clinical data have also signal records including the foetal heart rate and the maternal uterine contraction in addition to the clinical data that we used for this research. Trying other techniques of neural network for studying the foetal heart rate (FHR) to detect the foetal hypoxia and looking at the T/QRS ratio of the foetal ECG during the labour was another challenge for the researchers [160]. Therefore, for future work, we intend firstly, to examine the same six classifiers with different parameters and study diagnosis of foetal hypoxia using both signal and clinical data and finding other possible relationships between the foetal umbilical blood profile variables with the foetal heart rate and uterine contraction to improve the early detection of the foetal hypoxia before labour by detecting the hypoxic state in the antepartum stage of the pregnancy not only in the intrapartum stage.

Although, some studies have found a correlation between T/QRS ratio and umbilical artery (UA) pH (as in MaeLachlan et al study 1992 [162], Others have failed to establish any significant relationship [161]. Even in the study of MaeLachlan et al which did show a correlation between T/QRS ratio, umbilical artery pH level, the sensitivity for detecting acidaemia was only 29 %, compared to 76% for a pathological CTG [162].

Therefore, to overcome this limitation, our future study intends to examine different types of classifiers as a useful tool for categorization of multivariate data [163] and analyse the individual foetal CTG signal data in addition to the clinical umbilical blood profile data.

Establishing different parameters (FHR, uterine contraction) or ratios and relating these results to the acid-base status of the foetus is considered a vital method in detecting hypoxia and preventing further neurological and physical foetal damage.

References

[1] V. S. Talaulikar and S. Arulkumaran, "Maternal, perinatal and long-term outcomes after assisted reproductive techniques (ART): implications for clinical practice," *Eur J Obstet Gynecol Reprod Biol,* vol. 170, pp. 13-9, Sep 2013.

[2] C. S. Bobrow and P. W. Soothill, "Causes and consequences of fetal acidosis," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 80, pp. F246-F249, 1999.

[3] M. A. Hasan, M. Reaz, M. Ibrahimy, M. Hussain, and J. Uddin, "Detection and processing techniques of FECG signal for fetal monitoring," *Biological procedures online,* vol. 11, p. 263, 2009.

[4] E. Wiberg-Itzel, C. Lipponer, M. Norman, A. Herbst, D. Prebensen, A. Hansson*, et al.*, "Determination of pH or lactate in fetal scalp blood in management of intrapartum fetal distress: randomised controlled multicentre trial," *bmj,* vol. 336, pp. 1284-1287, 2008.

[5] A. M. Vintzileos, D. J. Nochimson, A. Antsaklis, I. Varvarigos, E. R. Guzman, and R. A. Knuppel, "Comparison of intrapartum electronic fetal heart rate monitoring versus intermittent auscultation in detecting fetal acidemia at birth," *American journal of obstetrics and gynecology,* vol. 173, pp. 1021-1024, 1995.

[6] I. Ingemarsson, E. Ingemarsson, and J. A. Spencer, *Fetal heart rate monitoring: a practical guide*: Oxford University Press, USA, 1993.

[7] J. Bretscher and E. Saling, "pH values in the human fetus during labor," *American journal of obstetrics and gynecology,* vol. 97, pp. 906-911, 1967.

[8] D. Tuffnell, W. L. Haw, and K. Wilkinson, "How long does a fetal scalp blood sample take?," *BJOG,* vol. 113, pp. 332-4, Mar 2006.

[9] K. G. GOLDABER, L. C. GILSTRAP III, K. J. LEVENO, J. S. DAX, and D. D. McINTIRE, "Pathologic fetal acidemia," *Obstetrics & Gynecology,* vol. 78, pp. 1103-1107, 1991.

[10] M. Westgren, K. Kruger, S. Ek, C. Grunevald, M. Kublickas, K. Naka*, et al.*, "Lactate compared with pH analysis at fetal scalp blood sampling: a prospective randomised study," *BJOG: an international journal of obstetrics & gynaecology,* vol. 105, pp. 29-33, 1998.

[11] L. S. James, I. Weisbrot, C. Prince, D. Holaday, and V. Apgar, "The acid-base status of human infants in relation to birth asphyxia and the onset of respiration," *The Journal of pediatrics,* vol. 52, pp. 379-394, 1958.

