Classification Techniques Using EHG Signals for Detecting Preterm Births

IBRAHIM OLATUNJI IDOWU

BSc, MSc, AHEA

A thesis submitted in partial fulfilment of the

requirements of Liverpool John Moores University

for the degree of Doctor of Philosophy

May 2017

ABSTRACT

Premature birth is defined as an infant born before 37 weeks of gestation and can be sub-categorized into three phrases; late preterm delivery between 34 and 36 weeks of gestation; moderately preterm between 32 and 34 weeks, and extreme preterm less than 28 weeks of gestation.

Globally, the rate of preterm births is increasing, thus resulting in significant health, development and economic problems. The current methods for the detection of preterm birth are inadequate due to the fact that the exact cause of premature uterine contractions leading to delivery is mostly unknown. Another problem is the interpretation of temporal and spectral characteristics of Electromyography (EMG), which is an electrodiagnostic medicine technique for recording and evaluating the electrical activity produced by uterine muscles during pregnancy and parturition – significant variability exists among obstetric care practitioners.

Apart from a small number of potential causes for preterm birth, such as medication, uterine over-distension, preterm premature rupture of membranes (PPROM), intrauterine inflammation, precocious foetal endocrine activation, surgery, ethnicity and lifestyle, there is still a large amount of uncertainty about their specific risks. Hence, it is currently very difficult to make reliable predictions about preterm delivery risk. There has also been some evidence that the analysis of uterine electrical signals, collected from the abdominal surface, could provide an independent and easier way to diagnose true labour and detect the onset of preterm delivery. Early detection opens up new avenues for the development of an automated ambulatory system, based on uterine EMG, for patient monitoring during pregnancy.

This can be made possible through the use of machine learning. The essence of machine learning is the utilisation of previously recorded data outcomes to train algorithms to stimulate software learning elements. Such learned models can, as a result, be used to detect and predict the early signs associated with the onset of preterm birth.

Therefore in this thesis, Electrohysterography signals are used to classify uterine activity associated with preterm birth. This is achieved using an open dataset, which contains 262 records for women who delivered at term and 38 who delivered prematurely. Several new features from Electromyography studies are utilized, as well as feature-ranking techniques to determine their discriminative capabilities in detecting term and preterm records. The results illustrate that the combination of the Levenberg-Marquardt trained Feed-Forward Neural Network, Radial Basis Function Neural Network and the Random Neural Network classifiers performed the best, with 91% for sensitivity, 84% for specificity, 94% for the area under the curve and 12% for the mean error rate. Applying advanced machine learning algorithms, in conjunction with innovative signal processing techniques and the analysis of Electrohysterography signals shows significant benefits for use in clinical interventions for preterm birth assessments.

Word of Inspiration

"In the name of God, the Merciful, Passionate, Beneficent and Magnificent."

"My thoughts about life aren't just about life itself, but life is about anything that related and surrounded by it. I said many times be not afraid of your shadow Son, but walk along the path that leads to success. Life is not about who you are but is always about what you are capable of giving to, yourself and the people around you. I always said be optimistic, visionary, focus and believe in yourself.

When I look back today as a father, my past, present and my future, nothing is greater than the defined greater Mighty God. My past is gone I am greater than, my present is here today, catching up with, and my future is unknown greater than me whilst still working hard to obtain success in life to make better future for my family.

Son I knew you have a dream, and your imagination keep going far, far from your thinking, your abilities, skills, passions power of knowledge will continue to increase gradually, I can see the brighter light at the end of the road for you.

Son remember when you take a step in life, you walk along with your shadow, nothing to be afraid off, just get on and it will always leads you to a success path, shadow of passion, shadow of future and shadow of life, then you'll be inspired by the theme legacy.

My inspirational thoughts about legacy of life from my perspective background, is to give what it take to get what you want in life. Through honesty, enthusiasm, equality, passion, energetic, positive, sociable and endurance, sky is the limit for success, diplomacy of life goes around life within life circle."

Alhaj Imam Isiaka Rabiu Idowu El Saki (1963-2016)

ACKNOWLEDGEMENTS

First and foremost, all praise is due to Allah (s.w.t) for granting me the ability, strength and determination to complete my PhD. Very special thanks goes out to my Father Alhaj Imam Isiaka Rabiu Folorunsho Idowu El-Saki. May his Soul Rest in Peace and May Allah Grant him Jannatul Firdaus (Ameen Ya Robil-alameen) for all the financial assistant, encouragement and words of wisdom during the completion of this thesis – this will forever be valuable to me.

My gratitude also extends to my director of studies, Dr Paul Fergus without whose support and advice I would not have had the opportunity to undertake this PhD research work. Dr Paul always quotes me Rocky Balboa "*Let me tell you something you already know. The world ain't all sunshine and rainbows. It's a very mean and nasty place, and I don't care how tough you are, it will beat you to your knees and keep you there permanently if you let it. You, me, or nobody is gonna hit as hard as life. But it ain't about how hard you hit. It's about how hard you can get hit and keep moving forward; how much you can take and keep moving forward. That's how winning is done! Now, if you know what you're worth, then go out and get what you're worth. But you gotta be willing to take the hits, and not pointing fingers saying you ain't where you wanna be because of him, or her, or anybody. Cowards do that and that ain't you. You're better than that*". Dr Paul dedicated his precious time to my research work and provided discussions on ideas that eventually became a part of this thesis. I am so grateful to him for polishing my writing skills which has led to the publication of this research work in well-known conferences and journals. In addition, very special thanks go to Dr Abir Husain and Dr Chelsea Dobbins for reviewing my work and giving me valuable feedback.

Many thanks to the technical support group in the Department of Computer Science (CS), LJMU for all of their computer and technical assistance throughout my PhD; and to the office staff Tricia Waterson, Carol Oliver, Elizabeth Hoare, Beverley Williams and Lucy Tweedle for all the instances in which their assistance helped me along the way. Thanks also goes out to those who provided me with advice at times of critical need; Mr Hamed Idowu, Dr Nonso Nnamoko, Mrs Mulenga Kapasa, Dr. Haya Mohmmad Alaskar, Dr Nathan Shone, Mr. Tajudeen Dauda, Mr Sherif Essa, Mr Lanre Okeya Mr Robert Keight, Mr Stephen Akingbe, Ms Áine MacDermott, Mrs Amal Fadak, Dr William Hurst, Mr lanre Sangobunmi, Mr. Mark Welshy, Mr Mohammed Alghamdi, Dr Shamaila Iram and Mr Mohammed Khalaf. I would also like to thank my fellow PhD students in the department of CS at LJMU for the technical debates we had. I would also like to thank Natural Breaks Merseyside Staffs, OCS staffs, for exchanges of knowledge, sharing of skills, and allowing me to vent off my frustrations during my PhD program, which helped enrich the experience.

Finally, my heartfelt thanks go to my family, who deserve great thanks for the immense support they have given me throughout my PhD. My fiancé Kanwal Jabeen Mahmood deserves my unending gratitude for her help and innumerable tolerance, patience, understanding and assistance during my PhD studies. I thank both of my mothers' Mrs Aminat Abimbola Omoluabi and Mrs Diane Idowu for their continuous advice and guidance. A big thank you to my brothers (Ismail, Ali and Adam) and sisters (Halima, Munirat, Melissa, and Layla) for their true understanding and sacrifice in uncountable ways.

This thesis is dedicated to my Father (Alhaj Imam Isiaka Rabiu Folorunsho Idowu El Shaki 1963-2016) and My Supervisor (Dr Paul Fergus)

CONTENTS

LIST OF FIGURES

LIST OF TABLES

THESIS ACRONYMS

ABBREVIATIO N	MEANING
AAC	Average Amplitude Change
AANN	Advanced Artificial Neural Network
AB	AdaBoost
ANN	Artificial Neural Network
ARV	Average Rectified Value
AUC	Area Under the Curve
BN	Bayesian Network
BPXNC	Back-Propagation trained Feed-Forward Neural Network Classifier
CLM	Cervical Length Measurement
DASDV	Difference Absolute Standard Deviation Value
DCE	Digital Cervical Examination
DRBMC	Discriminative Restricted Boltzmann Classifier
ECG	Electrocardiogram
EHG	Electrohysterography
EMD	Empirical Mode Decomposition
EMG	Electromyography
FFNN	Feed-Forward Neural Network
FLNN	Functional Link Neural Network
FN	False Negative
FP	False Positive
FR	Frequency Ratio
HONN	High Order Neural Networks
IEHG	Integrated EHG
IMF	Intrinsic Mode Functions
IPC	Intrauterine Pressure Catheters
KDE	Kernel Density Estimation
KL	Kullback-Leibler
LD	Log Detector
LMNC	Levenberg-Marquardt Trained Feed-Forward Neural Network Classifier
LNN	Linear Neural Network
MAV	Mean Absolute Value
MAVSLP1	Mean Absolute Value Slope1
MDS	Multidimensional scaling
MFL	Maximum Fractal Length
MLA	Machine Learning Algorithm
MLP	Multi-Layer Perceptron

PUBLICATIONS RESULTING FROM THIS THESIS

1. Paul Fergus, Ibrahim Olatunji Idowu, Abir Jaffar Hussain, Chelsea Dobbins. (2016). Advanced artificial neural network classification for detecting preterm births using EHG records. Neurocomputing, 188, 42–49. doi:10.1016/j.neucom.2015.01.107

2. Ibrahim Olatunji Idowu., Paul Fergus, Abir Jaffar Hussain, Chelsea Dobbins, Khalaf Mohammed, Casana Eslava Raul, Robert Keight. (2015). Artificial intelligence for detecting preterm uterine activity in gynaecology and obstetric care. Proceedings - 15th IEEE International Conference on Computer and Information Technology215–220. doi:10.1109/CIT/IUCC/DASC/PICOM.2015.31

3. Ibrahim Olatunji Idowu, Paul Fergus, A. H. & D. C. (2015). Advanced artificial neural network classification algorithms for detecting preterm births using EHG records. In Liverpool John Moores University Research Week. Liverpool.

4. Ibrahim Olatunji Idowu, Paul Fergus, Abir Jaffar Hussain, Chelsea Dobbins, Al-askar Haya (2014). Advance Artificial Neural Network Classification Techniques Using EHG for Detecting Preterm Births. The 8th International Conference on Complex, Intelligent, and Software Intensive Systems (CISIS'14), 8(1).

5. Ibrahim Olatunji Idowu, Paul Fergus, Abir Jaffar Hussain, Chelsea Dobbins (2014). Evaluation of Advanced Artificial Neural Network Classification and Feature Extraction Techniques for Detecting Preterm Births Using EHG Records. Intelligent Computing. (10th Inter., Vol. 8590, p. pp 309–314). Taiyuan, China: Springer. doi:10.1007/978-3-319-09330-7.

6. Ibrahim Olatunji Idowu, Qi Shi, Madjid Merabti, Kashif Kifayat. (2012). Ad-Hoc Cloud Networks: A Probabilistic Model for Vulnerability Detection in Critical Infrastructure Using Bayesian Networks. In 13th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, Liverpool, UK, 2012 (pp. 1– 6), Liverpool, UK

Chapter 1 Introduction

Preterm birth, also known as premature birth or delivery, is described by the World Health Organisation (WHO) as the delivery of babies who are born alive, before 37 weeks of gestation. In contrast, term births are the live delivery of babies after 37 weeks, and before 42 weeks. According to the WHO, worldwide in 2010, preterm deliveries accounted for 1 in 10 births and is now the second leading cause of death in children under 5 years, after pneumonia (WHO 2012). In 2009, in England and Wales, 7% of live births were also preterm. Preterm birth has a significant adverse effect on the new born, including an increased risk of death and health defects. The severity of these effects increases, the earlier the delivery is. Approximately 50% of all perinatal deaths are caused by preterm delivery (Offiah et al. 2012), with those surviving often suffering from afflictions, caused by the birth. These include impairments to hearing, vision, the lungs, the cardiovascular system and non-communicable diseases; up to 40% of survivors of extreme preterm delivery can also develop chronic lung disease (Greenough 2012). These afflictions contribute to the lifelong disabilities of individuals who are born preterm and their families.

1.1 Cause of Preterm Births

Preterm births can occur for three different reasons. According to (Offiah et al. 2012), roughly one-third are medically indicated or induced; delivery is brought forward for the best interest of the mother or baby. Another third occurs because the membranes rupture, prior to labour (PPROM). Lastly, spontaneous contractions (termed preterm labour or PTL) can develop. However, there is still a great deal of uncertainty about the level of risk each factor presents, and whether they are causes or effects. Nevertheless, in (Offiah et al. 2012) some of the causes of preterm labour, which may or may not end in preterm birth, have been discussed. These include infection, over-distension, burst blood vessels, surgical procedures, illnesses and congenital defects of the mother’s uterus and cervical weakness. Further studies have also found other risk factors for preterm (Rattihalli et al. 2012; Steer 2005). These include health and lifestyle choices, cervical and uterine abnormalities, recurrent antepartum haemorrhage, illnesses and infections, any invasive procedure or surgery, underweight or obese mothers, ethnicity, social deprivation, long working hours/late nights, alcohol and drug use, and folic acid deficiency.

1.2 Effect of Preterm Birth on the Infant

Survivors of preterm birth suffer with neuro-developmental or behavioural defects, including cerebral palsy, and motor and cognitive impairment. In addition, preterm births also have a detrimental effect on families, the economy, and society as a whole. In 2009, the overall cost to the public sector, in England and Wales, was estimated to be nearly £2.95 billion (Mangham et al. 2009; Bulletin 2011), while in 1994, in the United States alone, it was estimated that, of the \$820 million spent on hospitalising women with suspected preterm labour, \$360 million was for women who did not actually deliver during their stay. As well as patients being incorrectly diagnosed with preterm labour (false positive results), 20% of patients displaying symptoms or preterm labour were given false negative results: they were denied early admittance, but eventually went on to deliver prematurely (Mangham et al. 2009). In 2001, the cost of care increased rapidly. According to a nationwide survey carried out by (Mangham et al. 2009), hospital costs for preterm infants in the United States was \$12.4 billion.

1.3 Treatment Strategies

Developing a better understanding of preterm deliveries can help to create preventative strategies and thus positively mitigate, or even eradicate, the effects that preterm deliveries have on babies, families and healthcare services. As well as investigating preterm deliveries, several studies have explored preterm labour (the stage that directly precedes the delivery). In spite of these studies, there is no internationally agreed definition for preterm labour.

Nonetheless, in practice, women who experience regular contractions, increased vaginal discharge, pelvic pressure and lower backache tend to show Threatening Preterm Labour (TPL). While this is a good measure, (Mangham et al. 2009), suggest that clinical methods for diagnosing preterm labour are insufficient.

Following a medical diagnosis of preterm labour, only 50% of all women with preterm actually deliver, within seven days (Offiah et al. 2012). In support of this, (McPheeters et al. 2005), carried out a similar study that showed 144 out of 234 (61.5%) women diagnosed with preterm labour went on to deliver at term. This can potentially add significant costs, and unnecessary interventions, to prenatal care. In contrast, false-negative results mean that patients requiring admittance are turned away, but actually go on to deliver prematurely (Lucovnik, Kuon, et al. 2011).

One possible approach to mitigate many of these concerns is to utilise advances made within the machine learning community. This thesis examines the use of Electrohysterography (EHG) signals, feature engineering and machine learning algorithms to classifying term and preterm births. This has been achieved by 1) filtering the raw EHG signals to remove unwanted artefacts, 2) generating features from EHG records to classify preterm and term delivery records, 3) use different sized training sets to test the effect that these proportions have on the results, 4) and use dataset resampling strategies to balance the data. The aim is to providing a viable solution that addresses this global health problem by managing the human gestation period better to improve health and reduce costs.

1.4 Scope of the Research

The aim of this thesis is to develop a new robust methodology that provides a decision support system that helps obstetricians and midwifes make objective decisions about pathological outcomes that may occur during pregnancy. Globally, the rate of preterm births is increasing, which presents significant health, developmental and economic problems

(Fergus et al. 2013). At present, most methods for preterm birth classification, at an early stage, are inadequate and subjective (Lucovnik, Kuon, et al. 2011). Current approaches that measure the process of labour, such as Tocodynamometer, Intrauterine Pressure Catheters (IPC), Cervical Length Measurement (CLM), Fetal Fibronectin, and Digital Cervical Examination (DCE), have numerous problems. For instance, they only measure the onset of labour indirectly and do not detect changes that are characteristic of true labour. Consequently, their classification values for term or preterm delivery are poor. There has however been increasing evidence to show that the analysis of uterine electrical signals, from the abdominal surface, could provide an objective and easy way to diagnose true labour and detect the onset of preterm labour (Garcia-Gonzalez et al. 2013; SMS Baghamoradi 2011; Carré et al. 1998). This has been termed EHG analysis.

Uterine EHG signals refers to the electrical activity that is captured from the uterus during pregnancy, using electrodes. These signals are very useful as they reveal a great deal of information about uterine contractions. For term deliveries, true labour only starts within 24 hours. However, for preterm deliveries, it may start within 7 to 10 days prior to delivery. The change in EHG activity, from non-labour to labour, is dramatic. Therefore, it is expected that a raw dataset of EHG signals would be useful for further analysis. Furthermore, the implementation of different classification algorithms is required so that a conceptual understanding that relates to any physiological changes and differences in pregnancy before labour delivery can be developed.

The dataset that has been used in our research to satisfy this requirement is the Term-Preterm Electrohysterogram (TPEHG) dataset, which can be obtained from a freely-accessible online repository (PhysioNet 2012a). The TPEHG dataset was originally built and used in a study by Fele-Žorž et al. in (Fele-Žorž et al. 2008). In this work, an analysis of the separability of term and preterm delivery EHG records, using various linear and non-linear signal processing techniques, has been undertaken. While the approach focuses on advanced statistical analysis techniques, machine learning techniques are not considered or compared to uncover detailed patterns and non-linear relationships within EGH signals. This is a challenge that will be addressed in this thesis.

Another challenge is the extraction of features from EHG signals during uterus muscle contractions. In many classification tasks the given features are not sufficient to achieve acceptable classification performance. A transformation of the features may yield new features that are more highly correlated with the class value. In addressing this issue a variety of feature extraction techniques are considered in this thesis.

Within the national healthcare services, for many developed countries, there is an immediate need to move from a system focused on treating medical conditions to one that can predict the onset of such conditions (Blencowe et al. 2012; Ren et al. 2015). Therefore, understanding the early signs of a condition and implementing countermeasures are seen as a viable way of reducing spiralling national healthcare costs. However, to date attempts at utilizing computational approaches to achieve sufficient classification have not achieved the high discrimination accuracy that a clinical application requires. In our study, we propose a new analytical approach for assessing the risk of preterm delivery using EHG recordings and advanced machine learning techniques.

1.5 Aims and Objectives

The aim of this thesis is to provide a robust data processing pipeline for signal processing, feature engineering and machine learning to classify term and preterm records. Seen as a decision support system, the approach posited will be evaluated to determine whether it can improve on obstetrician and midwife decision making processes during pregnancy. In other to satisfy the research aim, a number of objectives will need to be considered and these are defined as follows:

- Gain in-depth understanding of TPEHG dataset and conduct initial exploratory data analysis. Any candidate dataset must provide sufficient samples for appropriate data modelling in decision support systems.
- Investigate the filtering and pre-processing processes for the raw EHG signals. Signal artefacts need to be removed to ensure that only the signal directly correlated with the condition under investigation is present and appropriate for feature engineering tasks.
- Select relevant features for extraction from raw EHG signals. Features beyond those currently used in previous works need to be considered and selected based on their ability to maximise the decision boundary between term and preterm records.
- Empirical evaluation of algorithms for classification of term and preterm records. Dependent on the characteristics of the data several classifiers need to be considered and their discriminative capacity evaluated using the feature space generated and compared with previous works to determine their efficacy.
- Address class imbalance limitation evident in the TPEHG dataset. Skewed class distributions pose significant problems in machine learning tasks and thus strategies to address this limitation need to be evaluated.

1.6 Novel Contributions

This thesis proposes a new methodology for EHG signal processing for discriminating between term and preterm records. Our approach provides a robust data pipeline for signal processing; feature engineering; and data modelling using machine learning. On this basis this study claims several novel contributions which are discussed in turn in the following subsections.

1.6.1 Literature Review

- An up to date literature review of current works within the field of preterm birth is presented. This includes a detailed discussion of condition, treatment strategies, including successes and limitations associated with current medical practice. The review also includes a detailed discussion on the use of technology as a support tool for obstetricians and midwifes and the efficacy of such approaches. The state of the art is presented in new and advanced treatment interventions which includes machine learning and how this approach is being studied as an early intervention strategy. This review collectively presents a current holistic view of preterm birth within the scientific community at the point of completing this thesis.

1.6.2 Signal Processing

- Signal processing and feature engineering techniques (extraction and selection) for EHG analysis. State in the art in signal processing is adopted with a discussion on signal filtering and noise analysis. A detailed study of existing features within EHG that are currently used in obstetric care is presented and a robust evaluation of Electromyography (EMG) features used within other domains. Many EMG features have not previously been used in preterm birth monitoring solutions – we are the first to do so. The results show that several EMG features have better discriminative capacity than those typically used in EHG analysis.

1.6.3 Machine Learning

- Use of advanced machine learning algorithms in EHG analysis and preterm birth classification. In this thesis we evaluate several machine learning strategies and provide a detailed evaluation with regards to their detection capabilities using EHG signals. We are the first to provide a broad comparison of machine learning algorithms within the domain of preterm birth analysis. Using this approach we have found several classifiers that classify significantly better than classifications made by obstetricians and midwifes and many other studies reported in the literature.

1.6.4 Objective Measurement for Decision Support

- Improved objective measures for use in decision support systems that significantly outperform obstetrician and midwife prediction of pathological outcomes. We rely solely on the EHG signal to make a distinction between term and preterm deliveries which removes subjectivity from the decision making processes. This provides a significant aid to existing solutions that measures foetus state, but leaves the interpretation of normal and pathological outcomes to medical practitioners. Using machine learning in an ambulatory system provides an additional objective support tool for clinicians to aid the decision making processes.

1.7 Thesis Structure

Chapter 1 of this thesis provides an overview of the problem domain, namely preterm birth and the inefficiencies associated with current treatment and medical interventions. It highlights that the classification capacity of medical practitioners is low and its overall impact on reducing the number of preterm deliveries is ineffectual. The chapter argues that new technological interventions are required, specifically ambulatory decision support systems, to aid the decision making process carried out by medical practitioners in obstetric care. In doing so the challenges are presented, which include advanced signal processing and feature engineering techniques, machine learning, and robust objective measures for the classification of different preterm birth outcomes. The scope of the project is clearly described and the aims and objectives are discussed and the novel contributions this leads to are claimed. Finally, the chapter is concluded with an outline of the thesis structure.

In Chapter 2 we describe what preterm birth is and the impact this condition has on the baby the family and national healthcare costs. This is followed by a discussion on uterine EHG and is use as a clinical tool in obstetric care. It provides a detailed discussion on how EHG data collection is performed and how raw signals are processed. There is a heavy reliance on exploratory data analysis and how specific tools are used in this thesis to provide quick insights in the specific data characteristics. The discussion ends with a detailed account on what kinds of features are typically extracted from EHG raw signals. Finally the chapter is concluded with a summary of the chapter.

Chapter 3 discusses machine learning and the different models that are evaluated using the TPEHG dataset in this thesis. Models that are discussed include different artificial neural network architectures, random forest classifier, support vector machine, naïve bayes, and decision tree classifier. This is followed by a discussion on the background works in machine learning within term and preterm classification tasks. The chapter is then concluded with an overview of the chapter.

A detailed discussion of our methodology is presented in Chapter 4 and the core data pipeline modules used to process signals, extract and select features, train and evaluate machine learning algorithms and validate the results produced by trained machine learning models, are presented. A detailed theoretic framework is provided for all key stages in the methodology ensuring a robust and non-ambiguous approach is presented.

The results are presented in Chapter 5. This includes a discussion on single-direction filtered signals, feature extraction and selectin strategies, oversampling, and classification. Classification results include single model evaluations and classifier combinations. This chapter is then concluded with a detailed discussion of the results. The thesis is concluded in Chapter 5, which provides a summary of the contributions made in this thesis, before the future work is presented.

The results are discussed in Chapter 6 followed by a summary of the contributions made. The thesis is concluded in Chapter 7 and future work is presented.

Chapter 2 Background

2.1 Introduction

This section provides an overview of the work carried out in the main research areas relevant to this thesis, which includes a discussion on preterm birth and its impact, EHG and its use in foetal monitoring. This is followed by a discussion on EHG signal processing and feature extraction strategies.

2.2 Preterm Birth

Preterm birth is defined by the WHO as any birth that occurs before 37 weeks of the gestational period, i.e. the period of gestation <259 days (WHO 2012). Term births are the delivery of babies between 259 – 294 days of the gestation period. More than 1 in 10 of the world's babies are born prematurely and 1 million die as a result of their prematurity (Blencowe et al. 2012). Premature births are the second leading cause of death in children under the age of 5 and for those that survive, they are often left with a lifetime of significant disability (Liu et al. 2012).