[12] C. V. Ananth and A. M. Vintzileos, "Epidemiology of preterm birth and its clinical subtypes," *J Matern Fetal Neonatal Med,* vol. 19, pp. 773-82, Dec 2006.

[13] J. A. Low, C. Panagiotopoulos, and E. J. Derrick, "Newborn complications after intrapartum asphyxia with metabolic acidosis in the preterm fetus," *American journal of obstetrics and gynecology,* vol. 172, pp. 805-810, 1995.

[14] H. S. Dweck, W. Huggins, L. P. Dorman, S. A. Saxon, J. W. Benton, and G. Cassady, "Developmental sequelae in infants having suffered severe perinatal asphyxia," *American journal of obstetrics and gynecology,* vol. 119, pp. 811-815, 1974.

[15] M. K. Lal. (2014). *Pediatric Respiratory Acidosis*. Available: https://emedicine.medscape.com/article/906545-overview

[16] A. C. o. O. Practice, "ACOG Committee Opinion No. 348, November 2006: Umbilical cord blood gas and acid-base analysis," *Obstetrics and gynecology,* vol. 108, p. 1319, 2006.

[17] R. Liston, D. Sawchuck, and D. Young, "Intrapartum Fetal Surveillance (Chapter 2) in: SOGC Clinical Practice Guideline No 197," *J Obstet Gynecol Canada,* vol. 29, pp. S25-S44, 2007.
[18] J. Thomas and S. Paranjothy, "Royal College of Obstetricians and Gynaecologists Clinical Effectiveness Support Unit. National Sentinel Caesarean Section Audit Report," ed: London: RCOG press, 2001.
[19] D. W. Rurak, *Fetal oxygenation, carbon dioxide homeostasis, and acid-base balance*: New York: Oxford University Press, 1994.
[20] A. Huch, R. Huch, H. Schneider, and G. Rooth, "Continuous transcutaneous monitoring of fetal oxygen tension during labour," *British journal of obstetrics and gynaecology,* vol. 84, p. 1, 1977.
[21] D. Singer, "Neonatal tolerance to hypoxia: a comparative-physiological approach," *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology,* vol. 123, pp. 221-234, 1999.
[22] K. M. Godfrey and D. J. Barker, "Fetal nutrition and adult disease," *The American journal of clinical nutrition,* vol. 71, pp. 1344s-1352s, 2000.
[23] D. Ayres-de-Campos, "Acute Fetal Hypoxia/Acidosis," in *Obstetric Emergencies*, ed: Springer, 2017, pp. 7-25.
[24] M. B. Holm, "The fetal response to acute perinatal hypoxia and the pathophysiological mechanisms behind hypoxic-ischemic encephalopathy," 2012.
[25] R. S. Pattar, A. Raj, and B. C. Yelamali, "Incidence of multiorgan dysfunction in perinatal asphyxia," *International Journal of Contemporary Pediatrics,* vol. 2, pp. 428-432, 2017.
[26] J. Parer and E. Livingston, "What is fetal distress?," *International Journal of Gynecology & Obstetrics,* vol. 34, pp. 389-389, 1991.
[27] F. A. Manning, L. D. Platt, and L. Sipos, "Antepartum fetal evaluation: development of a fetal biophysical profile," *American journal of obstetrics and gynecology,* vol. 136, pp. 787-795, 1980.
[28] S. B. Thacker and R. L. Berkelman, "Assessing the diagnostic accuracy and efficacy of selected antepartum fetal surveillance techniques," *Obstetrical & gynecological survey,* vol. 41, pp. 121-141, 1986.
[29] K. Marcdante and R. M. Kliegman, *Nelson Essentials of Pediatrics E-Book*: Elsevier Health Sciences, 2014.
[30] T. K. Butt, R. Farooqui, and M. Khan, "Risk factors for hypoxic ischemic encephalopathy in children," *J Coll Physicians Surg Pak,* vol. 18, pp. 428-432, 2008.
[31] V. Chudacek, J. Spilka, P. Janku, M. Koucky, L. Lhotska, and M. Huptych, "Automatic evaluation of intrapartum fetal heart rate recordings: a comprehensive analysis of useful features," *Physiol Meas,* vol. 32, pp. 1347-60, Aug 2011.
[32] M. V. S. Talaulikar and S. Arulkumaran, "Persistent Challenge of Intrapartum Fetal Heart Rate Monitoring," *Dasgupta's Recent Advances in Obstetrics and Gynecology,* vol. 9, p. 68, 2012.
[33] J. Spilka, "Complex approach to fetal heart rate analysis: A hierarchical classification model," *Czech Technical University, Faculty of Electrical Engineering, Prague,* pp. 35-47, 2013.
[34] N. Jukelevics, *Understanding the dangers of cesarean birth: Making informed decisions*: ABC-CLIO, 2008.
[35] G. S. of Gynecology, "S1-Guideline on the Use of CTG During Pregnancy and Labor: Long version–AWMF Registry No. 015/036," *Geburtshilfe und Frauenheilkunde,* vol. 74, p. 721, 2014.