According to (Philip Baker 2006), there are two main reasons why babies are born prematurely. The first is due to pre-labour premature rupture of the membrane (PPROM), which is the spontaneous onset of labour. PPROM occurs during pregnancy when the breakage of the amniotic sac (which is commonly known as the mother waters) occurs more than one hour before the onset of labour. The second is based on medical reasons e.g. uterine bleeding leading to placenta previa and placental abruption which causes the foetal membranes to rapture prematurely and trigger preterm labour (Philip Baker 2006).

Other causes and risk factors of preterm birth are the stretching of the uterus. This can occur when a pregnant woman is having two or more babies (an excessive amount of amniotic fluid around the babies occurs and this leads to stretching of the uterus triggering preterm labour). Bacterial infection in the uterus can also stimulate the production of substances that can initiate the onset of uterine contractions leading to preterm labour. The psychological stress of a pregnant woman can also lead to the release of hormones which can cause preterm labour (McPheeters et al. 2005).

These risk factors are considered high during pregnancy and it is essential to protect both the mother and the baby by bringing the labour forward through the process of Caesarean. Outlining the risk factors for the prediction of preterm birth is essential. The maternal history of preterm births is one of the risk factors that is mostly driven by genetic and environmental risk factors (Goldenberg et al. 2008; Blencowe et al. 2013a; Muglia Louis 2010; Plunkett & Muglia 2008). Moreover, uterine over distension with multiple pregnancies has also been regarded as a vital risk factor associated with preterm birth. For instance, there is a high risk of preterm birth with multiple pregnancies compared to single births (Blondel et al. 2006).

Naturally, the occurrence of multiple pregnancies differs among ethnic groups, with reported rates between 1 in 40 in West Africa and 1 in 200 in Japan. However, a large contributor to the incidence of multiple pregnancies has been the result of an increase in the availability of assisted conception in high income countries; thus, this has brought about a proportional increase in the number of births of twins and triplets in many of these countries. For instance, England and Wales, France and the United States have reported a 50% to 60% increase in the twin rate between the mid-1970s and 1998, with some countries (e.g. Republic of Korea) reporting a larger increase in uterine distension with multiple preterm birth pregnancies (Blondel & Kaminski 2002). Recently, this trend has been curtailed as a result of policies put in place limiting the number of embryos that are transferred during in vitro fertilization in some countries i.e. Korea (Marttila 2005), while others countries e.g. Nigeria, India, China etc. continue to report an increase in multiple birth rates (Lim 2011; Martin et al. 2009).

Infection also play a crucial role in preterm birth. For instance, some infections, such as urinary tract, malaria, bacterial vaginosis, Human Immunodeficiency Virus (HIV), and syphilis, have all been associated with preterm birth (Lawn et al. 2010). Furthermore, other conditions, such as cervical insufficiency, that result from ascending intrauterine infection and inflammation with secondary premature cervical shortening, have also been associated with preterm birth (Lee et al. 2008).

Furthermore, some lifestyle factors also contribute to preterm birth. These include, stress, excessive physical work or long times spent standing, smoking, excessive alcohol consumption, as well as periodontal disease and Folic acid deficiency (Lawn et al. 2010). As it can be seen preterm birth is a complex condition with many factors contributing to preterm deliveries. Obtaining the values for different risk factors is complex and information is often obtained through interviews with the patient. This leads to inaccurate data, due to subjectivity, patient confidence about sensitive information (e.g. sexual infections) and errors when recalling information (Smith et al. 2009).

Preterm birth is more common in boys, around 55% of all preterm births are male (Zeitlin 2002). Furthermore, males have a higher rate of mortality compared to girls, born at a similar gestation period (Kent et al. 2011). The role of ethnicity in preterm birth (other than through twinning rates) has been highly debated. However, since the 1970's, various population-based studies have proven that there is evidence backing a variation in normal gestational length within ethic groups (Ananth et al. 2006). While these differences have been associated with socioeconomic and lifestyle factors in some studies, a genetic role has been suggested by recent studies as another cause. For instance, babies of black African ancestry tend to be born earlier than Caucasian babies and have less respiratory distress (McPheeters et al. 2005; Philip Farrell 1975), lower neonatal mortality and do not need special care, unlike Caucasian babies (Alexander et al. 2003).

The number and causes of provider-initiated preterm birth are more variable. Worldwide, the highest affected countries have the lowest coverage of pregnancy monitoring and low caesarean birth rates (less than 5% in most African countries). Nevertheless, recent studies in the United States, have uncovered that more than half of all provider-initiated preterm births, at 34 to 36 weeks of gestation, were carried out in the absence of a strong medical indication (Gyamfi-Bannerman et al. 2011). Due to errors in gestation age assessment, unintended preterm birth can also occur with the elective delivery of a baby that was thought to be term (Neelanjana, Mukhopadhaya 2007).

Furthermore, underlying medical conditions can also cause preterm birth. These can be divided into maternal and foetal conditions. Of the latter, these include, severe preeclampsia, placental abruption, uterine rupture, cholestasis, foetal distress and foetal growth restriction with abnormal tests and are some of the more important direct causes recognized (Ananth et al. 2006). Underlying maternal conditions, such as renal disease, hypertension, obesity, diabetes, and eclampsia, can also contribute to preterm birth. The global epidemic of obesity and diabetes is thus, likely to become an increasingly important contributor to global preterm birth. In one region in the United Kingdom, 17% of all babies born to diabetic mothers were preterm; more than double the rate in the general population (Steer 2005).

2.3 Impact of Preterm Birth

Preterm birth has significant adverse consequences for new-borns, with risk of death and severe health defects. The severity of these effects, in most cases, usually leads to a more premature delivery. For instance, about one half of perinatal deaths are caused by preterm birth (Fergus, Idowu, Ibrahim, Hussain, Abir Jaffar, Dobbins, et al. 2014). Moreover, the proportion of neonatal deaths to preterm births is inversely related to neonatal mortality rates. This often occurs in low income countries, where there are higher rates of neonatal mortality, e.g. India, Pakistan, and Nigeria. These countries have a high rate of mortality as a result of infections, such as sepsis, pneumonia, diarrhoea and tetanus, and intrapartum-related birth asphyxia (Nations et al. 2005). The reason is simply due to a lack of health care facilities and programs for pregnant women, which affects the majority of low-income countries.

The economic cost of preterm birth is enormous. In terms of mitigation against neonatal mortality, and the complex nature of intensive health care required, treatments that are needed for women who are susceptible to delivering preterm are frequently experienced in some countries (Nicholson et al. 2000; Blencowe et al. 2013b). In England and Wales, for example, the percentage of live born babies weighing 2500g or less increased from 6.79% in 1990 to 7.28% in 1996 (Petrou 2003; Chiswick 1985). The percentage rates in Petrou et al. (Petrou 2003), indicates that incidences of preterm babies has increased at a faster rate. For example, the percentage of live born babies weighing less than 1500g in England and Wales increased from 0.96% in 1990 to 1.22% in 1996 and that of live born babies weighing less than 1000g from 0.34% to 0.49%.

While in the United States, one of the most common reasons for antenatal hospitalizations is preterm labour and delivery. The data from Nicholson et al. (Nicholson et al. 2000) indicates that from 1996, pregnancy complication surveillance systems found that preterm labour accounted for 33% of all hospitalizations before delivery in California. National costs for preterm labour, undelivered, were more than $360 million. Incremental costs for preterm labour with early delivery, compared with term delivery, ranged from $21 million to $191 million. This shows that preterm labour can lead to multiple hospitalizations during pregnancy, maternal morbidity, and neonatal mortality.

In high-income countries, the interventions to reduce neonatal deaths are divided into two health-system programmes: maternal-health programmes (covering pregnancy, childbirth, and early neonatal care) and child-health programmes (covering treatment from infancy into childhood). Addressing neonatal mortality requires continuity between these elements of care, which is lacking in many settings. Little care for the neonate often occurs and thus a fraction of the attention in either maternal or child-health programmes is often received. The utmost gap in care often decreases during the first week of life, when most neonatal and maternal deaths occur, as most often there is no access to formal health-care systems at home (Nations et al. 2005).

Preterm birth has profound economic consequences globally. For instance, the cost of a pregnant women's discharge from hospital is minimal compared with the greater cost not only associated with healthcare but also in education (Johnson et al. 2015). This is also a challenging impact of preterm birth because children born preterm are at risk of neuro-developmental pathological conditions resulting from diseases such as, cerebral palsy, vision, and hearing impairments (Saigal, Saroj, Doyle 2008). For example, in a British study up to a third of 7-year-old children born between 32–35 weeks' gestation were reported by their teachers to have difficulties in motor skills, speaking, writing, mathematics, behaviour, and physical education (Huddy et al. 2001). Thus elective deliveries of infants born near term are not without increased risk of mortality and morbidity, which has significant implications for educational services and costs (Saigal, Saroj, Doyle 2008; Huddy et al. 2001). Despite this, educational professionals are lacking educational awareness and formal knowledge in dealing with the educational needs of preterm children.

Another implication of preterm births is hospital re-admissions of preterm children in the weeks after their discharge. In (Escobar et al. 1999), Escobar discovered that more than half of children born prematurely are always re-admitted to hospital, at least once, in the first one to two years of life, mostly as a result of respiratory syncytial virus and other high health illnesses. Even at 10–12 years of age, children who had been born before 26 weeks' gestation had a greater tendency for specialist services, such as physician visits, occupational or physical therapy, nursing or medical procedures, and compensatory needs, than children at full gestational period. A reduction in both the prevalence of health disorders and the use of health-care resources by preterm infants has been reported in (Saigal et al. 2001). This study showed that, by the time premature children reach adulthood, some of them suffer from growth problems, which requires medical attention. Preterm infants require intensive care, which raises the cost of hospital care on the number of days spent in hospital. According to (Strasburger 2006), preterm birth costs England and Wales almost £2.95 billion a year.

In most scenarios, families' are affected by the impact of preterm births and their pathological condition has an enormous negative psychosocial and emotional effect. This persists during the first 2 years of life and by the age of 3 there is often no difference in distress symptoms, with parenting stress remaining greater than normal birthweight infants.

With all this in mind, having a better understanding of preterm birth deliveries can help to create the right decision and prevention strategies to reduce the negative effects associated with preterm birth pathological conditions on families, the economy and healthcare services.

2.4 Uterine Electrohysterography (EHG)

The process of giving birth to offspring, both at term and preterm, involves activation of the Myometrium. The Myometrium represents the middle layer of the uterine wall, consisting mainly of uterine smooth muscle cells (also called uterine myocytes), but also of supporting stromal and vascular tissue. Its main function is to induce uterine contractions (Cookson Victoria 2010; Garfield et al. 1998; Catalin, Buhimschi 1996). The use of EHG can help in recording uterine electrical activity from the abdominal surface to diagnose true labour (Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013).

Several events in the uterine muscle precede labour. Excitability of cells increases due to changes in transduction mechanisms and synthesis of various proteins, including ion channels and receptors for uterotonins (Tezuka et al. 1995; Fuchs et al. 1984). At the same time, systems that inhibit myometrial activity, such as the nitric oxide system, are down regulated, leading to withdrawal of uterine relaxation (Garfield et al. 1998). Electrical coupling between myometrial cells also increases, and an electrical syncytium allowing the propagation of action potentials from cell to cell is formed (Blondel 1996; Leitich et al. 1999). These changes are required for effective contractions that result in the delivery (expulsion) of the foetus.

Previous studies have established that the electrical activity of the myometrium is responsible for myometrial contractions (Kirk Riemer 1998; Kuriyama Hiroshi 1961). Extensive studies have been done to monitor uterine contractility using the electrical activity measured from electrodes placed directly on the uterus (Fele-Žorž et al. 2008; Pajntar et al. 1987). Fele-Žorž et al. found that uterine EHG can be monitored noninvasively from the abdominal surface. Measuring uterine EHG activity produces similar results to tocodynamometry and Intrauterine Pressure Catheter (IPC) systems. In addition, EHG can identify the transition from non-labour to labour states in the myometrium. Many studies have shown that different uterine EHG parameters can indicate myometrial properties that distinguish physiological contractions from true term and preterm labour, which is something contraction-monitoring devices cannot do (Ivan Verdenik, Marjan Pajntar 2001).

In the context of EHG analysis raw signals from the myometrial muscle are typically obtained using bipolar electrodes adhered to the abdominal surface. These are spaced at a horizontal, or vertical, distance of 2.5 cm to 7 cm apart. Most studies use four electrodes, although one study utilizes two (Greenough 2012). In a series of other studies, sixteen electrodes were used (McPheeters et al. 2005), and high density grids of 64 small electrodes were used in (Steer 2005). The results show that EHG may vary from women to women. This is dependent on whether she is in true or false labour, and whether she will deliver at term, or prematurely. Since the late 70s, one theory suggests that gap junctions are the mechanisms responsible. Nevertheless, more recently it has been suggested that interstitial cells or stretch receptors may be the cause of propagation (Nicholson et al. 2000). Gap junctions are groups of proteins that provide channels of low electrical resistance between cells. In most pregnancies, the connections between gap junctions are sparse, although gradually increasing, until the last few days before labour. A specific pacemaker site has not been conclusively identified, although, there may be a generalised propagation direction, from the top to the bottom of the uterus (Lucovnik, Kuon, et al. 2011).

Before analysis or classification tasks, EHG signals, in their raw form, need pre-processing. Pre-processing can include filtering, de-noising, wavelet shrinkage or transformation and automatic detection of bursts. Recently, studies have focused on filtering EHG signals using a bandpass filter between 0.05Hz and 16Hz (Léman et al. 1999; Ivan Verdenik, Marjan Pajntar 2001; Maner 2003). In other studies filters as high as 50Hz have been used (Garfield et al. 1998). Some argue that using EHG with a wide range of frequencies is not recommended, since interference affects the signal.

2.4.1 EHG Feature Extraction

Zardoshti et al. (Zardoshti Wheeler 1993) evaluated a number of features commonly used when dealing with EHG signals. These included integrated absolute value, zero crossings and auto-regression coefficient. However, despite their good discriminant capabilities, a precise frequency threshold for accurate contraction and delivery classification, over different patients, could not be determined. Fergus et al. (Fergus et al. 2013), conducted a broad study of techniques for analysing the features of the EHG signal where, features such as peak frequency, median frequency, root mean square and sample entropy, performed particularly well when discriminating between term and preterm records, with several of the classification models used to validate the approach reporting very good results.

However, it is in Electromyography (EMG) that we find some new and interesting works. In one such study, Lucovnik et al. (Lucovnik, Maner, et al. 2011) investigated whether uterine EMG could be used to evaluate Propagation Velocity (PV). In this study, the electrical signals of the uterus were measured both in labour and non-labour patients who delivered at term and prematurely. The results indicate that, the combination of Power Spectrum (PS) and PV peak frequency parameters yielded the best predictive results in identifying true preterm labour. However, only one dimension of propagation is considered at a time, which is based on the estimation of time delays between spikes (Lucovnik, Maner, et al. 2011). In comparison, Lange et al. (Lange et al. 2014) estimate the PV of the entire EHG bursts that occurs during a contraction. This has been achieved by calculating the bursts corresponding to a full contraction event. The results illustrate that the estimated average propagation velocity is 2.18 (60.68) cm/s. No single preferred direction of propagation was found (Lange et al. 2014).

Meanwhile, (Alamedine et al. 2013) presented three techniques to identify the most useful features relevant for contraction classification. These included linear features, such as peak frequency, mean frequency and root mean square, and nonlinear features, such as the Lyapunov exponent and sample entropy (Alamedine et al. 2013). In order to choose the most suitable features that represent contractions, feature selection algorithms have been used. This process used a Binary Particle Swarm-Optimization (BPSO) algorithm and calculated the Jeffrey Divergence (JD) distance. This is a Sequential Forward Selection (SFS) algorithm. The results show that the BPSO and SFS algorithms select features with the greatest discriminant capabilities. In this case, out of the six features considered, sample entropy produced the best results.

There has also been increased interest in the use of non-linear EMG and EHG signals to detect term and preterm labour earlier. In one example, (Diab et al. 2013) used four non-linear features to detect labour contractions (Diab et al. 2013). These features were time reversibility, sample entropy, Lyapunov exponents and delay vector variance. The results show that time reversibility produced the highest classification rate of 0.842 percent. In comparison, (Sim et al. 2014) used 26 features in their experiment. These include 18 time domain features and 8 frequency domain features. The features were extracted from 40 signals in the TPEHG database to determine the characteristic differences in uterine muscle activities between term and preterm delivery. The signals are divided into four groups, depending on the time of recording (before or after the 26th week of gestation) and the length of gestation (term delivery $\geq$ 37 weeks and preterm delivery $<$ 37 weeks). The results illustrated significant differences between term and preterm records before 26 weeks when Frequency Ratio (FR) and Mean Absolute Value Slope1 (MAVSLP1) have been used. While other features, such as Willison amplitude (WAMP), Slope Sign Change (SSC), and 3rd Spectral Moments (SM3) show substantial differences between preterm and term delivery data recorded during the later period of gestation.

Ye-Lin et al. (Ye-Lin et al. 2014) developed a tool that provides automatic segmentation of EHG recordings, whilst distinguishing between uterine contractions and other artefacts. This was achieved by using an algorithm that generates the Tocography (TOCO) signal, derived from the EHG, and detects windows with significant changes in amplitude. In order to develop the classifier, a total of eleven spectral, temporal, and nonlinear features were extracted from the EHG signal for 12 women, classed, by experts, as being in the first stages of labour. The combination of characteristics that led to the highest degree of accuracy in detecting artefacts was then determined. Using only seven features, the results produced a precision of 92.2%. This study determined that it is possible to obtain automatic detection of motion artefacts in segmented EHG recordings.

While Venugopal et al. (Venugopal et al. 2014) attempted to analyse Surface Electromyography (SEMG) signals in patients with and without muscle fatigue, using Multiple Time Window (MTW) features. In the experiment, SEMG signals were recorded from the muscles in the biceps brachii of fifty volunteers. Using four window functions (rectangular, Hamming, trapezoidal, and Slepian windows), eleven multiple time window features were acquired (Venugopal et al. 2014). These were selected using a genetic algorithm and information gain based ranking. In addition to this experiment, four different algorithms (naïve Bayes, support vector machines, k-nearest neighbour, and linear discriminant analysis) were also been evaluated. The results show that, under fatigue, there was a reduction in the mean and median frequencies of the signals. The k-nearest neighbour algorithm was the most precise in classifying the features, with a maximum accuracy of 93%. Meanwhile, (Vasak et al. 2013), studied whether uterine EMG can identify inefficient contractions. This can lead to first-stage labour and caesarean delivery in term nulliparous women, with the unplanned onset of labour. In this study, EMG was recorded during spontaneous labour in 119 such cases, with singleton term pregnancies in the cephalic position. Electrical activity of the myometrium, during contractions, is characterized by its power density spectrum (PDS) (Vasak et al. 2013). The diagnosis of labour has been made if the patient was in active labour, with no increase in dilation, for at last two hours. The data was analysed to calculate the Intra-class correlation coefficients. This has been achieved by comparing the variance of contraction characteristics, within subjects, to the variance between subjects. The result illustrated that mean peak frequency in women undergoing caesarean delivery, for first-stage labour, was significantly higher (0.55Hz), than in women delivering vaginally without (0.49Hz) or with (0.51Hz) augmentation of labour ($P = .001$ and $P = .01$, respectively). Augmentation of labour increased the mean PDS frequency when comparing contractions before and after the start of augmentation. This increase was only significant in women who eventually delivered vaginally. However, the paper fails to use additional aspects of intra-partum recordings in vitro analysis, testing the hypothesis of a link between an increase in peak frequency and lactic acidosis and impaired in vitro contractility. Furthermore, it also fails to consider other parameter analysis subsets (i.e. sample entropy, root mean square or wavelet). This could be because, depending on the dataset and parameter analysis equation, the use of different parameter analysis techniques is more challenging in getting meaningful EMG signals. Additionally, if these methods had been applied effectively, it would have led to greater classification results.

Fractal dimension was explored by (Maner et al. 2006) as a non-linear feature in order to determine the onset of labour. The fractal dimension is a measure of how fractal the signal is, that is, how much self-similarity or repetition of its own pattern it displays as a ratio of the scale of zoom. Lucovnik et al. (Lucovnik et al. 2010) showed promising results for separating patients (n=27) into those that would give birth spontaneously within 24 hours (n=14), and those that would not (n=13); all of the patients eventually had term deliveries. No machine learning algorithms were used, but different cut points of the fractal dimension were trialled to see what sensitivity and specificity would result. For example, for a fractal dimension cut-point of 1.220, the analysis technique had a sensitivity of 1.0 and a specificity of 0.231. As the study noted, a further study involving more patient data, and a head to head comparison of this technique against other linear techniques would prove useful.

2.5 Feature Selection

Feature selection is capable of improving the learning performance of classifiers, lowering computational complexity, building better generalizable models, and decreasing required storage. The dynamic nature and invaluable advantages of feature selection in the field of pattern recognition, statistics, machine learning and data mining communities is highly commendable (Włodzisław Duch, Tomasz Winiarski, Jacek Biesiada 2003; Shardlow 2007; Isabelle Guyon 2003). Feature selection is purposely designed to choose a subset of input variables by reducing features that are inappropriate or have limited predictive information. Feature selection has been used and proved to be effective both in theory and practice in enhancing learning efficiency, increasing predictive accuracy and reducing the complexity of learned results (Andrew Ng 1998). Thus, feature selection in supervised learning has a main goal of detecting a feature subset that produces higher classification precision. When the dimensionality of a domain increases, the number of features *N* raises, therefore finding an optimal feature subset is intractable and problems related to feature selection have proved to be *NP*-hard (Kirsopp et al. 2002). At this juncture, it is essential to describe traditional feature selection processes, which consists of four basic steps, namely, subset generation, subset evaluation, stopping criterion, and validation (Savoy 2013; Chen 2003).

Subset generation can be described as a search process that produces candidate feature subsets for evaluation based on a certain search strategy. Each candidate subset is assessed and related with the previous best one according to a certain evaluation. However, if the new subset is better, it replaces the best one. This process is repeated until a given stopping condition is satisfied. The next step is the categorization of features. This process determines the importance of any individual feature, neglecting their possible interaction. Categorizing methods are based on statistics, information theory, or on some function of a classifier's outputs (John 1997; Alamedine et al. 2013; Tang et al. 2014). Algorithms for feature selection fall into two broad classifications namely; a wrapper that uses the learning algorithm itself to evaluate the usefulness of features and filters that evaluate features according to heuristics based on general characteristics of the data (Shardlow 2007).

2.6 Exploratory Data Analysis in EHG

In statistics exploratory data analysis (EDS) is a powerful tool capable of providing very quick and very insightful information about EHG signals under investigation. It allows the main characteristics in EHG to be summarised, typically using visual representations. In data science EDS it is often used to see what the data can tell you before formal modelling or hypothesis or testing tasks are conducted. Each of the main EDS techniques utilised in this thesis are discussed in detail in the following subsections.

2.6.1 Box Plot

A box plot utilises standard techniques for presenting 5-number summaries, which consist of the minimum and maximum range values, the upper and lower quartiles, and the median (Frigge et al. 1989; Potter 2006). This technique is a quick way to summarize the distribution of data datasets. This, like many of the other techniques presented below, is described as a descriptive statistic that graphically depicts groups of numerical data through their quartiles. An example of a box plot is illustrated in figure 2.1.

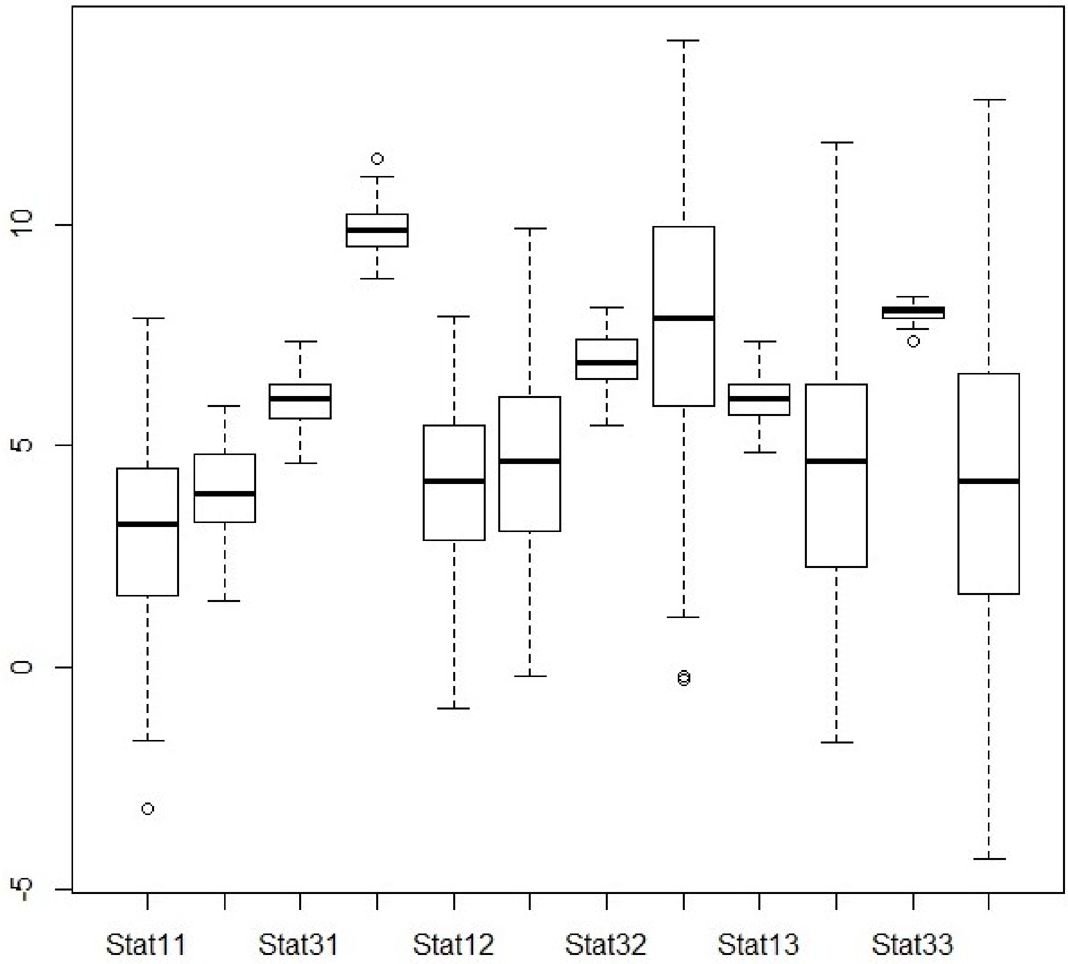


Figure 2.1: Boxplot Example

2.6.2 Histograms Plot

A histogram is a type of graph that shows the frequency distribution of data within equal intervals (Freedman, David, Robert Pisani 1997). Histograms use a function m_i that counts the number of observations that fall into each of the disjoint bins. The mathematical representation of histograms (see equation 2.1) is as follows. Let n be the total number of observations and k be the total number of bins, the histogram m_i count the number of observations. This technique has helped to visually illustrate the distributions of the selected features against the expected normal distributions estimated from the data.

$$n = \sum_{i=1}^{k} m_1 \tag{2.1}$$

Figure 2.2 provides a graphical example of a histogram plot.

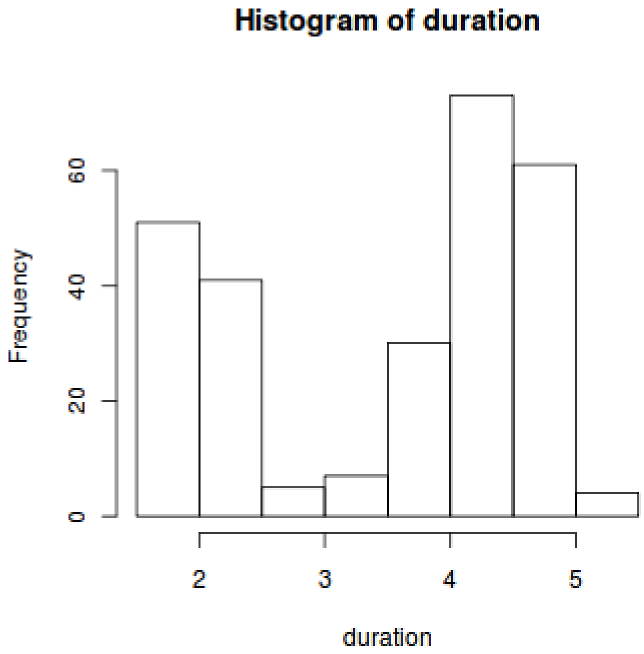


Figure 2.2: Histogram Example

2.6.3 QQ Plots

A q-q plot is a plot of the quantiles of the first data set against the quantiles of the second data set (Population 2012). The theoretical QQ plot (see equation 2.2) examines whether or not a sample $X_{1\ldots,}X_n$ has come from a distribution with a given distribution function $F\,(x)$. The plot displays the sample quantiles $X_{1\ldots,}X_{(n)}$ against the distribution quantiles $F^{-1}\,(p_1), \ldots, F^{-1}(\,p_n\,)$, where:

$$p_1 = \frac{i - {}^{1}/_{2}}{n}$$

(2.2)

Figure 2.3 provides an example of a q-q plot.

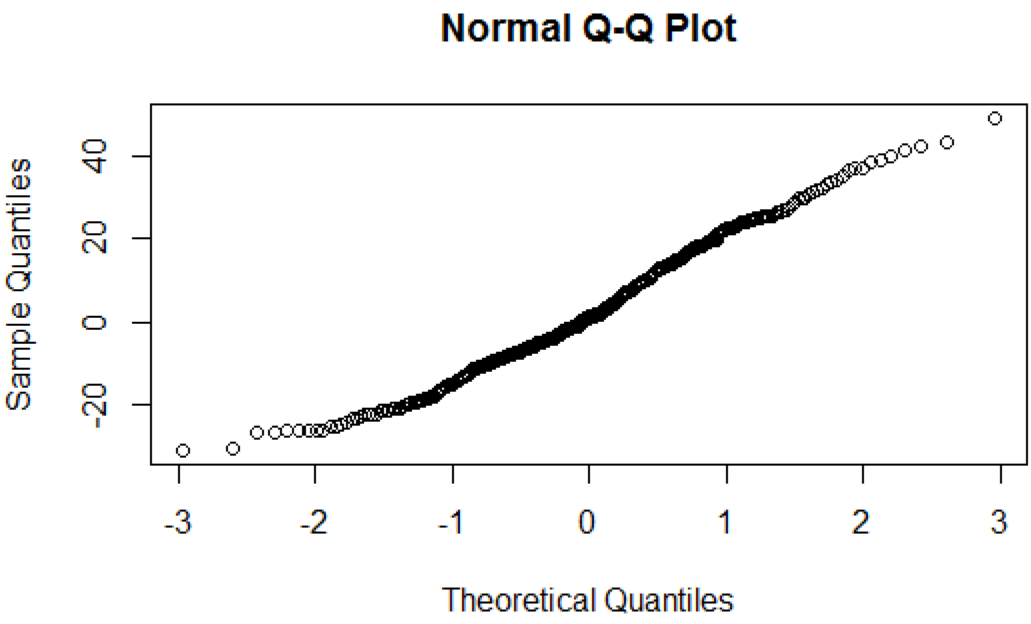


Figure 2.3: Q-Q Plot Example

2.6.4 Principle Component Analysis

Reducing feature vector dimensions, commonly known as feature reduction, helps to avoid over-fitting depending on the algorithm selected. Over-fitting is a serious problem for large-scale datasets with lots of missing values. It can also occur when analysing results without removing redundant information or eliminating noise, leads to non-reliable results. To remove the ambiguity of biased results, due to "feature redundancy," Principal Component Analysis (PCA) is often used. During PCA, three components are calculated, Eigenvalues, Eigenvectors, and Scores. Eigenvalues measure the amount of variation explained by each principle component, with the first coefficient being the largest, eigenvectors are a linear combination of the original variables and have a corresponding Eigenvalue, and scores are used in the bi-plot to represent the data by illustrating how close the features are to the first and second principle components (Powell Victor 2016; Diab et al. 2009; Wold et al. 1987). Through the exploration and visualization of PCA, this technique helps to emphasize variation and find patterns in data of high dimension. Again, figure 2.4 shows and example of this visualisation technique.

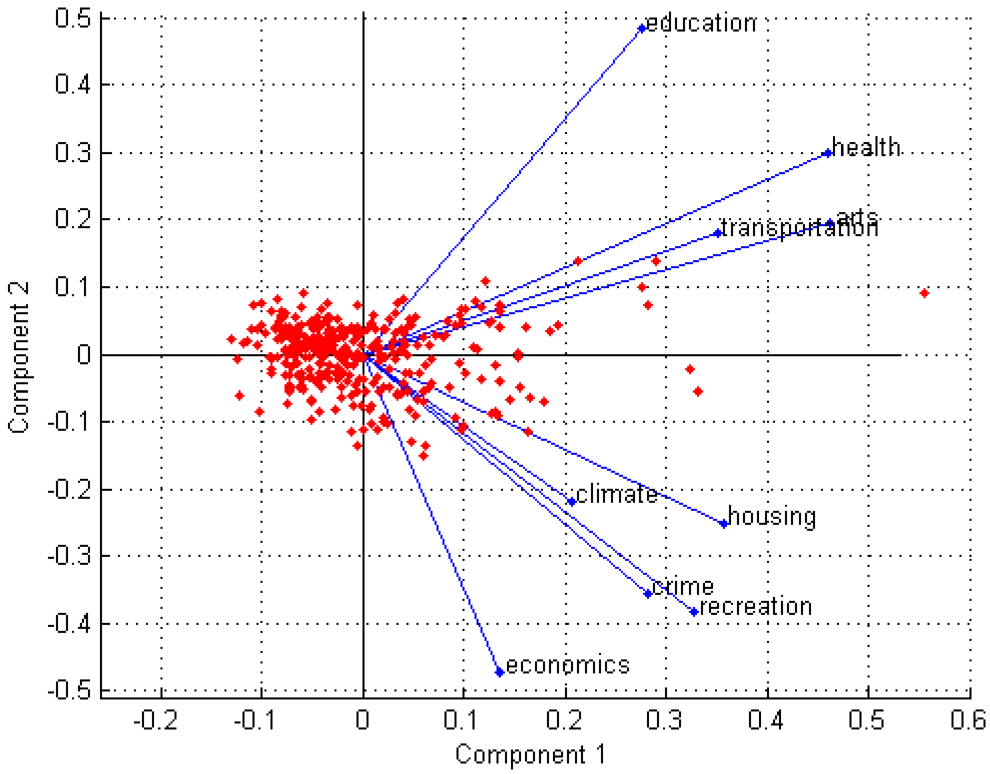


Figure 2.4: PCA Plot Example

2.6.5 Stochastic Neighbour Embedding

The Stochastic Neighbour Embedding (SNE) algorithm was introduced by Geoffrey Hinton et al. (Hinton Geoffrey 2002) and places objects in a low-dimensional space so as to optimally preserve neighbourhood identity; when extended it allows multiple low-d images of each object (Hinton Geoffrey 2002; Agrafiotis 2003; Dietterich 1997). Equation 2.3 describes the mathematical representation of the stochastic neighbour embedding algorithm. In this equation, i represents each object, j is each potential neighbour, p_{ij} is the computation of the asymmetric conditional probability, q_{ij} represents the low-dimensional counterparts of y_i and y_j the high-dimensional data points of x_i and x_j and E_i is the average over data samples i.

In this algorithm, a probability distribution is defined in the input space, based on the pairwise distances, to describe how likely it is that point i is the neighbour of point j. The same is done in the low-dimensional output or projection space. The algorithm then optimises the configuration of points in the output space, such that the original distribution of neighbours is approximated as closely as possible in the output space. The natural measure of approximation error between distributions is Kullback-Leibler (KL) divergence, which is averaged over all points. The SNE algorithm searches for the configuration of point y_i that minimizes KL divergence D between the probability distributions in the input and output spaces, averaged over all points (Venna Jarkko 2007; Laurens van der Maaten 2008). The cost function is:

$$E_{SNE} = E_i[D(p_i, q_i)] \propto \sum_i D(p_i, q_i) = \sum i \sum_{j \neq i} p_i j \log \frac{p_{ij}}{q_{ij}} \tag{2.3}$$

2.6.6 Correlation Scatter Matrix Plot

Correlation scatter matrix plots are used to represent a correlation between two variables (Pace 2012). There are two types of correlations, positive and negative. In positive correlation, as one variable increases, the other variable also increases. Similarly, as one variable decreases, the other variable also decreases. This means that the variables move in the same direction. Within negative correlation the variables move in the opposite direction. Therefore, as one variable increases, the other variable decreases. Alternatively, as one variable decreases, the other variable increases. Scatterplot matrices are used to approximately determine a linear correlation between multiple variables. This is particularly helpful in pinpointing specific variables that might have similar correlations. The mathematical representation of this technique is shown below in equation 2.4. r represents the Pearsons Correlation and is always a number between -1 and 1. $\frac{1}{n-1}$ represents the value of the covariance. In a positive scenario, $r > 0$ and in the negative case scenario $r < 0$. In the plot, the value of r near 0 indicates a weak linear relationship. The strength of the linear relationship increases when r moves away from 0 towards -1 or 1.

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \vec{X}}{s_x}\right)\left(\frac{y_i - \vec{Y}}{s_y}\right) \tag{2.4}$$

An example of a scatter matrix plot is illustrated in Figure 2.5.

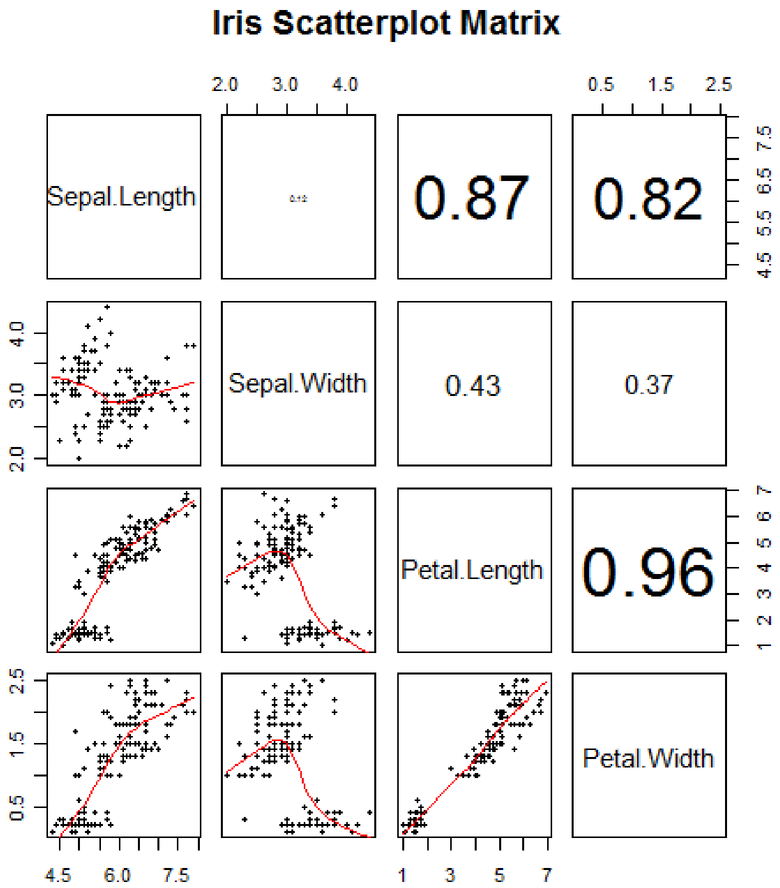


Figure 2.5: Scatter Matrix Plot Example

2.6.7 Kernel Density Estimation Plot

The Kernel Density Estimation (KDE) plot is another statistical EDA method that is used extensively in data analysis. This technique is a non-parametric representation of the Probability Density Function (PDF) of a random variable (Sheather 2004). This technique has gained wide interest over the last 20 years and has been applied to different research topics (Sheather 2004). For example, in (Baxter et al. 2000), the authors used this technique to estimate statistical analysis of lead isotope ratio data in archaeology. Tortosa et al. (Tortosa-Ausina 2002) used it on financial datasets to normalize the cost index estimation for different periods. KDE is used to avoid assumptions about the distribution of the data. This distribution is defined by a smoothing function and a bandwidth value that controls the smoothness of the resulting density curve. Its formula is given in equation 2.5 below.

$$\hat{f}_h(x) = \frac{i}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right); -\infty < x < \infty \tag{2.5}$$

Where n is the sample size, $K(.)$ is the kernel smoothing function, and h is the bandwidth. Figure 2.6 shows an example of kernel density estimation plot.

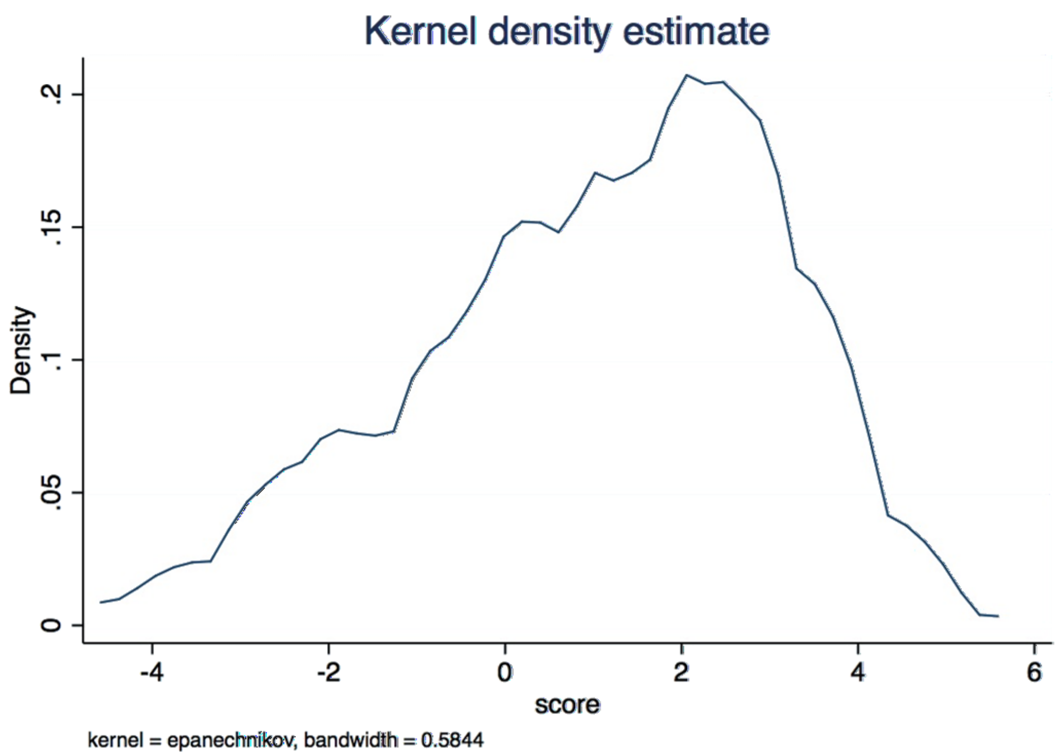


Figure 2.6: Kernel Density Estimation Plot Example

2.7 Summary

This chapter described what preterm birth is and the main cause and risk factors associated with it. It also explored the impact preterm birth has on the infant, the family and healthcare providers and presented a case for the use of uterine Electrohysterography (EHG) signals for diagnosing true labour. This was followed with a discussion on EHG feature engineering and how exploratory data analysis in EHG analysis, and other data processing domains, is often used to provide useful insights into data before formal modelling is conducted. The following chapter builds on the discussions in this chapter and presents information about machine learning and how it can be used to formally model data for the purpose of classification – in this instance classifying term and preterm birth records.

Chapter 3 Machine Learning

3.1 Introduction

Machine learning is the automatic assignment of a class to feature vectors previously extracted from raw signals. Machine learning algorithms are trained to classify these feature vectors where the class is not known. In order for a classifier to learn about a particular application domain, it is given a dataset to work with. This dataset is then divided into two separate sets – one for training and one for testing. In the training phase, a classifier is built. The classifier recognises patterns in the training data that corresponds to each of the classes modelled; it learns the distribution of features for each class and learns to generalize. In the testing phase, the classifier is given the test set, which are new feature vectors with no class labels; the trained models attempt to classify the unseen instances into each of the classes modelled. In machine learning algorithms, learning can be divided into 3 categories – supervised, unsupervised or reinforced. In supervised learning, the labels for each class are provided for the classifier during the training stage. In unsupervised learning (also known as clustering), the class labels are omitted and the classifier is asked to cluster the instances into groups where they display the same patterns or features, and hence each cluster may represent one class. In reinforcement learning, the classifier makes a classification of each instance and is given a score after each classification, to tell it how well it classified the instance. The classifier then adjusts its future actions accordingly (Stuart Russell 1995).

Many algorithms exist to perform modelling in this way, each with their own set of strengths and weaknesses. The remainder of this chapter discusses some of the popular machine learning algorithms currently being used in biomedical research.

3.2 Artificial Neural Networks

An artificial neural network (ANN) is a machine-learning technique that is modelled on the neural connections in our brains. ANN research began in 1940 and has since been used by scientists to resolve many different machine learning problems (Zhang 2000). Current research in this area is flourishing (Ghaffari et al. 2006a), (Xu, Xin and Tan, Hongzhuan and Zhou, Shujin and He, Yue and Shen, Lin and Liu, Yi and Hu, Li and Wang, Xiaojuan and Li 2014). The biological model of ANNs can be seen in Figure 3.1. There are approximately ten thousand million neurons in the human brain. Each neuron is connected to many thousands of other neurons, which can be trained to perform complex calculations. Click on the image for a hyperlink to the source of the publication.

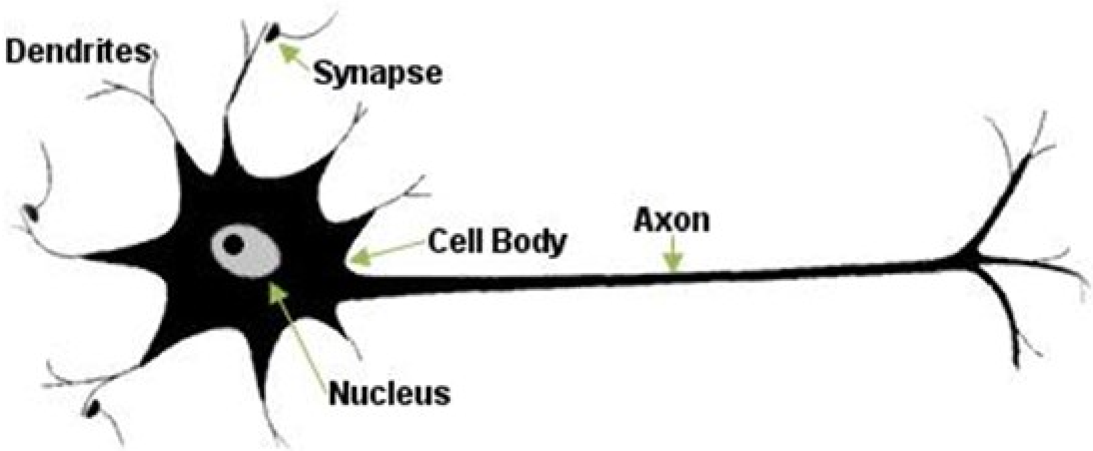


Figure 3.1: Biological Representation of a Neuron cell Processing (Anil K. Jain 1996)

Figure 3.1 represents the biological way the brain pre-processes information. The brains neural network architecture is composed of cell bodies or somas, and two types of branches (axon and dendrites). The cell body has a nucleus that contains information about transmissible traits and plasma that holds the molecular equipment for producing the material needed by the neuron. Neurons receive signals from each other through dendrites and transmit signals generated from the cell body to the axon. The output signals travel through the axon until they reach a synapse, where units determine whether they are passed in as inputs to the next neuron. The controlling power of the synapse can be adjusted to change the amount of signal that is allowed to pass from an axon to a dendrite (Anil K. Jain 1996). A good scenario that demonstrates how the brain works is the process of when we learn to read, write and understand speech. We recognise and distinguish these patterns when we are children through different learning examples. The ANN does it in the same way and learns from trained examples rather than being programmed. During the training period, the neural network can adapt itself, based on the examples of similar problems after sufficient training. Artificial Neural Networks are based on the connection of perceptron's, which is the abstract equivalent model of a neuron used in machine learning (Jain et al. 1996).

There are two common types of ANN structures that are mostly used by researchers to solve classification and prediction problems. These are a feed-forward neural network (FFNN) and a recurrent neural network (RNN). In feed-forward neural networks, the information from the input layer is transmitted down through the network layer until it reaches the output layer. The recurrent neural network incorporates recurrent links into its structure, such as a feedback connection, which makes them more dynamic. This is the main ANN considered in this thesis given their ability to store information for long periods.

Feed-forward neural networks are referred to as multilayer perceptron's (MLP). It is the most popular type of neural network used today when designing artificial neural network architectures. In the FFNN or MLP structure, the neurons are gathered into layers (Kuldip Vora 2014), (Hornik Kurt, Stinchcombe Maxwell 1989). The first and last layers are called the input and output layers, because they represent the inputs and outputs of the neural network. The remaining layers are called hidden layers. The hidden layer provides neural networks with additional learning capabilities to learn from patterns discovered in the dataset. There are two types of feed-forward neural networks. These are single layered and multi-layered ANNs (Kuldip Vora 2014; Zhen-Guo, Tzu-An Chiang 2011; Naik Arti 2012; Catley et al. 2006; Ghaffari et al. 2006b; Hornik Kurt, Stinchcombe Maxwell 1989).