[36] E. Draper, J. Kurinczuk, C. Lamming, M. Clarke, D. James, and D. Field, "A confidential enquiry into cases of neonatal encephalopathy," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 87, pp. F176-F180, 2002.
[37] R. K. Freeman, T. J. Garite, M. P. Nageotte, and L. A. Miller, *Fetal heart rate monitoring*: Lippincott Williams & Wilkins, 2012.
[38] S. Campbell, D. Griffin, J. Pearce, J. Diaz-Recasens, T. Cohen-Overbeek, K. Willson*, et al.*, "New Doppler technique for assessing uteroplacental blood flow," *The lancet,* vol. 321, pp. 675-677, 1983.
[39] A. Kurjak and F. A. Chervenak, *Textbook of perinatal medicine*: CRC Press, 2006.
[40] K. Okamura, T. Watanabe, S. Tanigawara, Y. Shintaku, H. Endo, M. Iwamoto*, et al.*, "Biochemical evaluation of fetus with hypoxia caused by severe preeclampsia using cordocentesis," *Journal of Perinatal Medicine-Official Journal of the WAPM,* vol. 18, pp. 441-447, 1990.
[41] J. A. Thorp, G. A. Dildy, E. R. Yeomans, B. A. Meyer, and V. M. Parisi, "Umbilical cord blood gas analysis at delivery," *American journal of obstetrics and gynecology,* vol. 175, pp. 517-522, 1996.
[42] L. C. Gilstrap, K. J. Leveno, J. Burris, M. L. Williams, and B. B. Little, "Diagnosis of birth asphyxia on the basis of fetal pH, Apgar score, and newborn cerebral dysfunction," *American journal of obstetrics and gynecology,* vol. 161, pp. 825-830, 1989.
[43] A. Georgieva, S. J. Payne, M. Moulden, and C. W. G. Redman, "Artificial neural networks applied to fetal monitoring in labour," *Neural Computing and Applications,* vol. 22, pp. 85-93, 2011.
[44] K. G. Rosén, S. Blad, D. Larsson, H. Norén, and N. Outram, "Assessment of the fetal bioprofile during labor by fetal ECG analysis," *Expert Review of Obstetrics & Gynecology,* vol. 2, pp. 609-620, 2007.
[45] H. Abdel-Aleem, A. Amin, M. Shokry, and R. Radwan, "Therapeutic amnioinfusion for intrapartum fetal distress using a pediatric feeding tube," *International Journal of Gynecology & Obstetrics,* vol. 90, pp. 94-98, 2005.
[46] R. Schapire, "Cos 511: Theoretical machine learning: Lecture 1," ed.
[47] J. G. Carbonell, "Learning by analogy: Formulating and generalizing plans from past experience," in *Machine learning*, ed: Springer, 1983, pp. 137-161.
[48] E. Alpaydin, *Introduction to machine learning*: MIT press, 2014.
[49] J. brownlee. (2014). *Discover Feature Engineering, How to Engineer Features and How to Get Good at It*. Available: https://machinelearningmastery.com/discover-feature-engineering-how-to-engineer-features-and-how-to-get-good-at-it/ [Accessed: 17.01.2017]
[50] W. I. D. Mining, "Data Mining: Concepts and Techniques," *Morgan Kaufinann,* 2006.
[51] M. Kantardzic, *Data mining: concepts, models, methods, and algorithms*: John Wiley & Sons, 2011.
[52] T. O. Ayodele, "Types of machine learning algorithms," in *New advances in machine learning*, ed: InTech, 2010.
[53] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*: MIT press, 2012.
[54] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning,* vol. 3, pp. 1-130, 2009.
[55] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks,* vol. 20, pp. 542-542, 2009.
[56] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* vol. 1: MIT press Cambridge, 1998.
[57] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning* vol. 135: MIT Press Cambridge, 1998.