In a single layered network, all the inputs are connected directly to the outputs, commonly known as a perceptron. Although the perceptron does have advantages over decision trees, many authors have, however, commented on the kind of problems the single-layered perceptron can handle, such as its inability to handle the XOR function (Kuldip Vora 2014). However, with modifications to the simple one-layer network, solutions have been found to overcome these limitations. This involves using non-linear activation functions and increasing the number of layers in the network. In a single perceptron, it can only deal with classes that can be separated on an x/y graph with a straight line (linearly separable) and binary outputs.

The second type of feed-forward neural network is the multi-layered network architecture, which involves the combination of several perceptron's to create a nonlinear decision boundary. A multi-layered network involves one or more hidden layers. The functions of these hidden layers are adjusted in response to training, in order to bring the correct response. The output values are compared with the correct answer to compute the value of some predefined error-function. This error is then fed back through the network. Using this information, the algorithm adjusts the weights such that the error decreases with each iteration and the neural model gets closer and closer to producing the desired output. This method uses the error back-propagation algorithm and is one of the simplest and most widely used algorithms by researchers interested in error correction (Kotsiantis 2007). Error correction can either be done through forward pass or backward pass. The architecture of this network is illustrated in Figure 3.2.

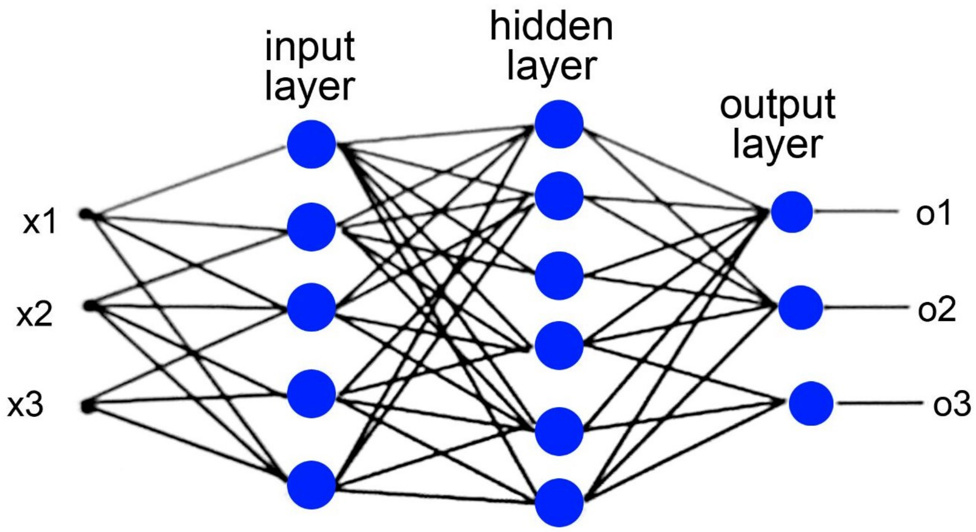


Figure 3.1: Feed-forward multi-layered neural network architecture

Figure 3.2 shows a sample feed forward neural network architecture for approximating a classification function that maps an input vector to more than one class. The network consists of N_s number of layers. The first layer is the input layer N_i, the second layer is the hidden layer N_h, and the output layer is N_u. Let us suppose the input is *x*= $[x_1, x_2\ x_3, \ldots, x_{ni}]$, the network output is y=$[O_1\ , O_2, O_3, \ldots, O_{nu}\ ]$ and the weights are represented as w_{wj}. The inputs are first passed to the input units in the input layer and then the outputs from the input units are passed to hidden layer until it reaches the last hidden layer units. The hidden layer outputs are passed to the output layer units. The optimized weights can be done through supervised learning, where the network learns from the large number of observations. Examples are usually provided one at a time. For each example, the actual vector is computed and compared to the desired output. Then, weights and thresholds are adjusted, relative to their contribution to the error made at the respective output. As previously mentioned, one of the most used methods is the backpropagation algorithm, in which the errors are propagated (error = the difference between actual and expected results) in the lower layers. However, the selection of the optimal number of hidden layers and hidden neurons for the required task is very challenging.

Although there are a large number of applications of the well-known MLP neural network, they suffer from difficulties, such as the determination of the optimal number of hidden units, and estimating the best weight values. The selection of these parameters is very important to improve the performance of neural networks. Furthermore, the MLP neural network is affected by some learning algorithm problems, such as over-fitting (Cao 2003). This means that the neural network can perfectly map between input and output in training data but it will not be able to sufficiently generalise its learning to new data.

3.2.1 Back-Propagation Trained Feed-Forward Neural Network Classifier

In the Back-Propagation trained Feed-Forward Neural Network Classifier (BPXNC), the network is trained to map a set of input data to outputs through iterative adjustments of the weights. The information from inputs is fed forward through the network to optimize the weights between neurons. Moreover, the optimization of the weights is made by backward propagation of the error during the training or learning stage. The BPXNC then reads the input and output values in the training dataset and changes the value of the weighted links to reduce the differentiation between the predicted and observed values. The error in prediction is reduced across several training cycles until the network reaches the best level of classification accuracy, while avoiding overfitting (Ghaffari et al. 2006b). The advantage of this classifier is the process of looking for the minimum of the error function in weight space using gradient descent. The combination of weights, which minimizes the error function, is considered to be a solution for learning problems the classifier.

3.2.2 Levenberg-Marquardt Trained Feed-Forward Neural Network Classifier

The Levenberg-Marquardt trained Feed-Forward Neural Network Classifier (LMNC) has similar functionalities to the BPXNC. However, it is much more memory intensive. The classifier provides a numerical solution to the problem of minimizing nonlinear functions over a space of parameters. Furthermore, during the training stage, training is stopped when the performance on an artificially generated tuning set of 1000 samples, per class, has been reached (based on k-nearest neighbor interpolation) and thereafter does not improve anymore (37steps 2013)-(Zhao et al. 2005).

Since LMNC is a feed-forward network, it is designed to approach second-order training speed, without having to compute the Hessian matrix. For example, in equation 3.1, J represents the Jacobian matrix, which contains the first derivatives of the network errors with respect to the weights and biases. μ represents the gradient computed and e represents the vector of the network errors.

$$\mathrm{x}_{K+1} = \mathrm{x}_{k} - [\mathrm{J}^{\mathrm{T}} - \mathrm{J} - \mu\mathrm{I}]^{-1}\mathrm{J}^{\mathrm{T}}\mathrm{e} \tag{3.1}$$

The scalar μ is zero; this is just Newton's method, using the approximate Hessian matrix. Newton’s method is quite fast and more accurate. When μ is large, this becomes gradient descent with a small step size. Thus, μ is decreased after each successful step (reduction in performance function) and is increased only when a cautious step would increase the performance function. In this way, the performance function is always reduced at each iteration of the algorithm. This algorithm appears to be the fastest method for training moderate-sized feed-forward neural networks (Zhao et al. 2005).

3.2.3 The Radial Basis Function Neural Network Classifier

The Radial Basis Function Neural Network Classifier (RBNC) is mostly used in complicated pattern recognition and classification problems. The classifier has three layers – the input layer, a hidden layer with non-linear radial basis activation function units and a linear output layer. The mapping properties of the RBNC can be modified through the weights in the output layer. The input can be modeled as a vector of real numbers, while the output of the network is then a scalar function of the input vector. This ANN classifier is restricted to only contain one hidden layer of units, whose output depends on a distance between the center of arbitrary transfer functions and a given pattern (37steps 2013).

3.2.4 The Random Neural Network Classifier

The Random Neural Network Classifier (RNNC) is a feed-forward neural network, which is inspired by the spiking behaviour of biophysical neurons (Gelenbe 1991). This classifier has attracted a lot of attention in the scientific community. This is because RNNC interacts by probabilistically exchanging excitatory and inhibitory spiking signals. The model is described by analytical equations. Its standard learning algorithm has low complexity and strong generalisation capacity even for a relatively small training data sets (Marx 2008). RNNC has one hidden layer of N sigmoid neurons, which exchanges positive and negative signals in the form of unit amplitude spikes. The input layer rescales the input features to unit variance; the hidden layer has normally distributed weights and biases with zero mean and standard deviation (37steps 2013). One of the advantages of using this classifier is that it is a standard learning algorithm that has a low complexity and strong generalisation capacity for a relatively small training dataset.

3.2.5 Linear Perceptron Linear Classifiers

The Linear Perceptron Linear Classifiers (PERLC) is the simplest type of neural network classifier and is trained with a supervised training algorithm. This classifier assumes that the true classes of the training data are available and incorporated in the training process. The input weights in this classifier can be adjusted iteratively by the training algorithm so as to produce the correct class mapping for the output. However, the problem with this classifier is that it does not have a hidden layer, therefore this leads to bias in result accuracy.

3.2.6 Voted Perceptron Classifier (VPC)

Voted Perceptron Classifier (VPC) is based on a paper by (Freund & Schapire 1996). The VPC is similar to the PERLC classifier and is trained with a supervised training algorithm. The classifier trains an ensemble of perceptron’s on the dataset. The training procedure performs a number of full sweeps through the training data. If a number of sweeps is not specified, 10 sweeps are performed (37steps 2013). In this classifier, the classification of new objects is performed by allowing the ensemble of perceptron's to vote on the label of each test point in the neural network. A VPC is simple to implement and has fast learning speeds. For example, a set of observations *O* can be given, where each observation is a pair *(x, y),* and $x \in R^{\eta}$ is a vector in a given ȩ-dimensional vector space and *y* is the label associated with the observation. Let's assume that the observation labels can have only one of the two values, 1 or -1. The algorithm makes use of the perceptron algorithm by starting with an initial prediction vector, $v = 0$, and predicts the label of the first observation instance x_1 to be $Q = sign(v_{x1})$. If this prediction is different from y_1 it updates the prediction vector to $v = v + y_1 x_1$. If the prediction is correct, then *v* has not changed. The process is then repeated with the next example. The advantage of using the VPC classifier is that it takes advantage of data that are linearly separable with large margins. This is because the algorithm builds on the iterative perceptron algorithm rather than solving quadratic programming problems (37steps 2013).

3.2.7 Discriminative Restricted Boltzmann Machine Classifier

Discriminative Restricted Boltzmann Machine Classifier (DRBMC) is similar to the LMNC and BPXNC neural networks and has been developed for a large variety of learning problems. DRBMC are usually used as feature extractors for other learning algorithm or to provide a good initialization for deep feed-forward neural network classifiers. The classifier uses an undirected generative model with a layer of hidden variables to model a distribution over visible variables. The model introduces binary stochastic latent variables in logistic regression, which turns it into a powerful non-linear model. The stochastic latent variables acts a lot like rectified linear units in neural networks and is trained with regularization using regularization parameters (37steps 2013).

3.2.8 Functional Link Neural Network Classifier

The Functional Link Neural Network Classifier (FLNN) is a High Order Neural Network (HONN) that utilizes higher combinations of inputs (Yoh-Han 1989)-(Pao 1992). HONNs were first introduced in (Giles Lee 1987). Such networks differentiate themselves from a feed-forward neural network by the presence of high order terms in the network. In a simple feed-forward network, the neurons are first order neurons, called linear neurons. This is because of the use of linear sums of its inputs in decision making. This linearity provides a hyper plane for decision making that restricts the capability of neurons to solve only linear discriminant problems (M.Guler 1994). In order to resolve this issue, a multilayer network with hidden units can be used, which can combine the outputs and generate nonlinear mappings. Another way to overcome the restriction to linear mappings is the introduction of higher order units to model nonlinear dependences. This provides better classification capability.

The use of HONNs utilizes a higher combination of learning task relationships between its inputs through summing operations. HONNs contain summing and product units that multiply their inputs in the network. These high order terms, or product units, can increase the information capacity of higher order networks in comparison to standard neural networks, with summation units only. Usually, but not necessarily, the output of a HONN is neither (1, 0) or (+1, -1). The architecture of a three input second order HONNs is shown in figure 3.3. Click on the image for a hyperlink to the source of the publication.

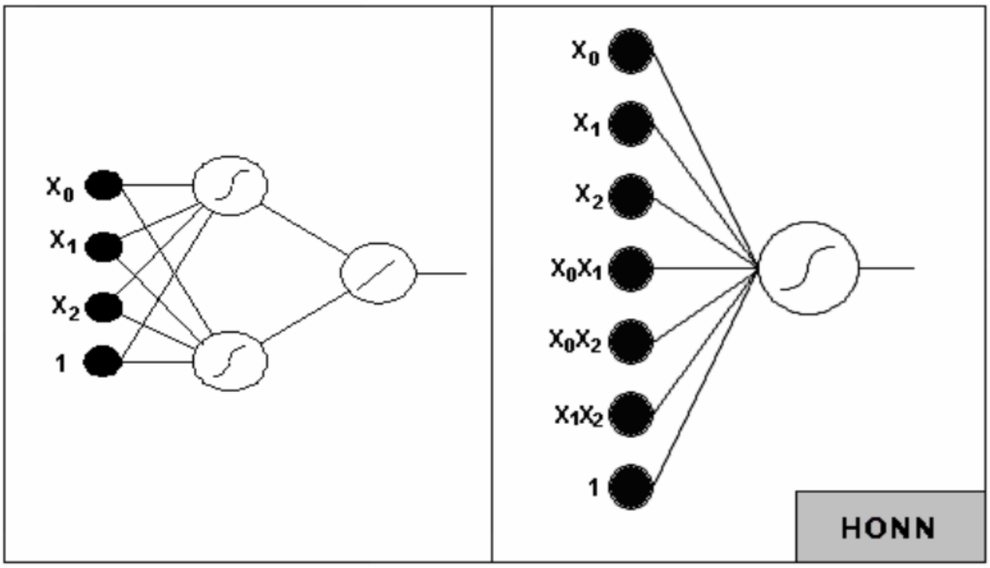


Figure 3.3: Higher Order neural network architecture (Christian et al., 2008)

HONNs make use of non-linear interactions between inputs. The networks therefore increase the input space into another space where linear separability is conceivable (Yoh-Han 1989). HONNs use joint activation between inputs, hence removing the task of establishing relationships between them during training. For this reason, hidden layers are not commonly used. The reduced number of free weights, compared with MLP, means that the problem of overfitting and local optima can be transferred to large a degree.

HONNs can accomplish similar performance tasks to that of standard multilayer neural networks using a single layer of trainable weights (Park et al. 2000). They are simple in their architecture and require fewer weights to learn the underlying equation when compared to ordinary feed-forward networks that deliver the same input output mappings. This means that the training time is potentially less than that of an MLP structure. There are three types of HONN models. These are the Functional Link Neural Network (FLNN), the Pi-Sigma Neural Network (PSNN), and the Ridge Polynomial Neural Network. However in this thesis we will just focus on Functional Link Neural Networks.

The FLNN created by Pao et al. in 1989 has been successfully used in many applications. For instance, it has been used in system identification in (Patra & Bornand 2010) where a computationally efficient Legendre Neural Network (LeNN) for identification of nonlinear dynamic systems was proposed due to its single-layer architecture. The LeNN offers much less computational complexity than that of a MLP. In (Dehuri & Cho 2010) this network architecture was used to solve classification problems and input feature selection to enhance the functional expansion genetically for the purpose of solving the problem of classification in data mining.

Dehuri, et al. proposed the single Hidden layer Functional link Neural Network (HFLNN), which aims to choose an optimal subset of input features by eliminating features with little or no predictive information which is a more compact classifier. While (Cass & Radl 1996) used FLNNs in process optimization and found that FLNNs can be trained much faster than MLPs without scarifying computational capability. This feature makes them more suitable in process modelling applications, where the ability to retrain or adapt to new data in real time is critical.

The network architecture of FLNNs is usually composed of a single layer network to handle linearly non-separable classes by increasing the dimensions of the input space using non-linear combinations of inputs. The architecture of FLNNs is a flat network, without any hidden layers, which makes the learning algorithm used in the network less complicated. In FLNNs, the input vector is extended with a suitably enhanced representation of the input nodes, thereby artificially increasing the dimension of the input space. The extended input data are then used for training, as for standard feed-forward neural networks. Basically, the inputs are transformed in a well understood mathematical way so that the network does not have to learn basic math functions (Pao 1992). The basic equation for calculating FLNN's output is illustrated below in equation 3.4. Click on the image for a hyperlink to the source of the publication.

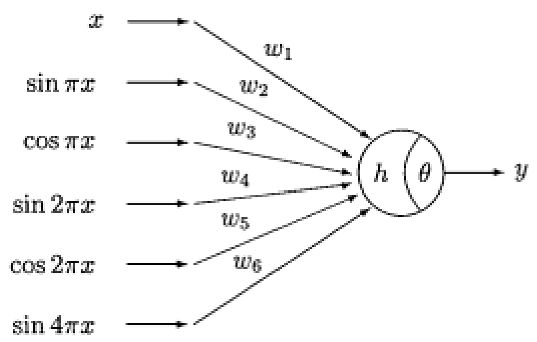


Figure 3.4: Function expansion network (Hongxing Li, C.L. Philip Chen 2000)

The network is formed by adding five other inputs to a one-input network, where x is regarded as a generating input and the other five are function transformations of the original one, w_1 is the adjustable threshold and y is the output.

$$y = w'_0 + w'_1 \, sin\pi x + w'_2 \, cos\pi x + w'_3 \, sin\pi x + w'_4 \, 2cos\pi x + w'_5 \, sin4\pi x \tag{3.2}$$

Functional link neural networks can be categorised into two models; functional expansion models and tensor (outer product) models. In the functional expansion model, the functional link acts on each node singly and may induce the same additional functionalities for each node in the input pattern. While in the tensor (outer product) model, each component of the input pattern multiplies the entire input pattern vector. The functional link in this model generates an entire vector from each of the individual components (Misra 2007). The same process may be described in terms of the formation of an outer product between two vectors, one being the original pattern vector and the other being the same vector augmented by an additional component of value unity.

Augmenting the vector allows the original pattern to be regenerated along with the higher order effects. The two models mentioned are capable of being used in the process of representation enhancement appropriately. However, we plan to use the tensor model because the effect of the nonlinear function transform is to change the representation of the input pattern so that, instead of being described in terms of a set of components [x_i], where $j \geq i$ it is described as $[x_i, x_j, x_k]$ or $[x_i x_j, x_i x_k, x_j x_k]$, where $k \geq j \geq i$ and so on. In a sense, no new information has been added, but joint activation has been made explicitly available to the network. Such functional transforms greatly increase the number of components in terms of which of the input patterns are described. We can simplify the enhanced pattern by omitting terms with two or more equal indices and also terms for which there is no correlation over an ensemble of input patterns.

3.3 Non ANN Classifiers

Looking at alternative models to brain inspired artificial neutral networks we now turn to a set of algorithms based on different mathematical principles. In some instances some of these models are less complex to train but in the same instance have shown results comparable to the most complex of ANN. Each of these is discussed in more detail in the following subsections.

3.3.1 Random Forest Classifier

Breiman et al. (Breiman 2001), proposed the Random Forest classifier, which adds an additional layer of randomness to bagging. In addition to constructing each tree using a different bootstrap sample of the data, random forests change how the classification or regression trees are constructed. In standard trees, each node is split using the best split among all variables. In a random forest, each node is split using the best among a subset of predictors randomly chosen at that node.

The Random Forest algorithm can be trained through the process of an ensemble of B trees $\{T_1(X), .., TB(X)\}$, where, $X = x_1 ... x_p$ p represents the dimensional vector of feature descriptors or properties associated with a class. The ensemble produces B outputs $\{Y_1 = T_1(X).., Y_B = T_B(X)\}$ where $Y_b, b = 1 ..., B$ is the prediction for a class by the b^{th} tree. Outputs of all trees are aggregated to produce one final prediction Y. For classification problems, Y is the class predicted by the majority of trees. In regression, it is the average of the individual tree predictions. In theory, given data on a set of n classes for training, D = $\{(X_1, Y_1)..,(X_n, Y_n)$, where $X_i, i = 1..,n$ is a vector of descriptors and Y_i is either the corresponding class label. Firstly, from the training data of n classes, draws a bootstrap sample (i.e., random sample, with replacement, n classes). Secondly, for each bootstrap sample, grow a tree with the following modification: at each node, choose the best split among a randomly selected subset of m_{try} (rather than all) descriptors. Here, m_{try} is essentially the only tuning parameter in the algorithm. The tree is grown to the maximum size (i.e., until no further splits are possible) and not pruned back. Thirdly, repeat the above steps until (a sufficiently large number) B trees are grown. When $m_{try} = p$ i.e., the best split at each node is selected among all descriptors, the random forest algorithm is the same as Bagging (Cutler et al. 2007; Svetnik et al. 2003; Pal 2005; Marx 2008; Breiman 2001; Biau 2012).

Random Forest algorithms can be very efficient, especially when the number of descriptors is very large. The efficiency of the algorithm, compared to that of growing a single decision tree, comes from two differences between the two algorithms. First, in the usual tree growing algorithm, all descriptors are tested for their splitting performance at each node, while random forest only tests m_{try} of the descriptors. Since m_{try} is typically very small (the square root of the number of descriptors for classification), the search is very fast. Therefore, at each node the Random Forest algorithm only sees m_{try}, rather than p descriptors.

Second, to get the right model complexity for optimal prediction strength, some pruning is usually needed for a single decision tree. This is typically done via cross-validation and can take up a significant portion of the computation. In regression, each tree is grown on the residuals of the previous trees. Prediction is done by weighted vote (in classification) or weighted average (in regression) of the ensemble outputs (Svetnik et al. 2003). Recently, Ham et al. (Ham et al. 2005), have applied Random Forests to the classification of hyperspectral remote sensing data. Their approach is implemented within a multi-classifier system arranged as a binary hierarchy. The obtained experimental results in (Ham et al. 2005) are good for a hyperspectral data set with limited training data.

3.3.2 Support Vector Machine Classifier

Vapnik et al. (Vapnik 2000) describe Support Vector Machine Classifiers (SVMC) as a statistical learning theory that uses a linear separating hyperplane. In other words, given labelled training data (supervised learning) the algorithm outputs an optimal hyperplane, which categorizes new examples. These are done through the algorithm properties, e.g. decision boundaries called a maximum margin separator, so that the distance between points of different instances, on either side of it, are as large as possible, improving generalisation. It also embeds 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. The mathematical equation of the algorithm is illustrated below in equation 3.3.

$$D = \left\{(x_i, y_i) \mid x_i \in R^p,\ y_i \in \{-1, 1\}\right\}_{i=1}^{n} \tag{3.3}$$

In this equation, D is the training data with a set of n points, the class label $y_i = \pm 1$ indicating the class to which the point x_i belongs and x_i is a p-dimensional vector; the SVMC algorithm builds a model by finding the maximum-margin hyper plane (gap) that divides the points $y_i = 1$ from $y_i = -1$; making it a non-probabilistic binary and linear classifier.

The hyperplane lies midway between the two margins as can be seen in figure 3.5. The SVM then has to learn where the optimal hyperplane is. One way to do this is to use gradient descent to search for different combinations of intercept and gradient. The purpose behind maximising the margin between points of different classes is to minimise the probability that the as-yet, unseen, unclassified points may fall on the wrong side of the hyperplane (Sotiris

2007). It has been proven that this improves generalisation, and therefore limits the maximum of the expected generalisation error. Click on the image for a hyperlink to the source of the publication.