[58] I. Aleksander and H. Morton, *An introduction to neural computing* vol. 3: Chapman & Hall London, 1990.
[59] R. Ho. (2012). *An Introduction to 6 Machine Learning Models* Available: https://dzone.com/articles/introduction-6-machine [Accessed: 12.01.2017]
[60] A. Mahdi, "The Application of Neural Network inFinancial Time Series Analysis and Prediction Using Immune System," *Liverpool John Moores University,* 2010.
[61] S. Haykin, "Neural networks a comprehensive introduction," ed: Prentice Hall, New Jersey, 1999.
[62] Y. Tsukamoto and A. Namatame, "Evolving neural network models," in *Evolutionary Computation, 1996., Proceedings of IEEE International Conference on*, 1996, pp. 689-693.
[63] A. Brabazon and M. O'Neill, *Biologically inspired algorithms for financial modelling*: Springer Science & Business Media, 2006.
[64] A. Knowles, "Higher-Order and Pipelined Neural Networks for Time Series Prediction of Currency Exchange Rates," *A Master Thesis, Liverpool John Moores University,* 2005.
[65] R. A. ogly Aliev and R. R. Aliev, *Soft computing and its applications*: World Scientific, 2001.
[66] L. Fausett, *Fundamentals of neural networks: architectures, algorithms, and applications*: Prentice-Hall, Inc., 1994.
[67] K. Richmond, "Estimating articulatory parameters from the acoustic speech signal," 2002.
[68] S. Lawrence and C. L. Giles, "Overfitting and neural networks: conjugate gradient and backpropagation," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, 2000, pp. 114-119.
[69] C. Dunis and M. Williams, "Modelling and trading the EUR/USD exchange rate: Do neural network models perform better?," *Derivatives use, trading and regulation,* vol. 8, pp. 211-239, 2002.
[70] J. MacIntyre, "Applications of neural computing in the twenty-first century and 21 years of Neural Computing & Applications," *Neural Computing and Applications,* vol. 23, pp. 657-665, 2013.
[71] P. J. Werbos, *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting* vol. 1: John Wiley & Sons, 1994.
[72] P. J. Werbos, "Backpropagation: Basics and new developments," in *The handbook of brain theory and neural networks*, 1998, pp. 134-139.
[73] S. Haykin and R. Lippmann, "Neural networks, a comprehensive foundation," *International journal of neural systems,* vol. 5, pp. 363-364, 1994.
[74] J. Fulcher, "Computational intelligence: an introduction," in *Computational intelligence: a compendium*, ed: Springer, 2008, pp. 3-78.
[75] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, "Artificial neural networks in medical diagnosis," ed: Elsevier, 2013.
[76] A. J. Maren, C. T. Harston, and R. M. Pap, *Handbook of neural computing applications*: Academic Press, 2014.
[77] A. Wu, S. Wen, and Z. Zeng, "Synchronization control of a class of memristor-based recurrent neural networks," *Information Sciences,* vol. 183, pp. 106-116, 2012.
[78] N. F. Güler, E. D. Übeyli, and I. Güler, "Recurrent neural networks employing Lyapunov exponents for EEG signals classification," *Expert systems with applications,* vol. 29, pp. 506-514, 2005.
[79] A. Petrosian, D. Prokhorov, W. Lajara-Nanson, and R. Schiffer, "Recurrent neural network-based approach for early recognition of Alzheimer's disease in EEG," *Clinical Neurophysiology,* vol. 112, pp. 1378-1387, 2001.

[80] A. Petrosian, D. Prokhorov, R. Homan, R. Dasheiff, and D. Wunsch II, "Recurrent neural network based prediction of epileptic seizures in intra-and extracranial EEG," *Neurocomputing,* vol. 30, pp. 201-218, 2000.
[81] F. Visin, K. Kastner, K. Cho, M. Matteucci, A. Courville, and Y. Bengio, "Renet: A recurrent neural network based alternative to convolutional networks," *arXiv preprint arXiv:1505.00393,* 2015.
[82] E. D. Übeyli, "Analysis of EEG signals by implementing eigenvector methods/recurrent neural networks," *Digital Signal Processing,* vol. 19, pp. 134-143, 2009.
[83] M. Hüsken and P. Stagge, "Recurrent neural networks for time series classification," *Neurocomputing,* vol. 50, pp. 223-235, 2003.
[84] J. R. Chung, J. Kwon, and Y. Choe, "Evolution of recollection and prediction in neural networks," in *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, 2009, pp. 571-577.
[85] E. Turban, R. Sharda, and D. Delen, *Decision support and business intelligence systems*: Pearson Education India, 2011.
[86] L. Gonzalez, C. Angulo, F. Velasco, and A. Catala, "Unified dual for bi-class SVM approaches," *Pattern Recognition,* vol. 38, pp. 1772-1774, 2005.
[87] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.
[88] V. Vapnik, *The nature of statistical learning theory*: Springer science & business media, 2013.
[89] E. Osuna, R. Freund, and F. Girosit, "Training support vector machines: an application to face detection," in *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on*, 1997, pp. 130-136.
[90] M. M. Adankon and M. Cheriet, "Model selection for the LS-SVM. Application to handwriting recognition," *Pattern Recognition,* vol. 42, pp. 3264-3270, 2009.
[91] P. H. Chen, C. J. Lin, and B. Schölkopf, "A tutorial on ν-support vector machines," *Applied Stochastic Models in Business and Industry,* vol. 21, pp. 111-136, 2005.
[92] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques," ed, 2007.
[93] D. Han, L. Chan, and N. Zhu, "Flood forecasting using support vector machines," *Journal of hydroinformatics,* vol. 9, pp. 267-276, 2007.
[94] B. V. Ramana, M. S. P. Babu, and N. Venkateswarlu, "A critical study of selected classification algorithms for liver disease diagnosis," *International Journal of Database Management Systems,* vol. 3, pp. 101-114, 2011.
[95] M. Khalaf, A. J. Hussain, R. Keight, D. Al-Jumeily, R. Keenan, P. Fergus*, et al.*, "The utilisation of composite machine learning models for the classification of medical datasets for sickle cell disease," in *Digital Information Processing and Communications (ICDIPC), 2016 Sixth International Conference on*, 2016, pp. 37-41.
[96] R. T. Ionescu and M. Popescu, "Knowledge transfer between computer vision and text mining," *Advances in Computer Vision and Pattern Recognition. Springer International Publishing,* 2016.
[97] J. Brownlee. (2016, 17/01). *K-Nearest Neighbors for Machine Learning*. Available: https://machinelearningmastery.com/k-nearest-neighbors-for-machine-learning/ [Accessed by 17.01.2017]
[98] P. A. Jaskowiak and R. Campello, "Comparing correlation coefficients as dissimilarity measures for cancer classification in gene expression data," in *Proceedings of the Brazilian Symposium on Bioinformatics*, 2011, pp. 1-8.
[99] P.-N. Tan, M. Steinbach, and V. Kumar, "Classification: basic concepts, decision trees, and model evaluation," *Introduction to data mining,* vol. 1, pp. 145-205, 2006.