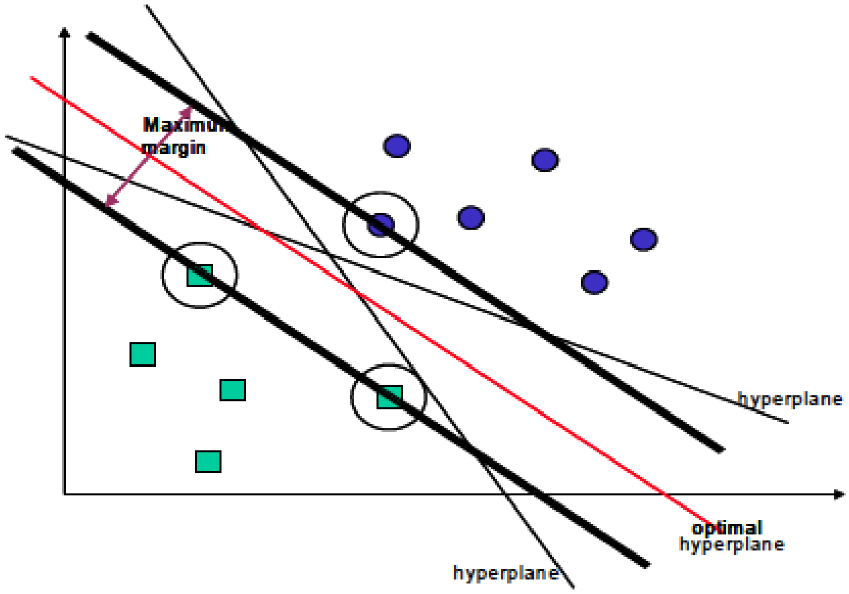


Figure 3.5: SVMC linearly separable set of 2D-points for two classes (Sotiris 2007)

3.3.3 Naive Bayes Classifier

Naive Bayes Classifiers (NBC) are very simple Bayesian networks and are composed of directed acyclic graphs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes). They have a strong assumption of independence among child nodes in the context of their parent (Sotiris 2007). The NBC theorem can be represented mathematically within equation 3.4 as:

$$p(x|y_j) = p(x_1|y_j), p(x_2|y_j), \dots\ p(x_n|y_j) \tag{3.4}$$

In this equation, x is the instance to be classified into class y_j. So $p(y_j|x)$ is the probability of instance x being in class y_j; $p(y_j|x)$ is the probability of generating instance x, given class y_j; $p(y_j)$ is the probability of occurrence of class y_j; and (x) is the probability of instance x

occurring. NBC is robust with missing values and has a short computational time for training. In addition, since the model has the form of a product, it can be converted into a sum through the use of logarithms with significant computational advantage.

3.3.4 Decision Tree Classifier

The decision tree approach involves arranging the features of a dataset into a hierarchical tree structure as shown in figure 3.6. Each node in a decision tree represents a feature in an instance to be classified, and each branch represents a value that the node can assume. Instances are classified starting at the root node and sorted based on their feature values. The tree is essentially a set of questions through which an instance is passed. The feature node queries what value the instance has for that particular feature. The branch taken next is dependent on the answer, which then leads to another question or a final classification (called a leaf node), when no more questions are asked. It should be noted that a particular feature, or attribute value, or leaf node can appear more than once within the tree. Click on the image for a hyperlink to the source of the publication.

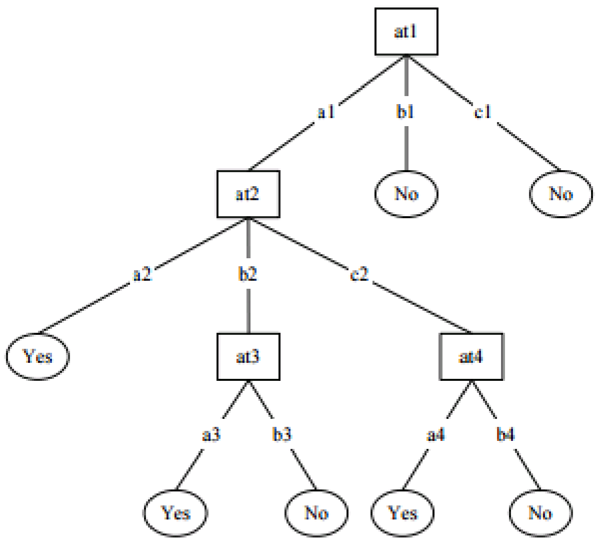


Figure 3.6: Decision tree (Sotiris 2007)

The method for finding the best nodes that separate the data is repeated for each node, until all sub-trees are created, and all the training data has been classified. The criteria for splitting may be based on information gain, purity or the Fisher Criterion (37steps 2013). To avoid overfitting of training data, the training algorithm can be stopped before it reaches this point, or the tree can be pruned-back. A stopping algorithm can be used for the former, which involves termination conditions, such as a threshold test for feature quality metrics. Elsewhere, in the post-pruning method, a check of the tree's performance is made and if necessary, pruned. The default is to have no pruning, but this is one of five pruning methods that can be selected. The number of branches coming out of each node is termed the branching factor or branching ratio, B. The branching factor B is normally two, which would make such trees binary trees. A binary tree can always be constructed by choosing the correct feature nodes and values.

3.3 Using Machine Learning to Classify Term and Preterm Records

Alaskar et al. (Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013) proposed a neural network that builds on the back-propagation network, called the self-organized layer inspired by immune algorithm (SONIA), to classify both term and preterm labour using EHG signals. The algorithm improves recognition and generalization in the back-propagation learning algorithm and produced an accuracy of 70.82% compared with other similar classification techniques.

While (SMS Baghamoradi 2011) used the TPEHG database (Fele-Žorž et al. 2008) to evaluate classification accuracy using sample entropy and thirty cepstral coefficients. The thirty cepstral coefficients were calculated from each signal recording and used in one feature set for classification, whilst for the second feature set was reduced to three cepstral coefficients only. The three cepstral coefficients were chosen by sequential forward selection and Fisher's discriminant. A multi-layer perceptron neural network was used to perform classification of the records into term and preterm records. It was found that the three cepstral coefficients gave the best classification accuracy of 72.73% (±13.5), as opposed to 53.11% (±10.5) for the full thirty coefficients, and 51.67% (±14.6) using sample entropy only. However, since the full thirty coefficients only presented a small improvement in classification accuracy, it is likely that the sequential forward selection and Fisher's discriminant had the most effect on the accuracy.

The k-NN algorithm was used by (Moslem Bassam, Diab Mohamad, Khalil Mohamad 2012) with an emphasis on Autoregressive (AR) modelling and wavelet transformation pre-processing techniques. The study focused on classifying contractions into three types using data obtained from 16 women. Group 1 (G1), were women who had their contractions recorded at 29 weeks, and then delivered at 33 weeks; Group 2 (G2) were also recorded at 29 weeks, but delivered at 31 weeks, and Group 3 (G3) were recorded at 27 weeks and delivered at 31 weeks. Using the k-NN algorithm, combined with the pre-processing method of AR, classification occurred against G1 and G2 and against G2 and G3. As well as this, an Unsupervised Statistical Classification Method (USCM), combined with Wavelet Transformation, was also used. The USCM adopted the Fisher Test and k-Means methods. The wavelet transformation, combined with USCM, provided a classification error of 9.5%, when discerning G1 against G2, and 13.8% when classifying G2 against G3. Using AR, k-NN provided a classification error of 2.4% for G1 against G2 and 8.3% for G2 against G3. In both classifications, the AR and k-NN methods performed better than the USCM. Furthermore, the classification accuracy for G1 and G2 was always lower than the equivalent G2 and G3 classifications. This suggests that it is easier to distinguish between pregnancies recorded at different stages of gestation than it is to distinguish between the times of delivery.

Support Vector Machines (SVM) have also been successfully used to classify term and preterm deliveries (Moslem Bassam, Khalil Mohamad, Diab Mohamad, Chkeir Aly 2011).

This classifier classifies contractions as either labour or non-labour, using different locations on the abdomen. Majority Voting (MV) decision fusion rules, including a Gaussian Radial Basis Function (GRBF), form the basis for classification. The feature vectors include the power of the EHG signal, and the median frequency. The support vector machine shows some promising results. For example, Moslem Bassam et al. (Moslem Bassam, Khalil Mohamad, Diab Mohamad, Chkeir Aly 2011) used a single SVM classifier, at one particular location on the abdomen The result indicated a 78.4% accuracy – the overall classification accuracy, for the combined SVM, was 88.4%. Finding the coefficients, for the decision boundary, occurs by solving a quadratic optimization problem.

Ren et al. (Ren et al. 2015) used the same dataset as Fele-Žorž et al. (Fele-Žorž et al. 2008). In this study, a new analytical approach was used for assessing the risk of preterm delivery, using EMG recordings. This method first employed Empirical Mode Decomposition (EMD) to obtain Intrinsic Mode Functions (IMF). The entropy values of both instantaneous amplitude and instantaneous frequency of the first ten IMF components are computed in order to derive ratios of these two distinct components as features. The accuracy of this approach was then compared to the proposed six different classifiers. These classifiers included an SVM), Random Forests (RF), MLP, AdaBoost (AB), Bayesian Network (BN) and Simple Logistic Regression (SLR). Three different electrode positions were then analysed for their prediction accuracy of preterm delivery in order to establish which uterine EMG recording location produced the optimal signal data. Ren et al. (Ren et al. 2015) illustrates a clear improvement in prediction accuracy of preterm delivery risk compared with previous approaches. Their results achieved an impressive maximum AUC value of 0.986 using signals from an electrode positioned below the nave.

3.4 Summary

This chapter provides a details discussion on machine learning algorithms. These were split into two major groups – artificial neural networks and those not inspired by the biological processes of the human brain. Several artificial neural networking strategies were discussed and the merits of each were outlined. Other non-ANN approaches were also presented and included a discussion on SVMs, Random Forests and Naïve Bayes algorithms. Combining the theoretical background of machine learning algorithms their application to EHG term and preterm record classification was also discussed. Here the state of the art works were discussed, which hare used in direct comparison with the results presented later in this thesis. In the following chapter we build on this background literature and describe the methodology adopted within this thesis.

Chapter 4 Proposed Methodology

4.1 Introduction

Machine learning systems follow a common set of principle operations, as illustrated in Figure 4.1. We briefly identify and clarify such principles to serve as an implementation agnostic view of the essential elements of machine learning. The usefulness of learning assumes a real world problem of interest, in the form of a system or phenomenon of interest, which can be measured to provide some form of observation. Such real world problem domains therefore act as a source of a finite observation, which can be used to estimate a reconstruction of systems for which the observations relate. The information encompassed can be operated upon inductively to construct a hypothesis of the underlying (presumed unknown) function, which acts to provide a generalised reflection of the original problem. Given the possession of such a model, useful estimates of the real world process can be realised using the model, in order to fulfil the requirements of a chosen application within a real world domain.

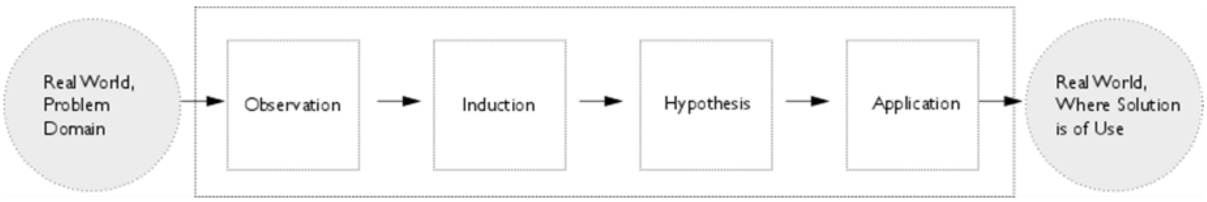


Figure 4.1: The process of machine learning

This chapter discusses the TPEHG dataset (PhysioNet 2012a), which contains the raw EHG signals necessary for our study. This also includes a detailed discussion on the data pre-processed tasks used, data segmentation, feature extraction, single classification techniques, stacked classifiers using bagging methods and several validation techniques to determine the overall accuracy of our experiments. Before these aspects are discussed further the following subsection presents a proposed methodological framework, which describes the different phases of our analyses of EHG signals.

4.2 Proposed Framework Architecture

In order to conduct our experiments using the TPEHG dataset, the proposed methodological framework is presented in Figure 4.2. These phases consist of raw EHG signals (data collection), signal pre-processing, feature extraction, oversampling with the synthetic minority over-sampling technique (SMOTE), generating test and training models, feature selection, classification, combining classifiers, validation, and the presentation of results. The remainder of the chapter will provide a more in-depth discussion of each of these processes within the proposed methodological framework.

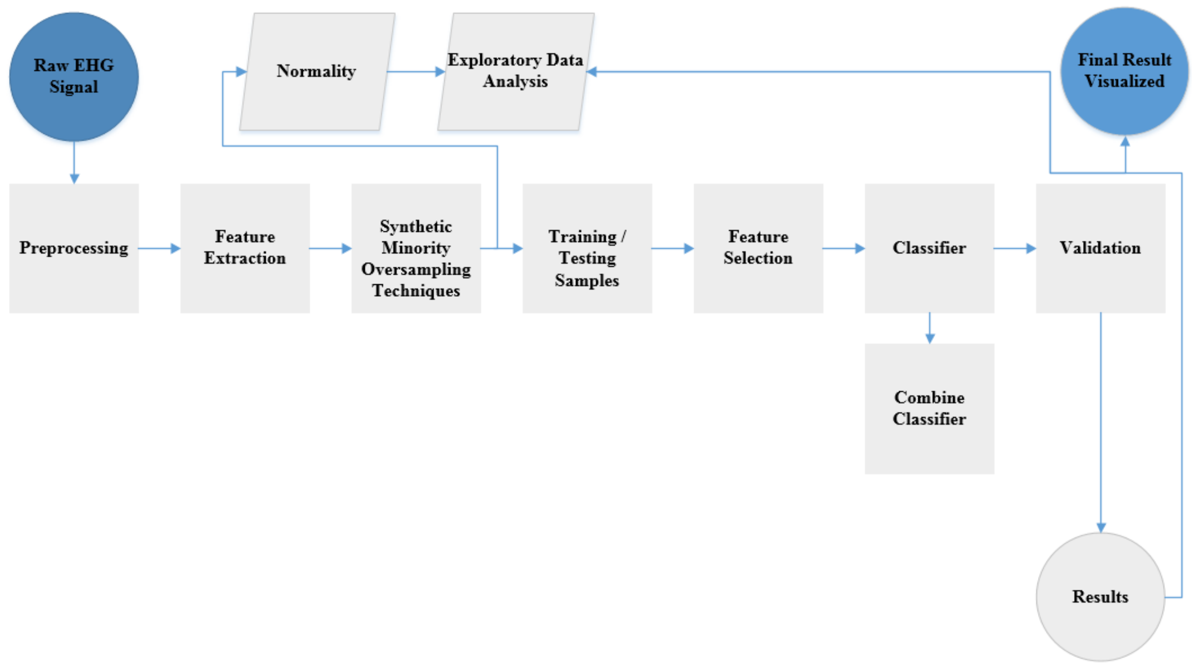


Figure 4.2: Methodology Phases

4.3 Data Collection

Several studies (Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013; Ye-Lin et al. 2014; Garcia-Gonzalez et al. 2013) have shown that the EHG signal may vary from woman to woman, depending on whether she is in true labour or false labour and whether she will deliver term or preterm. EHG provides a strong basis for objective classification and diagnosis of preterm birth. Many research studies have used EHG for classification or detection of true labour but not for the classification of term and preterm deliveries (Fele-Žorž et al. 2008; de Lau et al. 2014; SMS Baghamoradi 2011).

The EHG signal is affected by interference or noise, as the electrical activity from the uterus, that is detected by the electrodes, is very small, in the range of μV to 1mV (Carré et al. 1998). The type of interference affecting EHG signals include other electrical or physical interference from the mother or baby, such as the mother's physical movements, breathing, maternal cardiac activity, and foetal movements (Marque et al. 2007).

Other factors, which cannot be so easily cancelled out, are the physical differences between patients; the strength of the signal collected may be influenced by the varying amounts of fat or salinity level of skin that each patient has (Maner 2003). These types of factors are not easy to deal with, since they vary from patient to patient, and there appears to be no standardised way of compensating for these factors.

In this study we utilise the TPEHG database. The TPEHG records were collected from a general population of pregnant patients at the Department of Obstetrics and Gynaecology Medical Centre in Ljumljana, between 1997 and 2006. These records are publically available, via the TPEHG database, in Physionet. Records were collected from the general population of pregnant patients, as well as those admitted to the hospital with diagnosed preterm labour. In the TPEHG database, there are 300 records/recordings - one record per pregnancy. Each recording was approximately 30 minutes long, had a sampling frequency (fs) of 20Hz, and had a 16-bit resolution with an amplitude range ±2.5mV. Prior to sampling, the signals were passed through an analogue three-pole Butterworth filter, in the range of 0 to 5Hz. Records were either recorded early, <26 weeks (at around 23 weeks of gestation) or later, =>26 weeks (at around 31 weeks). In this experiment, linear and non-linear methods are used in both time and frequency domains, to improve the results obtained from classification algorithms.

4.4 Raw EHG Signals

In this research, the TPEHG dataset from Physionet is utilised to demonstrate the applicability of our approach (Fele-Žorž et al. 2008). Data was collected from four electrodes attached to the abdominal surface, with the navel at the symmetrical centre. Three signals were obtained simultaneously, per 'record', by recording through three different channels – Channel 1, Channel 2 and Channel 3. As shown in Figure 4.3 the Channel 1 signal was measured between E2 – E1, Channel 2 between E2 – E3 and Channel 3 between E4 – E3.

The EHG signals were recorded using four bipolar electrodes adhered to the abdominal surface and spaced at a horizontal, vertical, distance of 2.5cm to 7cm apart with the black circles representing the electrodes. The total number of records in the EHG dataset is 300 (38 preterm records and 262 term records). Each of the records were either recorded early, <26 weeks (at around 23 weeks of gestation) or later, =>26 weeks (at around 31 weeks).

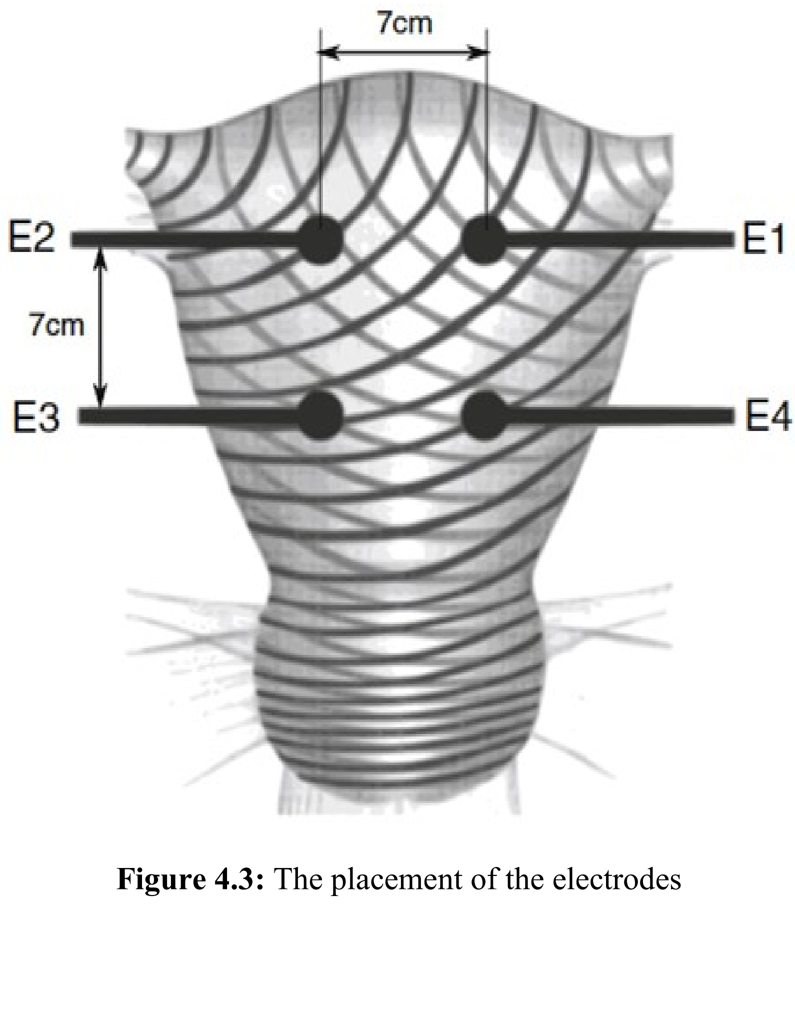


Figure 4.3: The placement of the electrodes

In this thesis only channel 3 was utilised based on the findings produced in previous studies that demonstrated the third signal gave the best discriminative capacity when classifying term and preterm records (Hussain et al. 2015; Fele-Zorz et al. 2008). Table 4.1 provides a detailed description of the TPEHG dataset.

Table 4.1: TPEHG Term and Preterm Dataset

Data	**Term Deliveries**		**Preterm Deliveries**		**All Deliveries**	**All Deliveries**
Recording Time	**# of Records**	**Mean/ Median Recording Weeks**	**# of Records**	**Mean/ Median Recording Weeks**	**# of Records**	**Mean/ Median Recording Weeks**
Early	**143**	**22.7/22.86**	**19**	**23.0/23.43**	**162**	**22.73/23.0**
Later	**119**	**30.8/31.14**	**19**	**30.2/30.86**	**138**	**30.71/31.14**
Total	**262**	**26.75/24.36**	**38**	**27.0/25.86**	**300**	**26.78/24.43**

Table 4.1 show the mean and median of the recoding week of the deliveries for both term and preterm deliveries. The recording time relates to the gestational age of the foetus, at the time the recording was made. The classification of these recordings as term or preterm deliveries were made retrospectively, after birth and followed the widely-used definition of preterm birth being under a fully-completed 37 weeks.

The four categories of recordings were therefore as follows:

- Early – Term: Recordings made early, which resulted in a term delivery
- Early – Preterm: Recordings made early, which resulted in a preterm delivery
- Late - Term: Recordings made late, which resulted in a term delivery
- Late – Preterm: Recordings made late, which resulted in a preterm delivery

In order to visually show the separation between term and preterm births in the dataset Figure 4.4 and 4.5 illustrate 2-D scatterplots for the Principle component and Stochastic Proximity Embedding algorithms.

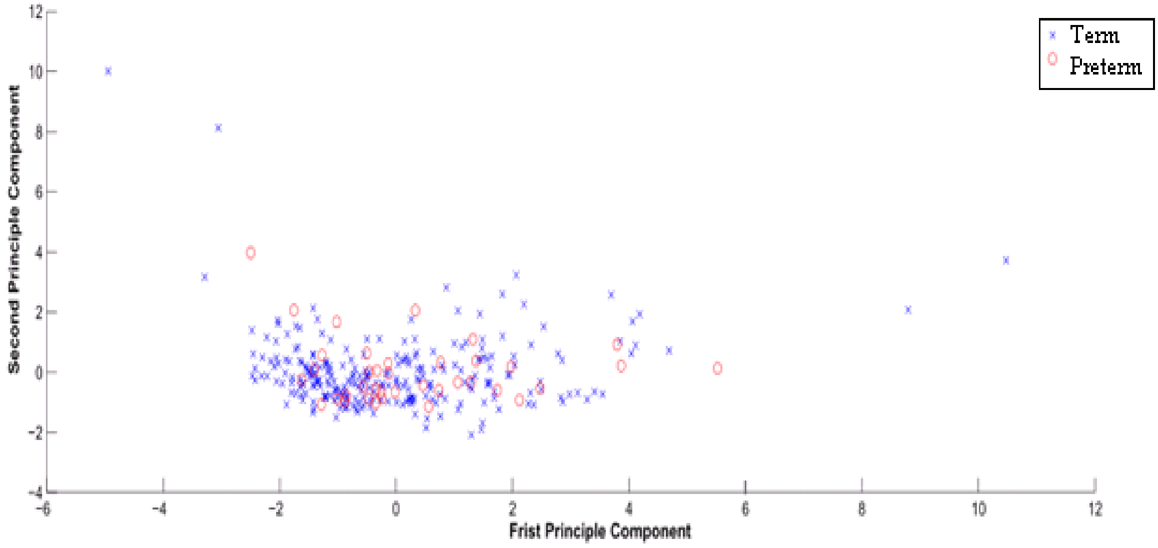


Figure 4.4: PCA for TPEHG Channel 3 – 0.34 Hz – 1 Hz Filtered Signal

This is useful to show the feature space mapping of both linear and nonlinear dimensions. In the Principle Component (PCA) and Stochastic Proximity Embedding (SPE) plots, in Figures 4.4 – 4.5, the blue crosses show term records, whilst the red circles represent preterm records.

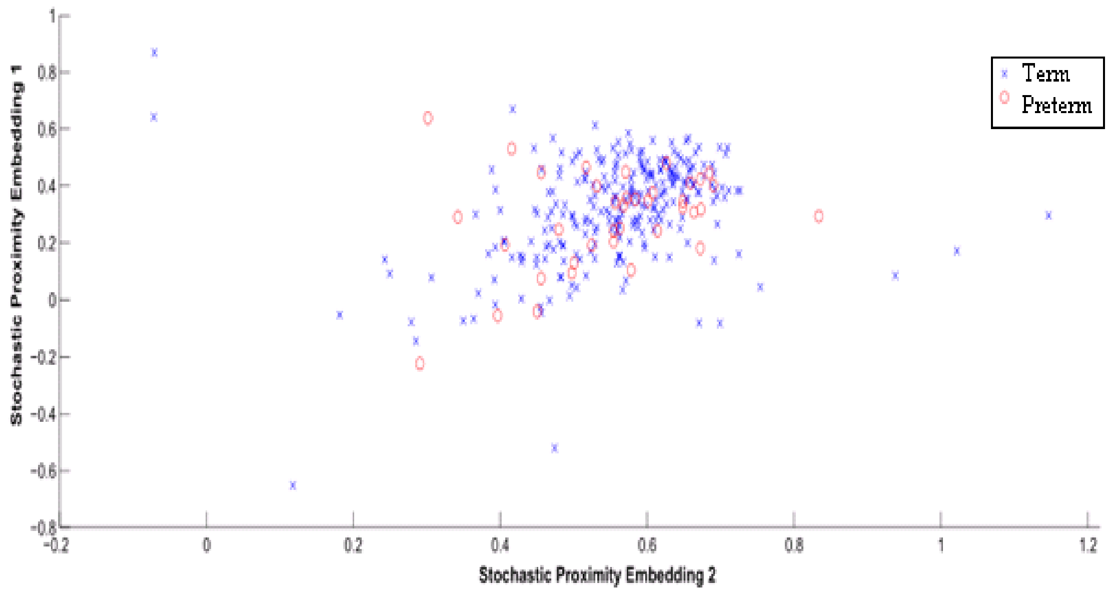


Figure 4.5: SPE for TPEHG Channel 3 – 0.34 Hz – 1 Hz Filtered Signal

In both plots, it shows the decision boundary of the two classes linearly (PCA) and nonlinearly (SPE). The results demonstrate that the data is not very well separable due to two distinct clusters in the middle of the patterns overlapping with each other.

Figure 4.6 shows a histogram of the two classes which clearly indicates that the distribution is significantly skewed in favour of the term records (term=262, preterm=38. This is known to cause significant problems in machine learning tasks and requires oversampling. This stage often follows the feature extraction stage which will be discussed later in this chapter.

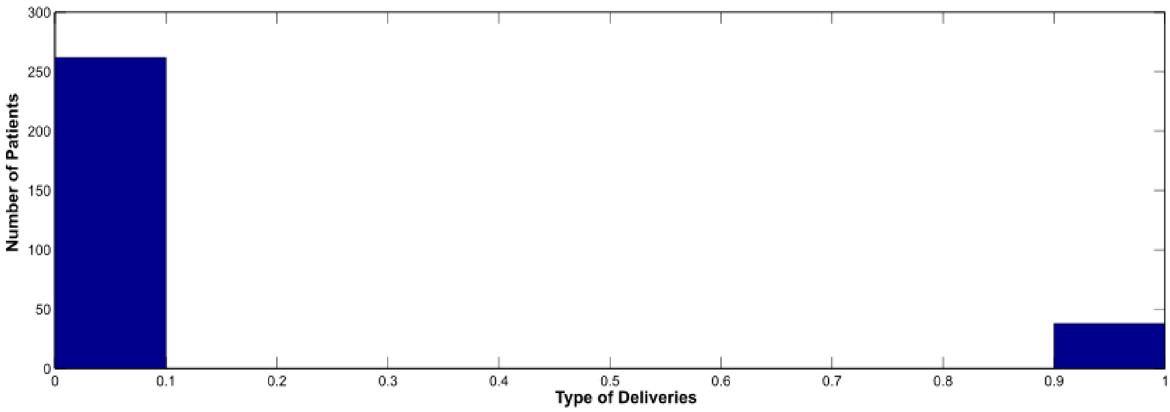


Figure 4.6: Distribution of deliveries in TPEHG Dataset

4.5 Pre-processing of EHG Signals

The collection of raw EHG signals is always temporal and so before analysis the raw data needs to be filtered. As raw recordings often contain unwanted noise, pre-processing is essential and often includes filtering and artefact removal. Each record in the dataset is approximately 30 minutes long, with a sample frequency of 20Hz, and a 16-bit resolution, with an amplitude range ± 2.5mV. A zero bandpass filter was created in Matlab, using the Filter Design & Analysis tool (FDAT) that is available in the Signal Processing Toolbox. The response type was 'Bandpass', and 'IIR Butterworth' was chosen as the Design Method. The order was set to eight, in an attempt to reduce the amplification effects. In this mode, where the order was explicitly specified, the attenuation of the signal magnitude was set at 3db. The sampling frequency (Fs) of 20Hz was inputted, since the data in the supplied TPEHG dataset was sampled at this rate. The pass band frequencies for Fc1 and Fc2 were 0.34 and 1 respectively.

In (Léman et al. 1999) EHG signals have been passed through various other Butterworth filter configurations that include 0.8-4Hz, 0.3-4Hz and 0.3-3Hz. The reason why channel 3 was selected is because Fele-Žorž et al. (Fele-Žorž et al. 2008), showed that the 0.3-3Hz filtered signals on channel 3 discriminated between preterm and term delivery records than the others. However, there was no appropriate filter to remove unwanted artefacts, such as maternal heart rate which is known to affect performance.

Uterine activity has been found to comprise both 'fast' and 'slow' signals. The fast waves represent the individual electrical signals firing, whilst the slow waves correspond to the resulting mechanical contractions. Slow waves exist between 0.03 and 0.3 Hz, and the fast waves exist between 0.3 and 3.0 Hz. Catalin et al. (Catalin, Buhimschi 1996) found in a study of 99 pregnant patients, that 98% of uterine electrical activity occurred in frequencies less than 1Hz, and that the maternal heart rate (ECG) was always higher than 1 Hz. Furthermore, 95% of the patients, measured had respiration rates of 0.33 Hz or less. Several other studies have adopted the same filtering scheme (Tong et al. 2011; Bassam Moslem et al. 2011). Therefore, in this thesis, the raw Channel 3 signal was chosen and filtered using a 0.34–1 Hz filter. This is to coincide with the findings of both (Fele-Žorž et al. 2008) and (Catalin, Buhimschi 1996).

Filtering in one direction was done with the filter function. The plot for one particular signal, $x873$ is shown below in Figure 4.7. The plot shows three different signals, one for a 0.3-3Hz filtered signal, which was supplied with the dataset, and the other two filtered with a 0.34-1Hz filter – one single-directional, the other zero-phased.

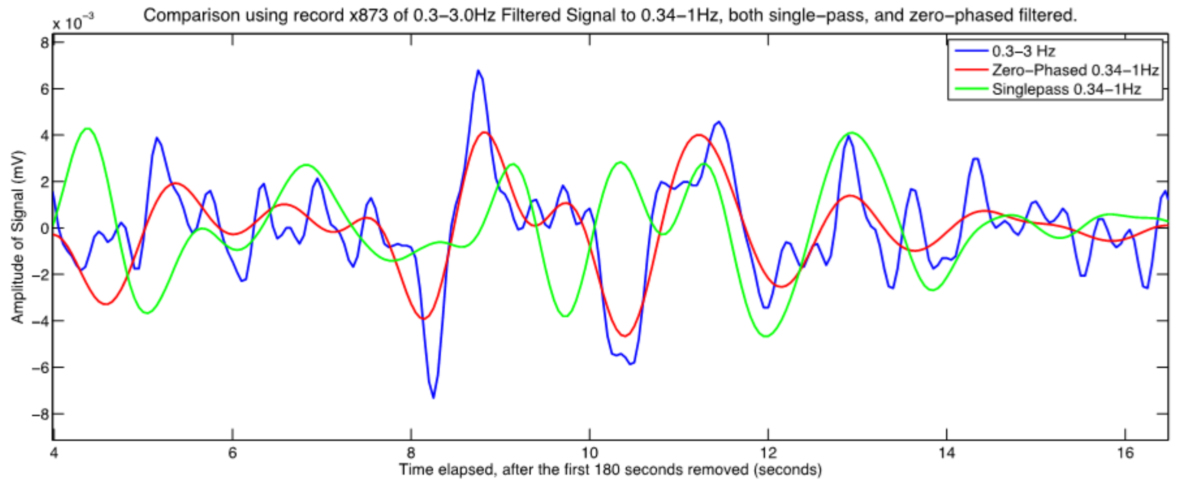


Figure 4.7: Record x873 Filtered Using three Different Methods

The plot shows two things that the 0.34-1Hz filter reduces the overall amplitude or peaks of the signal in comparison to the 0.3-3Hz filtered signal. The peaks are also smoothed out; there are less peaks and troughs in general. Figure 4.7 also shows that zero phase filtering puts the phase more in line with the 0.3-3Hz filtered signal, rather than the single-phase filter. In short, the zero-phase filtering appeared to be an appropriate filter strategy.

4.6 Feature Extraction from EHG Signals

Feature extraction transforms raw signals into more informative signatures that can assist in grouping different classes. In other words, features are synonymous of input variables or the attributes of a dataset that provide a good representation of a specific domain, related to the available measurement. In this thesis, several feature extraction techniques have been utilized (Phinyomark et al. 2009), (Fele-Žorž et al. 2008) and (Phinyomark, A. Nuidod, P.Phukpattaranont 2012). These are applied to the channel 3 records using the filter parameters previously discussed. Table 2, provides a summary of the mathematical proofs for each of the features used.

Table 4.2: Feature Extraction Techniques used in EMG

Equation Name	Equation Abbreviations
Integrated EHG	$\mathrm{IEMG} = \sum_{n=1}^{N} \lvert(x_n)\rvert$
Mean Absolute Value of EHG	$\mathrm{MAV} = \frac{1}{N}\sum_{n=0}^{N} \lvert x_n \rvert$

Simple Square Integral of EHG	$SI = \sum_{n=0}^{N} \lvert x_n \rvert^2$
Wavelet length of EHG Signal	$WL = \sum_{n=0}^{N-1} \lvert x_n - x_{n-1} \rvert$
Log Detector of EHG Signal	$LOG = e^{1/N \sum_{n=1}^{N} \log(\lvert x_n \rvert)}$
Root Mean Square of EHG Signal	$RMS = \sqrt{1/N \sum_{n=1}^{N} x_n^2}$
Variance of EHG	$VAR = \frac{1}{N} - 1 \sum_{n=1}^{N} x_n^2$
Difference Absolute Standard Deviation Value	$DAS = {}^{1}/_{N-1} \sum_{n=1}^{N-1} (x_{n+1} - x_n)^2$
Maximum Fractal Length of EHG Signal	$MFL = \log 10(\sqrt{\sum_{n=1}^{N-1} (x_n - x_{n+1})^2})$
Average Amplitude Change of EHG Signal	$AAC = \frac{1}{N} \sum_{n=1}^{N-1} \lvert x_{n+1} - x_n \rvert$
Peak Frequency of EHG Signal	$fmax = \arg(\frac{f_s}{N} \max_{i=0}^{N-1} P(i))$
Median Frequency	$fmed = i_m \frac{f_s}{N}, \quad \sum_{i=0}^{i=i_m} P(i) \doteq \sum_{i=i_m}^{i=N-1} P(i)$
Sample Entropy	$sampEn_{m,r}(y) = \begin{cases} -\log\left(\frac{c_m}{c_{(m-1)}}\right) & c_m \neq 0 \wedge c_{m-1} \neq 0 \\ -\log\left(\frac{N-m}{N-m-1}\right) & c_m = 0 \vee c_m = 0 \end{cases}$

In this list, x_n represents the n^{th} sample in the EHG signals in the segment; P represents the power spectrum (calculated using the Fast Discrete Fourier Transform), while N denotes the length of the EHG signal. The main difference between our work and (Phinyomark, A. Nuidod, P.Phukpattaranont 2012), (Phinyomark et al. 2009) is in the analysis of the electrical activity in the uterus, rather than in other muscle activity. Given that the uterus is a muscle, this study investigates whether techniques used to capture EMG activity can also be used in EHG analysis. It has been recognised that since there is some overlap between uterine electrical signals and noise and that the noise may also be non-linear, then more advanced signal processing methods may be required for analysis. Accordingly, we applied the feature definitions given in Table 2 to channel 3 of the EHG signals in the TPEHG database, yielding feature vectors that contain 13 features for each signal.

The feature extraction equations from Table 4.2 and initial exploratory data analysis visualisations are shown in figure 4.8 – 4.10. Using QQ plots, box plots, histograms and kernel density estimation plots, statistical analysis of the initial 300 TPEHG records are shown.

These set of EDS diagrams show the distribution of all the features. The names of the features in the plots below are abbreviated according to the following, Root Mean Square (RMS) Peak Frequency (FPeak), Median Frequency (FMed), Sample Entropy (Samp en), Integrated EMG (IEMG), Mean Absolute Value (MAV), Simple Square Integral (SSI), Wavelet Length (WL), Variance (V), Difference Absolute Standard Deviation Value (DASDV), Log Detector (LD), Average Amplitude Change (AAC), and Maximum Fractal Length (MFL).

Figure 4.8 shows the QQ plot quantiles of the first data set against the quantiles of the second data set. The red dotted lined shows that the best distribution fit for each feature set extracted from the TPEHG dataset. The QQ plots show outliers in RMS, FMed, FPeak, V, IEMG, MAV, SSI, AAC, MFL, LD, Samp en ,WL and DASDV where points deviate from the red reference line. The identified extreme outliers are unwanted artefacts and were removed from the feature space.

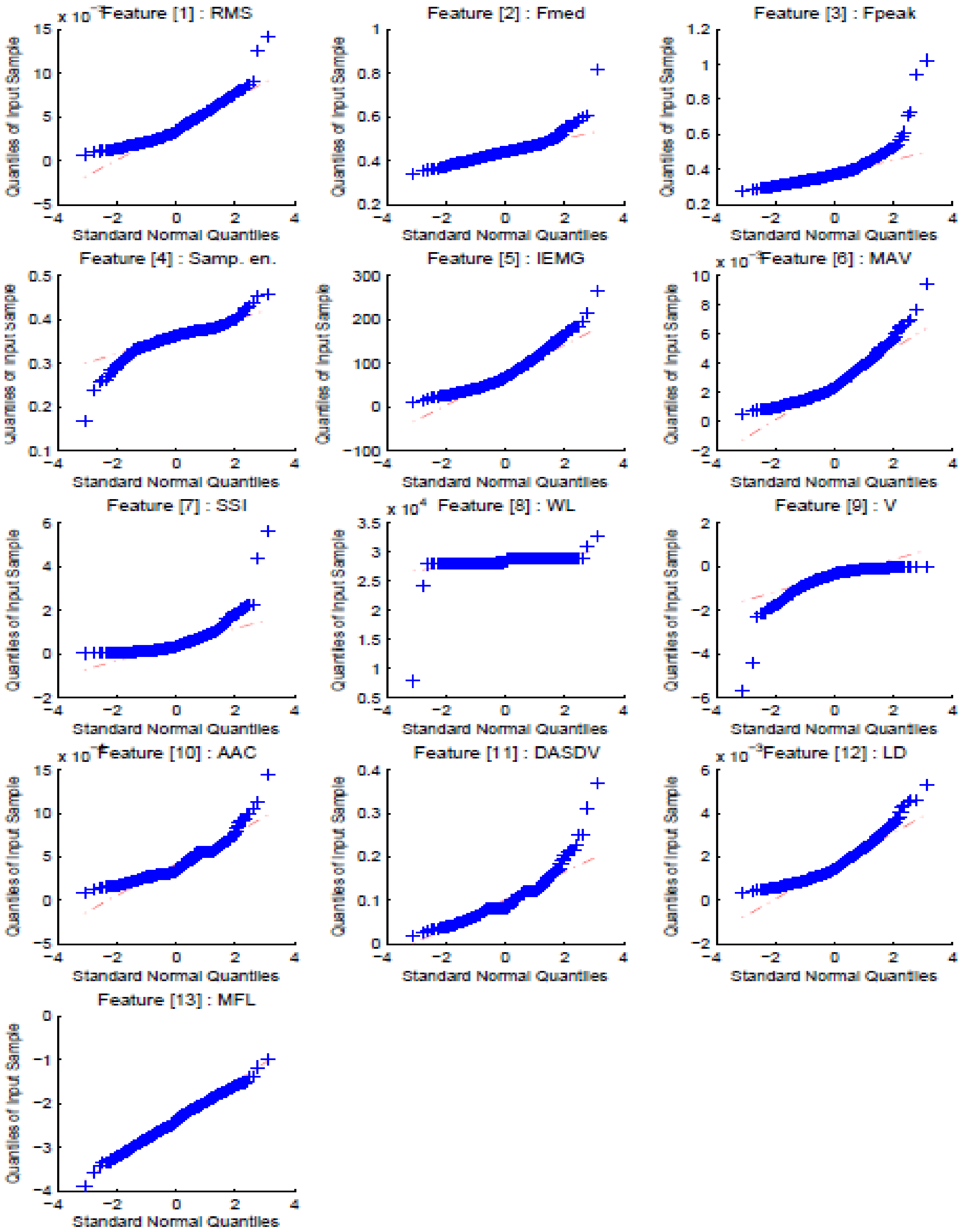


Figure 4.8: QQ plot of 300 TPEHG Feature Extractions

In Figure 4.9, it shows the distribution of all thirteen features extracted from the TPEHG records. The outcome results in both figure 4.9 and 4.10 correlate, because the data distributions for MAV, AAC and IEMG do not fit expected normal distributions, confirming the non-normality of the data, and the output of the Lilliefors tests. Since the minority of the features did not meet normality assumptions, it is improbable that the class densities would either.

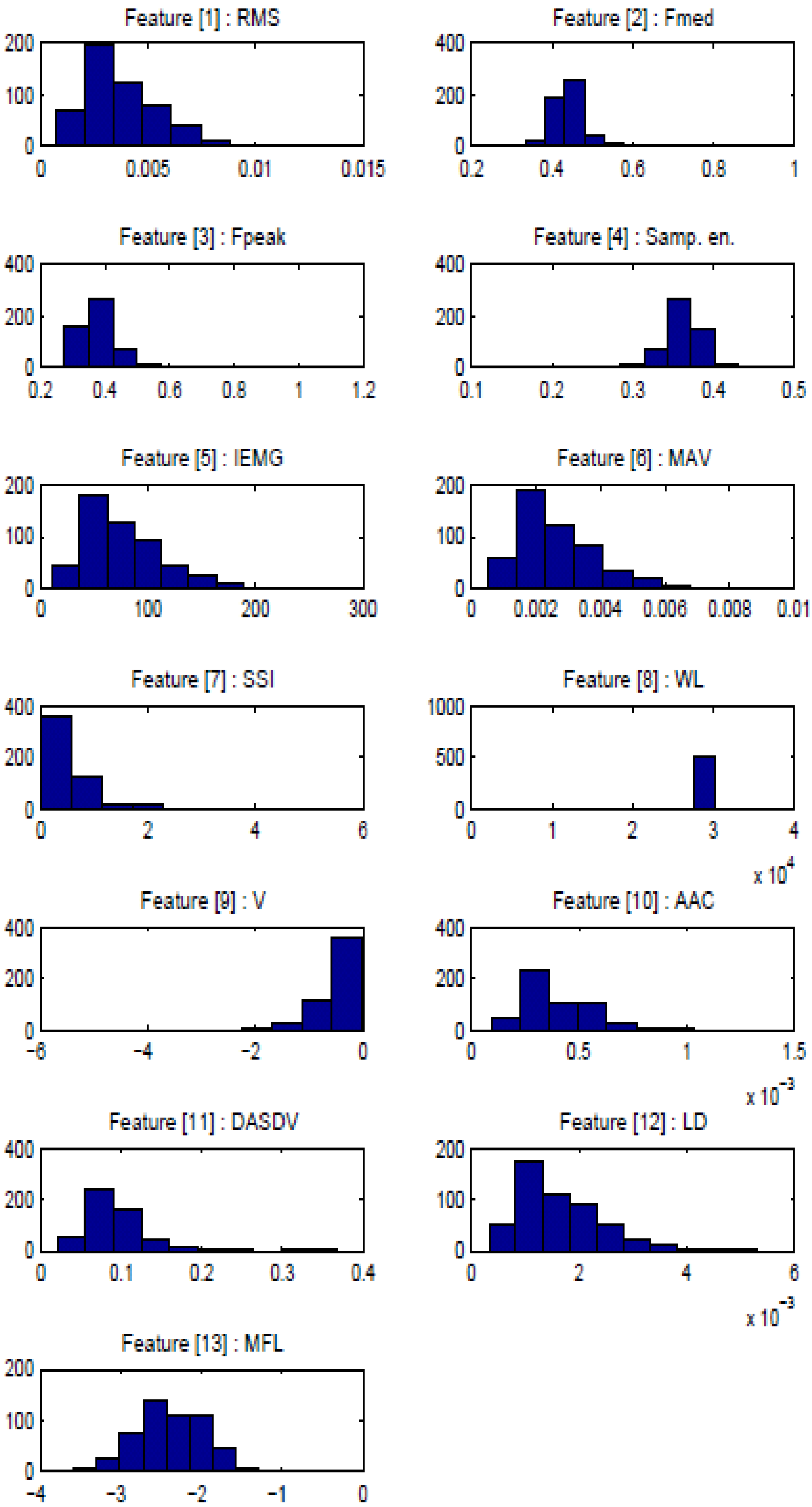


Figure 4.9: Histogram of 300 TPEHG Feature Extractions

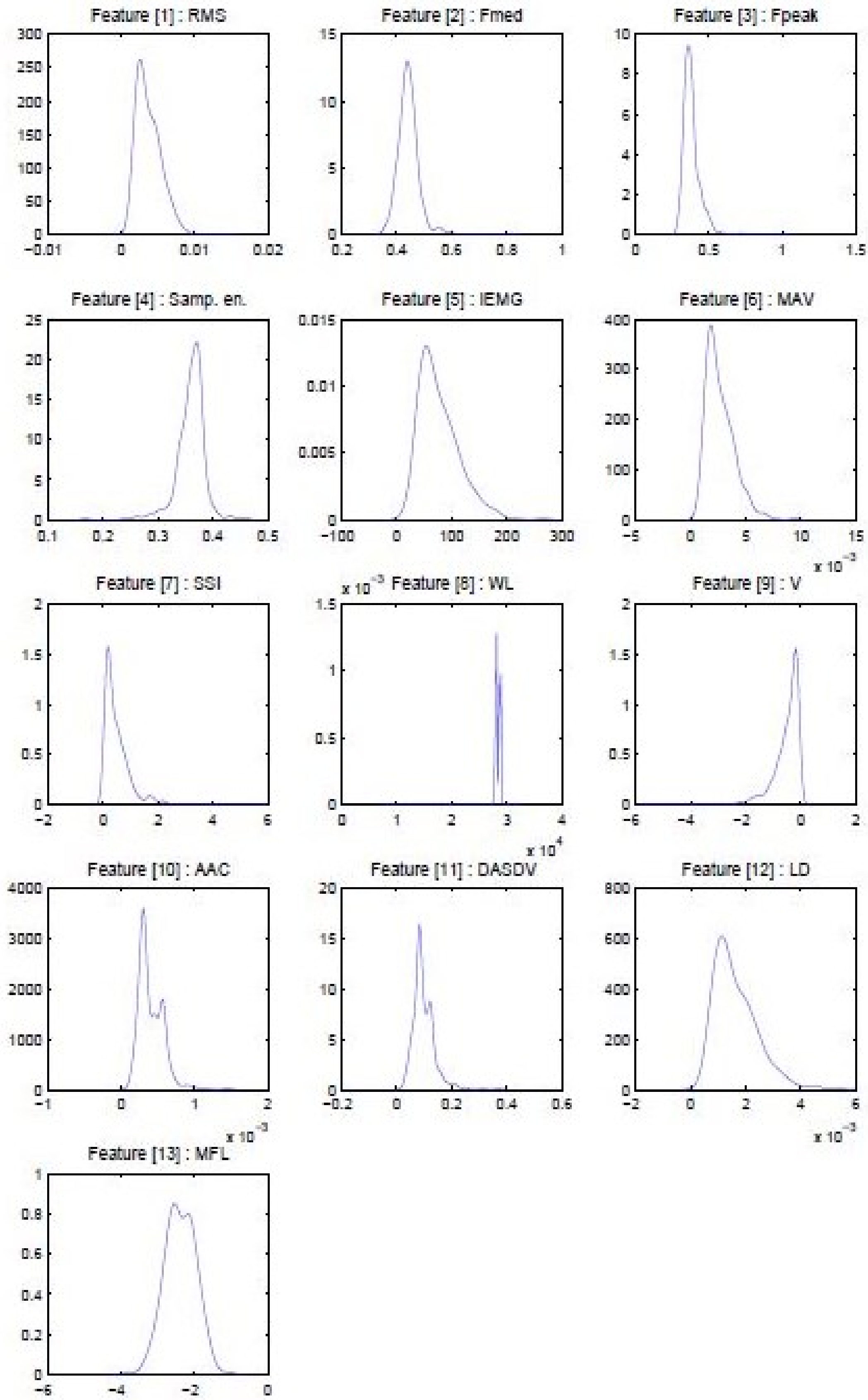


Figure 4.10: Kernel Density Estimation Plot of 300 TPEHG Feature Extractions

This is also show in the box plot in Figure 4.11. The results suggest a crossover for all of the features. Consequently, a feature selection strategy is needed to determine the discriminate capabilities of all features and suitable modelling strategies for feature subsets. This is discussed further in section 4.9.