[100] L. Rokach and O. Maimon, "Decision trees," in *Data mining and knowledge discovery handbook*, ed: Springer, 2005, pp. 165-192.
[101] Y.-Y. Song and L. Ying, "Decision tree methods: applications for classification and prediction," *Shanghai archives of psychiatry,* vol. 27, p. 130, 2015.
[102] L. Gordon, "Using classification and regression trees (CART) in SAS® enterprise miner TM for applications in public health," in *SAS Global Forum*, 2013, pp. 1-8.
[103] O. Pauly, B. Glocker, A. Criminisi, D. Mateus, A. M. Möller, S. Nekolla*, et al.*, "Fast multiple organ detection and localization in whole-body MR Dixon sequences," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2011, pp. 239-247.
[104] L. Breiman, "Random forests," *Machine learning,* vol. 45, pp. 5-32, 2001.
[105] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma*, et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, 2017, pp. 3149-3157.
[106] S. Hakan and A. Subasi, "Classification of Fetal state from the Cardiotocogram Recordings using ANN and Simple Logistic," 2012.
[107] A. MacLennan, "A template for defining a causal relationship between acute intrapartum events and cerebral palsy: international consensus statement. International Cerebral Palsy Task Force," *The Australian & New Zealand journal of obstetrics & gynaecology,* vol. 40, p. 13, 2000.
[108] G. Heller, R. Schnell, B. Misselwitz, and S. Schmidt, "Umbilical blood pH, Apgar scores, and early neonatal mortality," *Zeitschrift fur Geburtshilfe und Neonatologie,* vol. 207, pp. 84-89, 2003.
[109] M. Dijxhoorn, G. Visser, H. Huisjes, V. Fidler, and B. Touwen, "The relation between umbilical pH values and neonatal neurological morbidity in full term appropriate-for-dates infants," *Early human development,* vol. 11, pp. 33-42, 1985.
[110] H. Huisjes and J. Aarnoudse, "Arterial or venous umbilical pH as a measure of neonatal morbidity?," *Early human development,* vol. 3, pp. 155-161, 1979.
[111] P. P. van den Berg, W. L. Nelen, H. W. Jongsma, R. Nijland, L. A. Kollée, J. G. Nijhuis*, et al.*, "Neonatal complications in newborns with an umbilical artery pH< 7.00," *American journal of obstetrics and gynecology,* vol. 175, pp. 1152-1157, 1996.
[112] J. Wildschut, F. J. Feron, J. G. Hendriksen, M. van Hall, D. W. Gavilanes-Jiminez, M. Hadders-Algra*, et al.*, "Acid–base status at birth, spontaneous motor behaviour at term and 3 months and neurodevelopmental outcome at age 4 years in full-term infants," *Early human development,* vol. 81, pp. 535-544, 2005.
[113] J. Fahey and T. L. King, "Intrauterine asphyxia: clinical implications for providers of intrapartum care," *Journal of Midwifery & Women's Health,* vol. 50, pp. 498-506, 2005.
[114] G. L. Malin, R. K. Morris, and K. S. Khan, "Strength of association between umbilical cord pH and perinatal and long term outcomes: systematic review and meta-analysis," *Bmj,* vol. 340, p. c1471, 2010.
[115] G. Sykes, P. M. Molloy, P. Johnson, G. Stirrat, and A. Turnbull, "Fetal distress and the condition of newborn infants," *Br Med J (Clin Res Ed),* vol. 287, pp. 943-945, 1983.
[116] P. Steer, "Fetal monitoring—Present and future," *European Journal of Obstetrics & Gynecology,* vol. 24, pp. 112-117, 1987.
[117] P. BERG, S. Schmidt, J. Gesche, and E. Saling, "Fetal distress and the condition of the newborn using cardiotocography and fetal blood analysis during labour," *BJOG: An International Journal of Obstetrics & Gynaecology,* vol. 94, pp. 72-75, 1987.
[118] P. Yeh, K. Emary, and L. Impey, "The relationship between umbilical cord arterial pH and serious adverse neonatal outcome: analysis of 51 519

consecutive validated samples," *BJOG: An International Journal of Obstetrics & Gynaecology,* vol. 119, pp. 824-831, 2012.
[119] B. K. Strachan, D. S. Sahota, W. J. Wijngaarden, D. K. James, and A. M. Chang, "Computerised analysis of the fetal heart rate and relation to acidaemia at delivery," *BJOG: An International Journal of Obstetrics & Gynaecology,* vol. 108, pp. 848-852, 2001.
[120] M. Jezewski, R. Czabanski, J. Wróbel, and K. Horoba, "Analysis of extracted cardiotocographic signal features to improve automated prediction of fetal outcome," *Biocybernetics and Biomedical Engineering,* vol. 30, pp. 29-47, 2010.
[121] R. D. Keith, S. Beckley, J. M. Garibaldi, J. A. Westgate, E. C. Ifeachor, and K. R. Greene, "A multicentre comparative study of 17 experts and an intelligent computer system for managing labour using the cardiotocogram," *BJOG: an international journal of obstetrics & gynaecology,* vol. 102, pp. 688-700, 1995.
[122] G. Magenes, M. G. Signorini, and D. Arduini, "Classification of cardiotocographic records by neural networks," in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, 2000, pp. 637-641.
[123] P. A. Warrick, E. F. Hamilton, D. Precup, and R. E. Kearney, "Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography," *IEEE Transactions on Biomedical Engineering,* vol. 57, pp. 771-779, 2010.
[124] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark*, et al.*, "Physiobank, physiotoolkit, and physionet components of a new research resource for complex physiologic signals," *Circulation,* vol. 101, pp. e215-e220, 2000.
[125] C. X. Ling and V. S. Sheng, "Cost-sensitive learning," in *Encyclopedia of machine learning*, ed: Springer, 2011, pp. 231-235.
[126] J. Brownlee, "Machine learning mastery," *URL: http://machinelearningmastery. com/discover-feature-engineering-howtoengineer-features-and-how-to-getgood-at-it,* 2014.
[127] L. v. d. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research,* vol. 9, pp. 2579-2605, 2008.
[128] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training sets: one-sided selection," in *ICML*, 1997, pp. 179-186.
[129] F. Provost, "Machine learning from imbalanced data sets 101," in *Proceedings of the AAAI'2000 workshop on imbalanced data sets*, 2000, pp. 1-3.
[130] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Proceedings of the eleventh international conference on machine learning*, 1994, pp. 148-156.
[131] A. Estabrooks, T. Jo, and N. Japkowicz, "A multiple resampling method for learning from imbalanced data sets," *Computational intelligence,* vol. 20, pp. 18-36, 2004.
[132] L. K. Callaway, K. Lust, and H. D. McIntyre, "Pregnancy outcomes in women of very advanced maternal age," *Australian and New Zealand journal of obstetrics and gynaecology,* vol. 45, pp. 12-16, 2005.
[133] S. Berglund, C. Grunewald, H. Pettersson, and S. Cnattingius, "Risk factors for asphyxia associated with substandard care during labor," *Acta obstetricia et gynecologica Scandinavica,* vol. 89, pp. 39-48, 2010.
[134] J. Drage and H. Berendes, "Apgar scores and outcome of the newborn," *Pediatric Clinics of North America,* vol. 13, pp. 635-643, 1966.
[135] Z. Henderson and J. L. Ecker, "Fetal scalp blood sampling—limited role in contemporary obstetric practice: Part II," *Laboratory Medicine,* vol. 34, pp. 594-600, 2015.

[136] C. L. Winkler, J. C. Hauth, J. M. Tucker, J. Owen, and C. G. Brumfield, "Neonatal complications at term as related to the degree of umbilical artery acidemia," *American journal of obstetrics and gynecology,* vol. 164, pp. 637-641, 1991.
[137] J. M. Perlman and R. Risser, "Can asphyxiated infants at risk for neonatal seizures be rapidly identified by current high-risk markers?," *Pediatrics,* vol. 97, pp. 456-462, 1996.
[138] R. J. Portman, B. S. Carter, M. S. Gaylord, M. G. Murphy, R. E. Thieme, and G. B. Merenstein, "Predicting neonatal morbidity after perinatal asphyxia: a scoring system," *American journal of obstetrics and gynecology,* vol. 162, pp. 174-182, 1990.
[139] B. S. Carter, F. McNabb, and G. B. Merenstein, "Prospective validation of a scoring system for predicting neonatal morbidity after acute perinatal asphyxia," *The Journal of pediatrics,* vol. 132, pp. 619-623, 1998.
[140] K. P. Williams and A. Singh, "The correlation of seizures in newborn infants with significant acidosis at birth with umbilical artery cord gas values," *Obstetrics & Gynecology,* vol. 100, pp. 557-560, 2002.
[141] T. M. Goodwin, I. Belai, P. Hernandez, M. Durand, and R. H. Paul, "Asphyxial complications in the term newborn with severe umbilical acidemia," *American journal of obstetrics and gynecology,* vol. 167, pp. 1506-1512, 1992.
[142] J. A. Low, R. Galbraith, D. Muir, H. Killen, E. Pater, and E. Karchmar, "Factors associated with motor and cognitive deficits in children after intrapartum fetal hypoxia," *American journal of obstetrics and gynecology,* vol. 148, pp. 533-539, 1984.
[143] J. A. Low, R. S. Galbraith, D. W. Muir, H. L. Killen, E. A. Pater, and E. J. Karchmar, "Motor and cognitive deficits after intrapartum asphyxia in the mature fetus," *American journal of obstetrics and gynecology,* vol. 158, pp. 356-361, 1988.
[144] H. M. Sehdev, D. M. Stamilio, G. A. Macones, E. Graham, and M. A. Morgan, "Predictive factors for neonatal morbidity in neonates with an umbilical arterial cord pH less than 7.00," *American journal of obstetrics and gynecology,* vol. 177, pp. 1030-1034, 1997.
[145] B. M. Casey, K. G. Goldaber, D. D. McIntire, and K. J. Leveno, "Outcomes among term infants when two-hour postnatal pH is compared with pH at delivery," *American journal of obstetrics and gynecology,* vol. 184, pp. 447-450, 2001.
[146] D. M. Murray, G. B. Boylan, A. P. Fitzgerald, C. A. Ryan, B. P. Murphy, and S. Connolly, "Persistent lactic acidosis in neonatal hypoxic–ischaemic encephalopathy correlates with EEG grade and electrographic seizure burden," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 93, pp. F183-F186, 2008.
[147] L. Armstrong and B. Stenson, "Use of umbilical cord blood gas analysis in the assessment of the newborn," *Archives of Disease in Childhood-Fetal and Neonatal Edition,* vol. 92, pp. F430-F434, 2007.
[148] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R news,* vol. 2, pp. 18-22, 2002.
[149] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*: Elsevier, 2011.
[150] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition,* vol. 30, pp. 1145-1159, 1997.
[151] A. J. Aljaaf, D. Al-Jumeily, A. J. Hussain, P. Fergus, M. Al-Jumaily, and N. Radi, "Applied machine learning classifiers for medical applications: clarifying the behavioural patterns using a variety of datasets," in *Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on*, 2015, pp. 228-232.
[152] A. Aljaaf, D. Al-Jumeily, A. Hussain, T. Dawson, P. Fergus, and M. Al-Jumaily, "Predicting the likelihood of heart failure with a multi level risk assessment using decision tree," in *Technological Advances in Electrical, Electronics and*