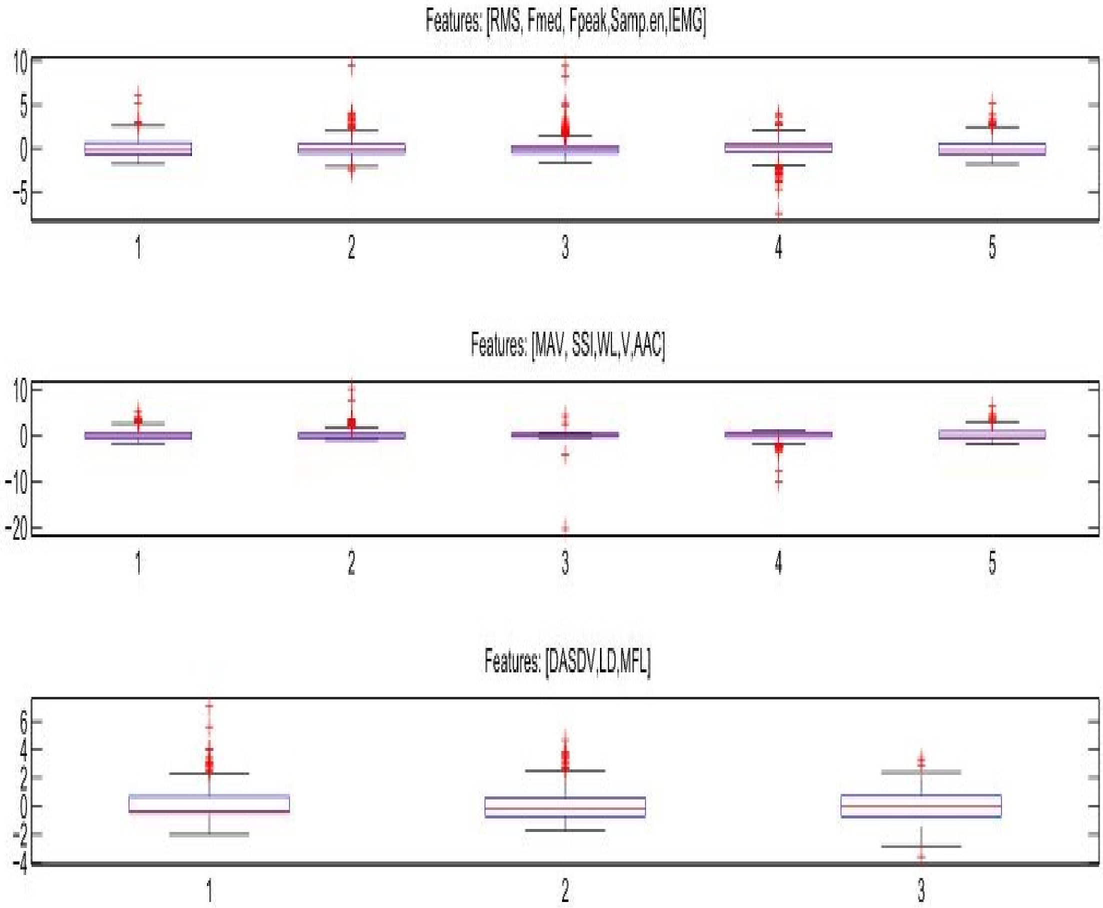


Figure 4.11: Box Plot of 300 TPEHG Feature Extractions

4.7 Synthetic Minority Over-Sampling Technique (SMOTE)

The dataset contains 262 samples for mothers who delivered at term and 38 that delivered prematurely. This shows that the preterm class has significantly less records than the term class. This poses significant problems in machine learning classification tasks (bias in the term class); therefore, the preterm EHG features need to be oversampled. This is important because an imbalanced dataset can cause ineffective classification, since the classifier does not have enough of the preterm class to learn from. Therefore, given that there are more term records, the probability of detecting a preterm record is low. To address this issue, the minority class (preterm) is oversampled using the Synthetic Minority Over-Sampling Technique (SMOTE). This technique is effective in solving class skew problems (Richman & Moorman 2000) and has been extensively utilised with the biomedical research community. Therefore, using the 38 preterm records, SMOTE is used to generate an additional 224 preterm records. This is indicated in figure 4.12 (preterm=262).

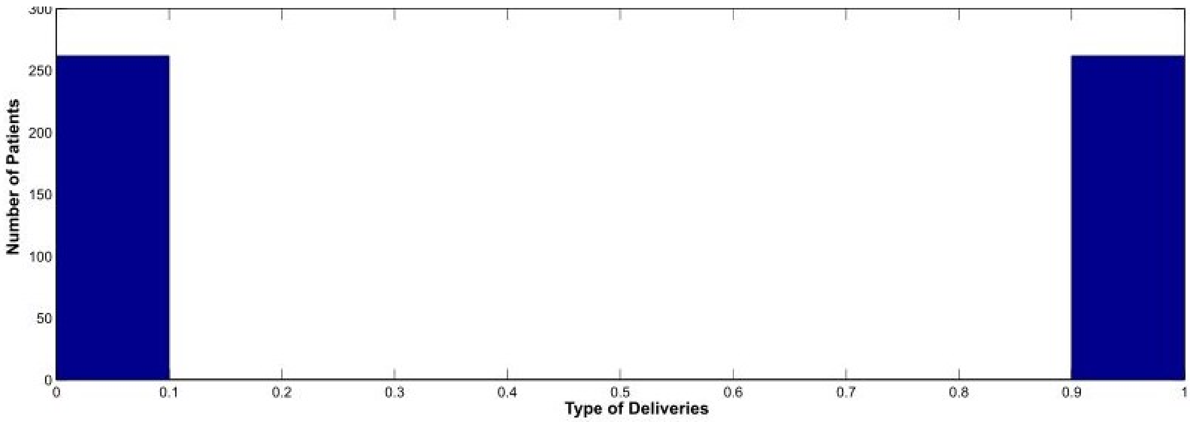


Figure 4.12: Histogram Oversample of TPEHG Dataset

4.8 Data Normalisation

Normalisation is required to manage bias and scale in the feature set. The size of the range over which a feature is represented is not generally an indicator of its importance relative to other features, meaning a difference in scale is of no consequence and can be discarded. The motivation for normalisation is to ensure a consistent standard of data is presented to the learning component, such that the learning process has as few outside factors to contend with as possible. The Min-Max scaling approach has been used to normalise our dataset. In this approach, the data is scaled to a fixed range - usually 0 to 1. The mathematical representation of this algorithm is shown below in equation 4.1:

$$X'_i = \frac{X_i - X_{minval}}{X_{maxval} - X_{minval}} . \left(r_{upper} - r_{lower}\right) + r_{lower,} \tag{4.1}$$

In this equation, X'_i is the transformed value, X_i is the raw value, X_{maxval} and X_{minval} are the original ranges of the feature, and $r_{upper} - r_{lower}$ denotes the target range.

We applied the above definition on an individual basis to each feature vector extracted from our dataset, yielding a final feature matrix with values ranging within 0 to 1.0.

4.9 Feature Selection

Using the features defined in Table 2, feature vectors have been generated. The literature reports that peak frequency, median frequency, sample entropy and root mean square have the most potential to discriminate between term and preterm records. Furthermore, the literature also reports that in EMG studies, the features described in Table 2 are equally as good at discriminating between different muscle activities. However, there is no mention of the uterus in many studies on EMG. To validate these findings, the discriminate capabilities of all the features reported in Table 2 (i.e. feature ranking) have been determined. This is achieved using several measures, including statistical significance, linear discriminant analysis using independent search (LDAi), linear discriminant analysis using forward search (LDAf), linear discriminant analysis using backward search (LDAb) and linear discriminant analysis using branch and bound search. At any node of the tree, the algorithm must make a finite decision and set one of the unbound variables. Using these measures, the features will be ranked, and the top uncorrelated features will be selected from the feature space. These features will be used in the classification stage to determine which set produced the greatest area under the curve (AUC), sensitivity and specificity values.

4.10 Classification

There has not been a consensus on the best classifier for all data domains. The choice of classifier depends on the dataset to some extent. The selection of a suitable classifier still involves trial-and-error processes. However, statistical validation can be used to guide this process (Domingos 2012; Elkan 2010). Although the choice of algorithm does always depend on the task at hand, the review of the classification methods highlights which ones might be most promising to classify term and preterm deliveries.

In this thesis eight advanced artificial neural network (AANN) classifiers and four non-AANN architectures are evaluated. These are the BPXNC, LMNC, PERLC, RBNC, RNNC, VPC, DRBMC, and the FLNN. Other types of classifiers that are not neural networks include the RFC, SVMC, NBC, and the DTC.

4.11 Single Classifier Framework

In our approach, two different classifier architectures are considered in isolation to evaluate which classifiers perform better. Each of these and the key configurations are described in Table 4.3.

Table 4.3: Classification Models Description

Model Designation	Description	Architecture	Training Algorithm	Role
LEVNN	Multilayer Perceptron, Trained using the Levenberg-Marquardt algorithm	Units: 13-2-2, tansig activations	Levenberg-Marquardt	Non-linear Comparison Model
RFC	Random Forest, Decision Tree Ensemble Classifier	13 inputs, 200 Trees, 2 outputs	Random feature bagging	Non-linear Comparison Model
SVM	Support Vector Machine	13 inputs, 2 outputs	Quadratic Optimisation	Non-linear Comparison Model
TREEC	Trainable Decision tree Classifier	13 inputs, 2 outputs	Decision tree induction	Non -linear Comparison Model
FLNN	Functional link neural network	Units: 13-30-2, tansig activations.	Gradient descent with momentum and adaptive learning rate backpropagation	Test model

RBNC	Feed-forward neural net including N sigmoid neurons.	13 inputs, 2 outputs	The classifier has radial basis units with only 1 hidden layer	Non-linear Comparison Model
NAIVEBC	Naive Bayes classifier	Naive Bayes classifier mapping	Divides each axis into 10 scalar numbers of bins counts the number of training examples for each of the classes in each of the bins, and classifies the object to the class that gives maximum posterior probability. Missing values will be put into a separate bin	Linear Comparison Model
PERLC	Trainable linear perceptron classifier	Units: 13-2, linear activations perceptron classifier mapping	Batch training with weight and bias learning rules	Linear Comparison Model
VPC	Combines an ensemble of perceptron's through voting procedure.	Units: 13-30-2, tansig activations	This Classifier performed by permitting the ensemble of perceptron's to vote in the neural network on the label of each test point.	Non-linear Comparison Model
DRBMC	Discriminative Restricted Boltzmann Machine classifier	The discriminative RBM can be viewed as a logistic regressor with hidden units. The discriminative RBM has N hidden units (default = 5)	It is trained with regularization using regularization parameter set has (default = 0).	Non-Linear Comparison Model

Each of these will be evaluated separately in the next chapter and compared with a combined classifier strategy that is discussed in the following subsection.

4.12 The Combined Classifier Framework

It has become clear that for more complicated data sets the classification results can be improved by combining them. For instance, in (Bassam Moslem et al. 2011), the authors use committee machines with an SVM as the component classifier in order to boost the classification accuracy of multichannel uterine EMG signals. The approach was applied on each channel and a majority voting rule was used in order to determine the final decision of the committee. The results indicate that committee machines exhibit performance unobtainable by an individual committee member on its own.

In our approach, we select particular base-level classifiers, and assemble them together. This involves combining the best classifiers that produce a consistent Area Under the Curve (AUC), Sensitivity, Specificity, Precision, F1 Score and Youden's J statistic (J Score) values. The findings in Ani et al. (Al-Ani, Ahmed and Deriche 2002) state that a classifier with a specific set of features may or may not be an appropriate option for another set of features. In other words, different classification algorithms will achieve a different degree of success for different kinds of applications. Therefore, combining classifiers can offer better complimentary information about the patterns to be classified than any single classifier. Our Combined Classifier Framework for training testing our classifiers is illustrated in Figure 4.13 and Figure 4.13 respectively.

In Figure 4.13, we show the full architecture of our combine classifier framework. The framework starts with a different set of primitive model pools containing machine learning algorithms. These are combined into a number of base model configurations drawn from the model pool, using bootstrap aggregation. These are used to collectively predict class probabilities.

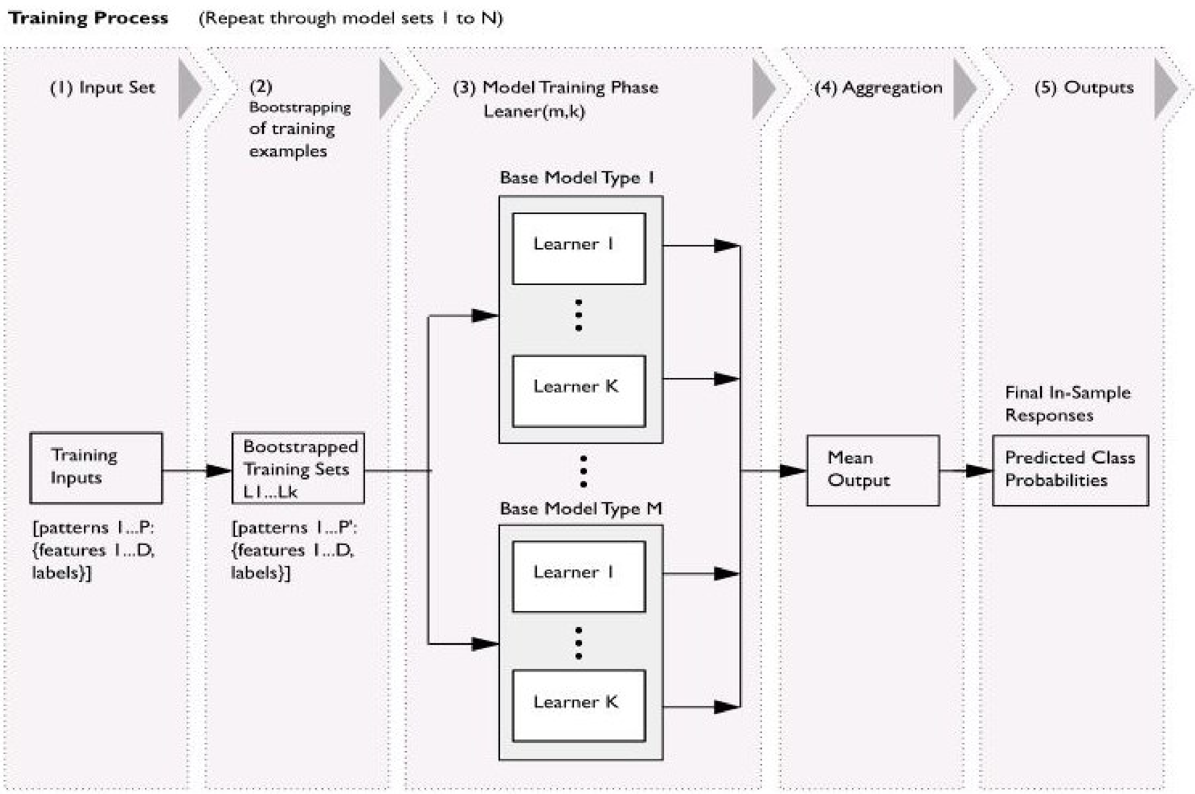


Figure 4.13: Training process of model

Bootstrap aggregation, which is also known as bagging predictors, was proposed in (Breiman 1996) and is a method for generating multiple versions of a predictor and using these to get an aggregated predictor. The aggregation averages over the versions when predicting a numerical outcome and does a plurality vote when predicting a class. The multiple versions are formed by making bootstrap replicas of the learning set and using these as new learning sets. The framework generates model combination sets where the primitive models and model combination sets are created. The base model, which involves the entire chosen algorithm list, is combined through the combinatorial sampling and modelling of the training dataset hundreds of times and averaging its predictions. This method helps to create better classification performance that is more resilient to noise, reduces variance and helps to avoid overfitting.

In Figure 4.13 the input sets $1 \ldots P$ and $1 \ldots D$ represent the patterns and feature labels respectively. The bootstrapping of the primitive model pool $1 \ldots K$ starts with a few training examples of $1 \dots P$ and $1 \dots D$, where model base $1 \dots M$ are represented. The mean outputs of the base models from $1 \dots M$ are aggregated and the in-sample of the predicted class probabilities outputted.

In the Testing process illustrated in Figure 4.14, the same mechanism as the training process is applied. As the set of training examples grows, the classifier improves, provided a limited number of negative examples are misclassified as positive, which could lead to deterioration of performance.

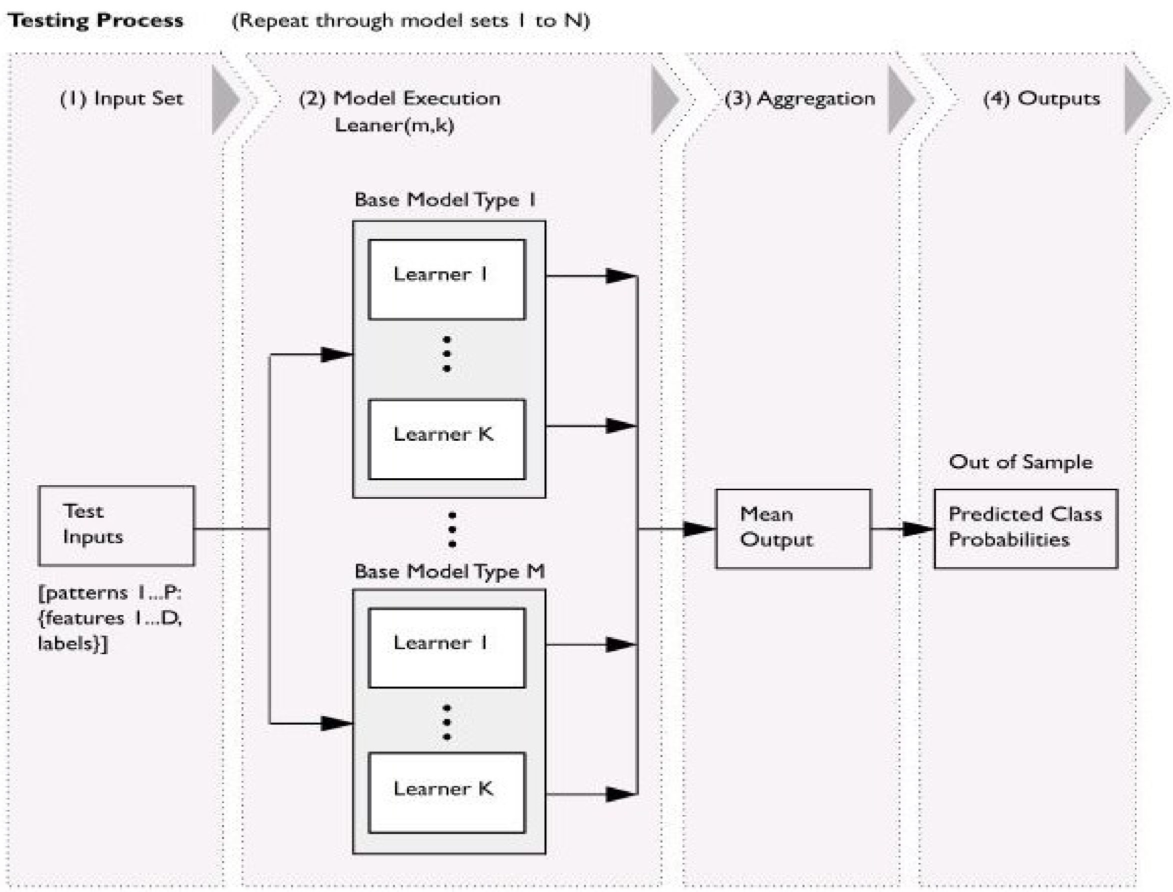


Figure 4.14: Testing process of model

We have a procedure for using the learning set to form a predictor $\varphi(x, L)$ if the input is x we predict D by $\varphi(x, L)$. Now, suppose we are given a sequence of learning sets $\{L_k\}$ each consisting of N independent observations from the same underlying distribution as L. Our task is to use $\{L_k\}$ to get a better predictor than the single learning set predictors $\varphi(x, L)$. The restriction is that all we are allowed to work with is the sequence of predictors $\{\varphi(x,\ L_k)\}$. If D is numerical, an obvious procedure is to replace $\varphi(x, L)$ by the average of $\{\varphi(x,\ L_k)\}$ over k, i.e. by $\varphi_A(x) = E_L\varphi(x, L)$ where $E_L$donates the expectation over L, and the subscript A in φ_A denotes aggregation. If $\varphi(x, L)$ predicts a class $j \in \{1, \dots, J\}$, then aggregate trains multiple K models on different samples (data splits) and averages their predictions by averaging the results of k models. In our framework, this is defined as:

$$\hat{g}_j{}^{*k}(.) = \ \hat{P}^*[Y^{*k} =.](j = 0, \dots, J-1) \tag{4.2}$$

$$\text{yielding an estimator for} \quad \hat{P}[Y = j|X =.] \tag{4.3}$$

Where $\hat{g}$ is the ensemble model, $\hat{P}$ the probability estimate, k the bootstrapped subset, j the classes, y the outcome per subset, and X the patterns per subset.

The benefit of using bootstrap aggregation is its use in non-linear modelling and generalisation that reaches beyond statistical inference to focus on class prediction (machine learning focuses on prediction rather than total explainability).

4.13 Validation Method

The dataset of 300 records (including the oversampled dataset) has been split using a holdout method, which involves keeping the training set totally independent of the test set (Kohavi 1995; Fergus et al. 2015). Different proportions of the dataset are for training and the remainder for testing. For example, when 80% of the dataset, i.e. 240 records, is used for training, the reaming 20% is used for testing – in this thesis 240 records are used for training and 60 for testing.

In order to determine the overall accuracy of each of the classifiers, several validation techniques are used. These include Holdout Cross-validation, which involves keeping the training set totally independent of the test set (Kohavi 1995; Fergus et al. 2015). To maintain generalisation, the training and test sets comprise randomly selected instances from the TPEHG dataset. Since the exact selection of instances, for training is random, it is necessary to repeat the learning and testing stage. The average performance obtained from 100 simulations is utilised. This number is considered, by statisticians, to be an adequate number of iterations to obtain an average (Salkind 2008). After each repetition, the error rate for each classifier is stored and the learning experience of the algorithm wiped so that it does not influence the next test. Producing several repetitions provides average error rates, standard deviations and performance values for each classifier.

K-fold Cross-validation is also used for experiments in this thesis. This technique involves splitting the training set into *k* subsets/folds (where *k* is an integer chosen by the investigator), training the classifier on all of the subsets, bar one; and then testing on that subset. This procedure is repeated until all the subsets have been tested, and the average error used as an estimate of the error of the classifier (Knopff et al. 2009). For example, if a 100-instance dataset was split into 10 subsets of 10 instances each the training set would consist of 90 instances, and the test set would comprise of 10 instances. This in effect increases the number of instances tested to 100, but also increases the computational time compared to hold-out methods, as the algorithm will be trained on 10*90 = 900 instances, as opposed to 80 in the hold-out method. Bias is of course introduced because the algorithm retains its experience from all instances trained on it until it has finished the whole 10 folds, which may be called one repetition. The error rate in the k-fold can be used as a more optimistic estimate of error rate than the hold-out method. A mean error rate and standard deviation can be obtained by using more than one repetition.

4.14 Performance Evaluation Parameters

The experiments in this thesis utilise out-of-sample (testing) and in-sample (training) diagnostics, involving sensitivity, specificity, precision, the F1 score, Youden's J statistic, AUC and classification accuracy.

Sensitivity (SN) and Specificity (SP) are considered suitable evaluation measures for classifiers producing binary outputs, which is the case in this thesis (Lasko et al. 2005). The accuracy of these tests is commonly assessed using measures of True Positives (TP), True Negatives (TN), False Positives (FP), and False Negative (FN) rates. The accuracy of the classification task is the proportion of the total number of predictions that were correct. The notation below in equations 4.4, 4.5 and 4.6, explain how the sensitivity, specificity and accuracy have been calculated.

$$Sensitivity = \frac{TP}{TP + FN} \tag{4.4}$$

$$Specificity = \frac{TN}{TN+FP} \tag{4.5}$$

$$Accuracy = \frac{Number\ of\ TP + Number\ of\ TN}{TP + FP + FN + TN} \tag{4.6}$$

Sensitivity depends only on measurements of preterm subjects (TP and FN), and specificity only on term subjects (TN and FP), so neither one depends on the occurrence of preterm in the test set. For this reason, they are popular measures to test accuracy. The AUC technique is used in our experiments because it provides an acceptable evaluation metric for each of the classifiers considered in this experiment. An AUC is defined in 4.7.

$$AUC = \frac{Sensitivity + specificity}{2}$$

(4.7)

Precision is the number of True Positives divided by the number of True Positives and False Positives. Put another way, it is the number of positive predictions divided by the total number of positive class values predicted. It is also called the Positive Predictive Value (PPV) and is defined in 4.8.

$$PPV = \frac{\mathrm{TP}}{TP+FP}$$

(4.8)

The F1 Score is a measure of a test's accuracy. It considers both recall and precision. Recall is a function of its correctly classified examples (true positives) and its misclassified examples (false negatives). Precision is a function of true positives and examples misclassified as positives (false positives). The F-score is evenly balanced when $\beta = 1$. It favours precision when $\beta > 1$, and recall otherwise (Sokolova et al. 2006). The F1Score is defined in 4.9

$$F^1\ Score = \frac{(\beta^2+1)*Precision*Recall}{\beta^2*Precision+Recall}$$

(4.9)

Youden's J statistic (J Score) was introduce in (Youden 1950). The algorithm is a single statistic that captures the performance of a diagnostic test. In our case, its value ranges from 0 to 1, and has a zero value when a diagnostic test gives the same proportion of positive results for groups with and without the disease. A value of 1 indicates that there is no FP or FN, i.e. the test is perfect. The index gives equal weight to FP and FN values, so all tests with the same value of the index give the same proportion of total misclassified results. The J Score is defined in 4.10.

$$Sensitivity + Specificity - 1$$

(4.10)

Using the Receiver Operating Curve (ROC) provides a graphical representation of the analysis of the cut off values for each of the classifiers, based on the sensitivity and specificity error rates and is defined in 4.11.

$$Roc = \frac{P(x\ positive)}{P(x\ negative)} \tag{4.11}$$

Where $P(x|C)$ denotes the conditional probability that a data entry has the class label C. A ROC curve plots the classification results from the most positive classification to the most negative classification (Sokolova et al. 2006).