Computer Engineering (TAEECE), 2015 Third International Conference on, 2015, pp. 101-106.

[153] T. R. Hoens and N. V. Chawla, "Imbalanced datasets: from sampling to classifiers," *Imbalanced Learning: Foundations, Algorithms, and Applications,* pp. 43-59, 2013.

[154] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*: Morgan Kaufmann, 2016.

[155] P. Cook, "The influence on foetal outcome of maternal carbon dioxide tension at caesarean section under general anaesthesia," *Anaesthesia and intensive care,* vol. 12, pp. 296-302, 1984.

[156] A. Peng, L. Blancato, and E. Motoyama, "EFFECT OF MATERNAL HYPOCAPNIA V. EUCAPNIA ON THE FOETUS DURING CAESAREAN SECTION," *Survey of Anesthesiology,* vol. 17, p. 366, 1973.

[157] W. L. Maner and R. E. Garfield, "Identification of human term and preterm labor using artificial neural networks on uterine electromyography data," *Annals of biomedical engineering,* vol. 35, pp. 465-473, 2007.

[158] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology,* vol. 278, pp. H2039-H2049, 2000.

[159] I. O. Idowu, P. Fergus, A. Hussain, C. Dobbins, and H. Al Askar, "Advance artificial neural network classification techniques using EHG for detecting preterm births," in *Complex, Intelligent and Software Intensive Systems (CISIS), 2014 Eighth International Conference on*, 2014, pp. 95-100.

[160] A. Stock, M. Rogers, A. Li, and A. Chang, "1 Use of the neural network for hypothesis generation in fetal surveillance," *Baillière's clinical obstetrics and gynaecology,* vol. 8, pp. 533-548, 1994.

[161] S. NEWBOLD, T. WHEELER, and F. CLEWLOW, "Comparison of the T/QRS ratio of the fetal electrocardiogram and the fetal heart rate during labour and the relation of these variables to condition at delivery," *BJOG: An International Journal of Obstetrics & Gynaecology,* vol. 98, pp. 173-178, 1991.

[162] N. A. MACLACHLAN, J. A. SPENCER, K. HARDING, and S. Arulkumaran, "Fetal acidaemia, the cardiotocograph and the T/QRS ratio of the fetal ECG in labour," *BJOG: An International Journal of Obstetrics & Gynaecology,* vol. 99, pp. 26-31, 1992.

[163] J. J. Liszka-Hackzell, "Categorization of fetal heart rate patterns using neural networks," *Journal of medical systems,* vol. 25, pp. 269-276, 2001.