4.15 Summary

This chapter has provided a detailed discussion on the methodology used in this thesis. This includes a description of the framework architecture and the subcomponents it comprises. These include data collection and pre-processing, feature engineering, oversampling, training and testing, classification, validation and evaluation. This proposed framework is used to generate the results which are presented in the following chapter.

Chapter 5 Results

5.1 Introduction

This chapter presents the results for each of the experiments considered in this thesis. The literature review identified that a variety of algorithms have been used in classification tasks, yet often producing lower results (Hussain et al. 2015; Iams 2003; Farl et al. 2015; Bassam et al. 2011). The machine learning problem specific to this study is the lack of preterm records compared to term records, and therefore an oversampling technique called SMOTE was implemented to address this issue. Several experiments have been considered to test the robustness of our framework architecture and analyse if it improves on the existing results reported in the literature. In our experiment, the TPEHG dataset already came with some pre-labelled classes. Therefore, we decided that supervised learning would be the most appropriate form of learning. This is because in supervised learning one is trying to find the connection between two sets of observations – in our case it is between term and preterm (Sotiris 2007). Some of the experiments utilise the clinical data for each of the women in the dataset which was made available in December 2012. Data items include the age of the women, parity (the number of previous births), abortions, weight, hypertension, diabetes, placental position, first and second trimester bleeding, funnelling and smoking. Several observations were removed because of missing data. This resulted in a new dataset containing 17 preterm records and 152 term records. Again, in order to balance the dataset, the preterm records have been oversampled using SMOTE to produce 153 preterm and 152 term samples. A new dataset is created that combines the real and synthetic observations (305 observations altogether). Each of these experiments are discussed in the following subsections.

5.1.1 Original Classification Results for 0.34-1 Hz Filter on Channel 3

This section presents the classification results for term and preterm delivery records. This has been achieved using the extracted feature set from the 0.34-1 Hz filter on Channel 3. Using the 80% holdout technique and k-fold cross-validation. This provides a baseline for comparison against all subsequent evaluations that have been performed, using the oversampled dataset, clinical data and the combination of classifiers.

The first evaluation uses the original TPEHG dataset, which contains 38 preterm and 262 term observations. The experiment was conducted using seven artificial neural networks. The performance of each classifier has been evaluated using the mean sensitivity, specificity, mean square errors (MSE), standard deviation, and AUC values. Each experiment has been repeated 30 times, with randomly selected training and test sets for each run. The classifier performance metrics is shown in Table 5.1.

Table 5.1: Original TPEHG Signal (262 Term And 38 Preterm)

Classifiers	Sensitivity	Specificity	AUC
BPXNC	**0.0000**	**0.9987**	**54%**
LMNC	**0.0667**	**0.9519**	**58%**
PERLC	**0.1619**	**0.8647**	**57%**
RBNC	**0.1286**	**0.9622**	**56%**
RNNC	**0.0667**	**0.9474**	**56%**
VPC	**0.0000**	**1.0000**	**50%**
DRBMC	**0.0000**	**0.9981**	**58%**

As it can be seen, the sensitivities (i.e. the ability to classify a preterm record), in this initial test, are low for all classifiers. This is expected since the dataset is unbalanced in favour of term observations, thus there are a limited number of preterm records from which the classifiers can learn. Consequently, specificities are much higher than sensitivities. Table 5.2, illustrates the results obtained from k-fold cross-validation. This has been used to determine whether the results from the holdout method can be improved.

Table 5.2: Original TPEHG signal (262 Term and 38 Preterm) cross-validation

	80% Holdout: 30 Repetitions		Cross Val, 5 Folds, 1 Repetition
Classifiers	**Mean Err**	**Standard Deviation**	**Mean Err**
BPXNC	**0.1278**	**0.0043**	**0.1333**
LMNC	**0.1602**	**0.0331**	**0.1767**
PERLC	**0.2243**	**0.1186**	**0.2400**
RBNC	**0.1434**	**0.0342**	**0.1333**
RNNC	**0.1641**	**0.0363**	**0.1567**
VPC	**0.1267**	**0.0000**	**0.1267**
DRBMC	**0.1283**	**0.0068**	**0.1267**

The k-fold cross-validation results, using five folds and one repetition, illustrate that k-fold cross-validation has improved the error rates, for some of the classifiers. However, these are negligible. Furthermore, the lowest error rates could not be improved below the minimum error rate expected, which is 12.67% (38 preterm / 300 deliveries).

5.1.2 Model Selection on Original Results for 0.34-1 Hz Filter on Channel 3

The Receiver Operating Characteristic (ROC) curve (see Figure 5.1) shows the cut-off values for the false-negatives and false-positive rates for each of the classifiers.

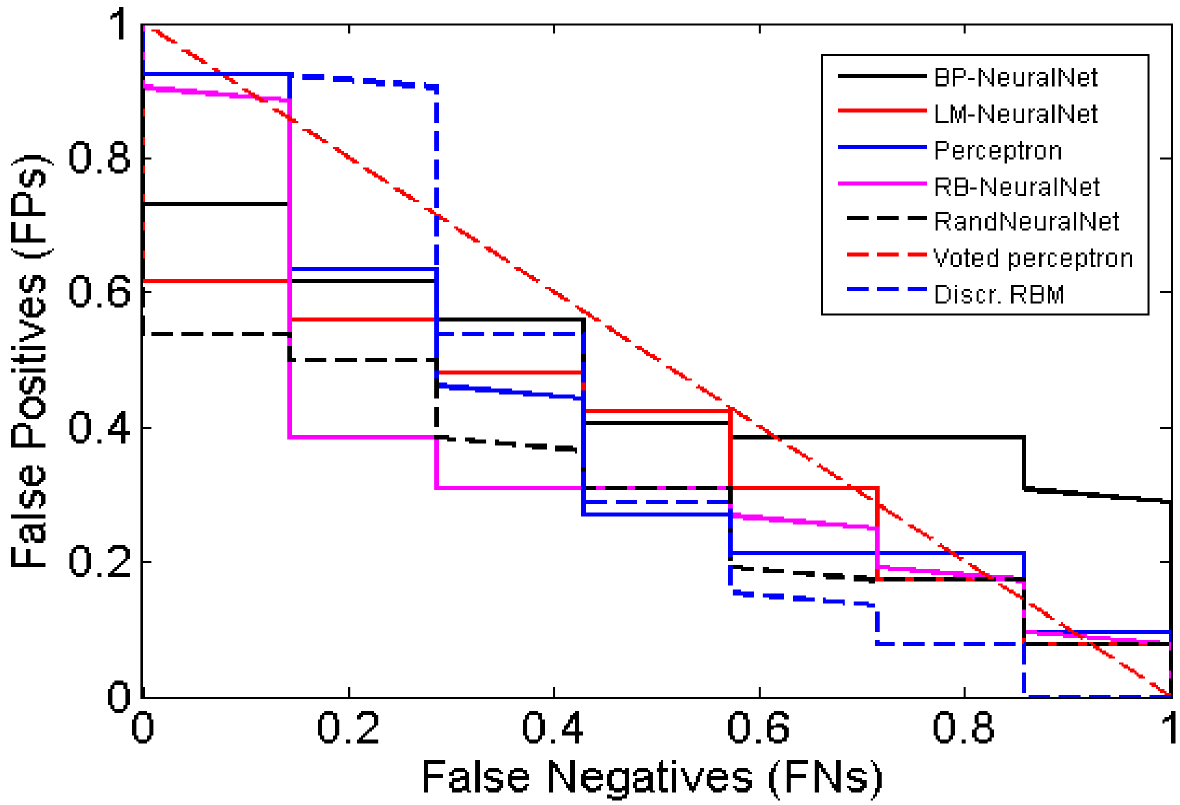


Figure 5.1: Received Operator Curve for 0.34-1Hz Signal of the Original TPEHG Dataset

As it can be seen in Figure 5.1, that none of the classifiers performed particularly well. The AUC values in Table 5.1 support these findings with very low accuracy values. The poor results demonstrate that the classification algorithms do not have enough preterm records to learn from, in comparison to term records. This was expected, as the dataset is unbalanced, and known to cause problems in machine learning tasks.

5.1.3 Results for 0.34-1 Hz TPEHG filter on Channel 3 – Oversampled

In order to improve the results, the preterm observations have been oversampled using the SMOTE technique. This algorithm allows the dataset to become balanced by oversampling the minority class (38 preterm records) to 262, which equals the 262 term samples already provided by the TPEHG database. A new dataset has now been generated that contains an even split between term and preterm records. Using this dataset, the experiment has been repeated a further 30 times.

5.1.4 Classifier Performance on Channel 3 – Oversampled

Table 5.3 illustrates that the sensitivities, for all of the algorithms, have significantly improved, while specificities have decreased. In addition, the AUC results also show a significant improvement in accuracy for all of the classifiers. In particular, the RBNC classifier has dramatically improved with an accuracy of 90%.

Table 5.3: SMOTE TPEHG signal (262 Term and 262 Preterm)

Classifiers	Sensitivity	Specificity	AUC
BPXNC	79%	58%	72%
LMNC	82%	69%	82%
PERLC	46%	67%	63%
RBNC	85%	80%	90%
RNNC	86%	72%	83%
VPC	98%	2%	50%
DRBMC	59%	55%	56%

Table 5.4 illustrates the resulting mean error rates of the oversampled dataset. As it can be seen, the mean error rates, produced by all of the classifiers, are lower than the cross-validation mean errors and the expected error rate, which is 262/524, i.e. 50%.

Table 5.4: SMOTE TPEHG signal (Term and Preterm) cross-validation

	80% Holdout: 30 Repetitions		Cross Val, 5 Folds, 1 Repetition
Classifiers	Mean Err	Standard Deviation	Mean Err
BPXNC	0.3144	0.0591	0.2977

LMNC	**0.2455**	**0.0489**	**0.2195**
PERLC	**0.4321**	**0.0624**	**0.4656**
RBNC	**0.1734**	**0.0424**	**0.1622**
RNNC	**0.2106**	**0.0451**	**0.2023**
VPC	**0.4984**	**0.0088**	**0.5000**
DRBMC	**0.4295**	**0.0376**	**0.4198**

5.1.5 Model Selection

The ROC curve (see Figure 5.2) illustrates the cut-off values for the false-negative and false-positive rates. Compared to Figure 5.1, there is a significant improvement in the accuracy of the classifiers. The values in Table 5.3 support these findings with LMNC, RBNC and the RNNC producing the best AUC, Sensitivity, and Specificity values.

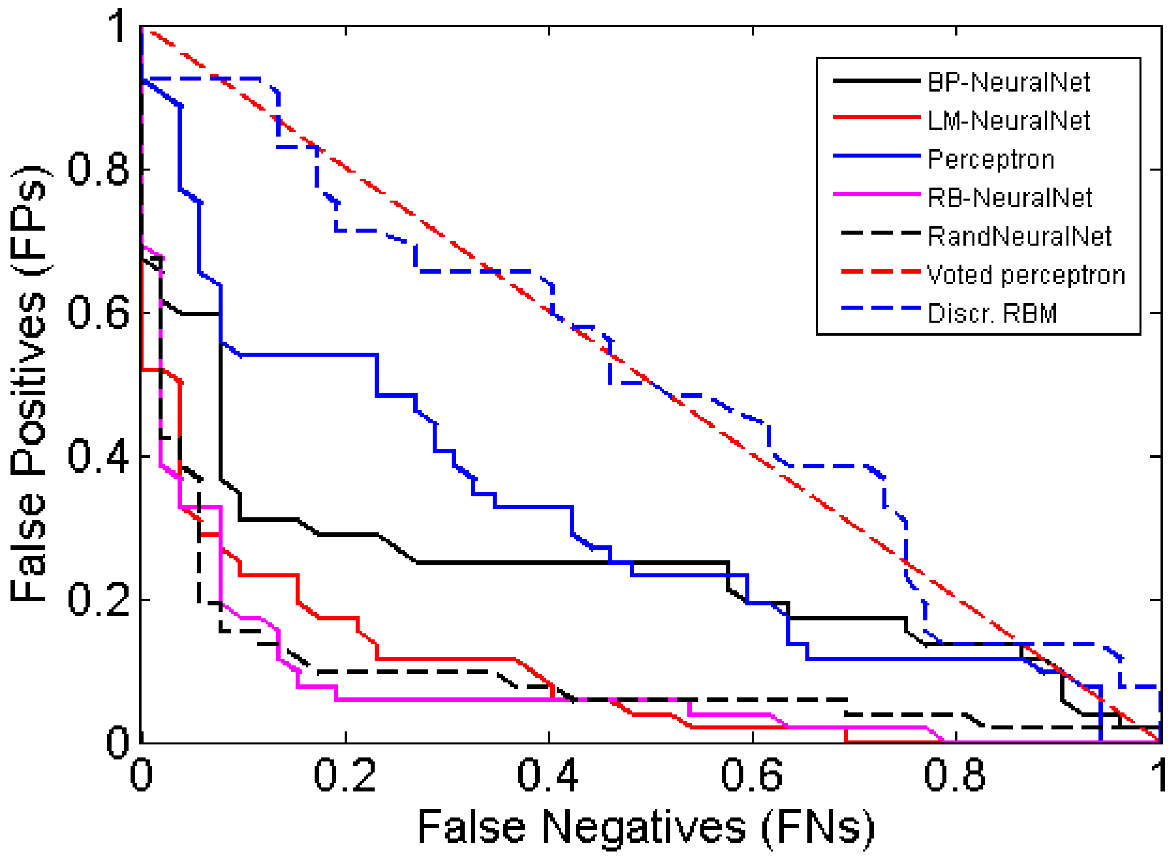


Figure 5.2: Received Operator Curve for the Oversampled 0.34-1Hz Signal TPEHG Dataset

5.1.6 Classifier Performance on Channel 3 combined with Clinical Data

In this experiment the clinical data is combined with our classification models. Table 5.5 illustrates that the inclusion of clinical data has improved the results slightly to those presented in Table 5.3. The LMNC and RBNC classifiers have now produced higher values for the AUC, Sensitivity and Specificity. This is despite having to reduce the size of the dataset to account for missing values in the clinical data (in the case of preterm 22 observations had to be removed; and in the case of term 110 observations had to be removed).

Table 5.5: TPEHG (152 Term and 153 Preterm) trained with Clinical Data

Classifiers	Sensitivity	Specificity	AUC
BPXNC	**64%**	**64%**	**68%**
LMNC	**85%**	**76%**	**85%**
PERLC	**53%**	**61%**	**64%**
RBNC	**87%**	**81%**	**91%**
RNNC	**87%**	**71%**	**84%**
VPC	**100%**	**0%**	**50%**
DRBMC	**56%**	**55%**	**52%**

Table 5.6 illustrates the resulting mean error rates of the dataset containing the clinical data. As it can be seen, the mean error rates produced by several of the classifiers, are much lower than the expected error rate, which is 153/304, i.e. 50%, and are comparable with the cross-validation mean errors.

Table 5.6: Signal and Clinical Data Validation

	80% Holdout: 30 Repetitions	Cross Val, 5 Folds, 1 Repetition

Classifiers	Mean Err	Standard Deviation	Mean Err
BPXNC	**0.3594**	**0.0839**	**0.3508**
LMNC	**0.1932**	**0.0710**	**0.1803**
PERLC	**0.4329**	**0.0674**	**0.3639**
RBNC	**0.1643**	**0.0365**	**0.1377**
RNNC	**0.2097**	**0.0460**	**0.1934**
VPC	**0.4984**	**0.0000**	**0.4656**
DRBMC	**0.4444**	**0.0561**	**0.4525**

5.1.7 Results with Additional Features and Clinical Data

Table 5.7 illustrates that the results have improved on those presented in Table 5.5, indicating that the additional features provide better separation between the two classes.

Table 5.7: TPEHG signal with Additional Features and Clinical Data

Classifiers	Sensitivity	Specificity	AUC
BPXNC	**67%**	**67%**	**70%**
LMNC	**95%**	**81%**	**88%**
PERLC	**56%**	**62%**	**65%**
RBNC	**70%**	**95%**	**94%**
RNNC	**88%**	**72%**	**87%**
VPC	**100%**	**0%**	**50%**
DRBMC	**61%**	**51%**	**51%**

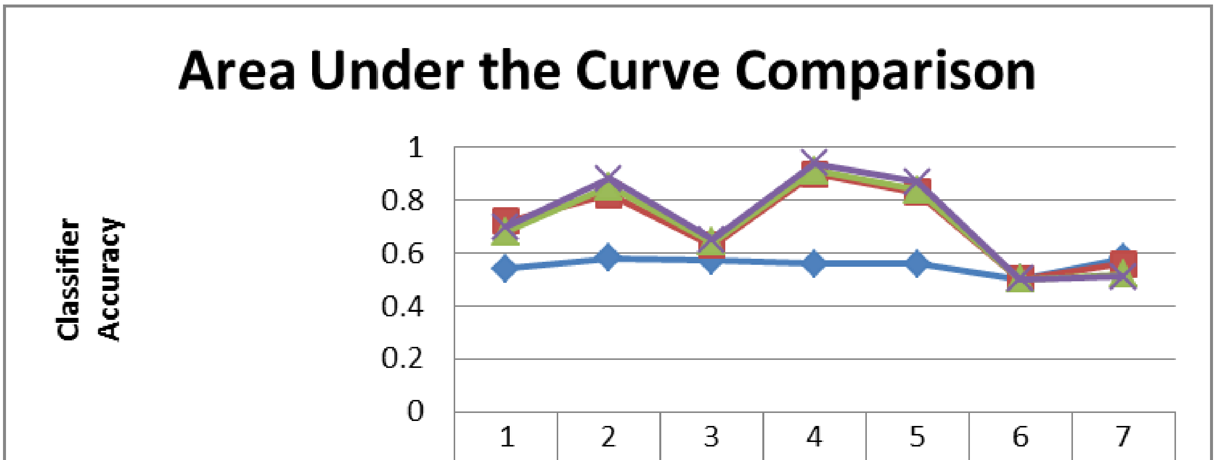
Building on our previous work in (Fergus et al. 2013), this experiment combines features from that work. These additional features are root mean squares, peak frequency and median frequency. Table 5.8 illustrates the resulting mean error rates of the dataset containing the clinical data. As it can be seen, the mean error rates, produced by several of the classifiers, are much lower than the expected error rate, which is 50%, and comparable with the cross-validation mean errors.

Table 5.8: TPEHG Signal with Additional Features and Clinical Data Cross-Validation

	80% Holdout: 30 Repetitions		Cross Val, 5 Folds, 1 Repetition
Classifiers	**Mean Err**	**Standard Deviation**	**Mean Err**
BPXNC	**0.3317**	**0.0838**	**0.3508**
LMNC	**0.1220**	**0.0560**	**0.1803**
PERLC	**0.4118**	**0.0542**	**0.3639**
RBNC	**0.1749**	**0.0406**	**0.1377**
RNNC	**0.1992**	**0.0451**	**0.1934**
VPC	**0.4984**	**0.0000**	**0.4656**
DRBMC	**0.4421**	**0.0527**	**0.4525**

5.1.8 Classifier Performance Comparison

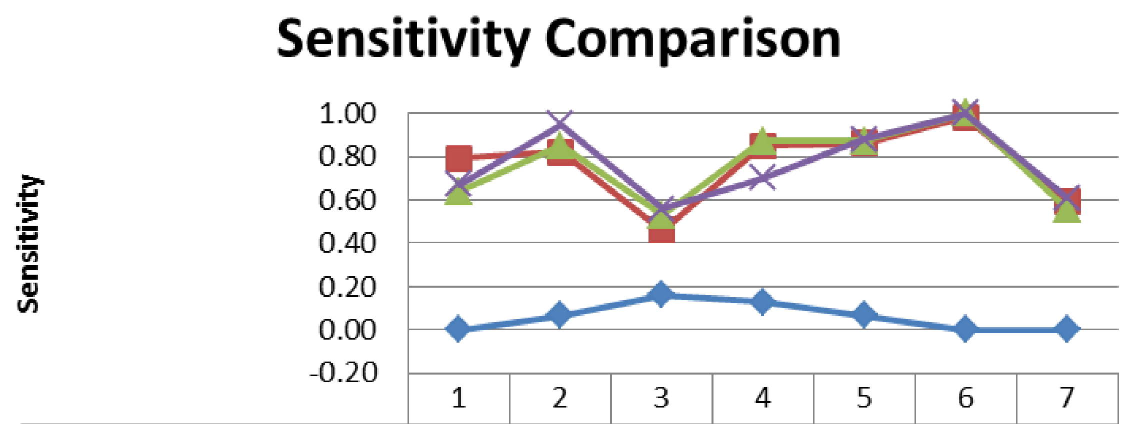
The results from the previously run experiments have now been compared in Figures 5.3, 5.4, and 5.5. As it can be seen in Figure 5.3, all of the classifiers have performed consistently under the four different strategies taken. However, the original unbalanced TPEHG dataset does provide the poorest results. This is due to the disparity between term and preterm observations. Interestingly, the linear and voted perceptron classifiers do not provide sufficient models for classification in any of the strategies used. This is a similar case for the Discriminative Restricted Boltzmann Machine classifier. The simulation results indicate that using the SMOTE oversampling technique, with clinical data and added features, provides the best AUC using the RBNC. This is followed closely by the LMNC and the RNNC. Comparison of AUC values using the Four Strategies is shown in Figure 5.3. Numbers one to seven represent BPXNC, LMNC, PERLC, RBNC, RNNC, VPC and DRBMC classifiers respectively.



	1	2	3	4	5	6	7
Original TPEHG	0.54	0.58	0.57	0.56	0.56	0.5	0.58
SMOTE TPEHG	0.72	0.82	0.63	0.9	0.83	0.5	0.56
SMOTE TPEHG Clinical	0.68	0.85	0.64	0.91	0.84	0.5	0.52
SMOTE TPEHG Clinical Add Feat	0.7	0.88	0.65	0.94	0.87	0.5	0.51

Figure 5.3: Comparison of AUC values using the Four Strategies

Figure 5.4 presents the sensitivities and hence the classifiers ability to classify preterm observations. The focus of our experiments is to try and improve sensitivity rates, as it is more important to predict preterm delivery, as opposed to miss-classifying a term pregnancy. As expected, the sensitivities are low using the original data. This is solely due to the majority of observations being term and only a small number of observations being preterm. The highest sensitivity readings have resulted from strategies 2, 3 and 4, using the LMNC and RBNC. This is consistent with the AUC values that have been depicted in Figure 5.3. Interestingly, the sensitivities are high for the VPC classifier, yet the findings are inconsistent with the very low AUC values for this classifier in Figure 5.3.



	1	2	3	4	5	6	7
Original TPEHG	0.00	0.07	0.16	0.13	0.07	0.00	0.00
SMOTE TPEHG	0.79	0.82	0.46	0.85	0.86	0.98	0.59
SMOTE TPEHG Clinical	0.64	0.85	0.53	0.87	0.87	1.00	0.56
SMOTE TPEHG Clinical Add Feat	0.67	0.95	0.56	0.70	0.88	1.00	0.61

Figure 5.4: Comparison of Sensitivity values using the Four Strategies

Lastly, Figure 5.5 illustrates the specificity results for each of the strategies that have been used. As expected, the specificity values for all classifiers, using strategy one, are high. Again, this is due to the unbalanced dataset (i.e. 262 out of the 300 observations were term). For the LMNC and the RBNC, the values are consistent with the previous figures. Interestingly, using strategy three and four, the LMNC performed better at classifying preterm records, whilst the RBNC classifier is better at classifying term.

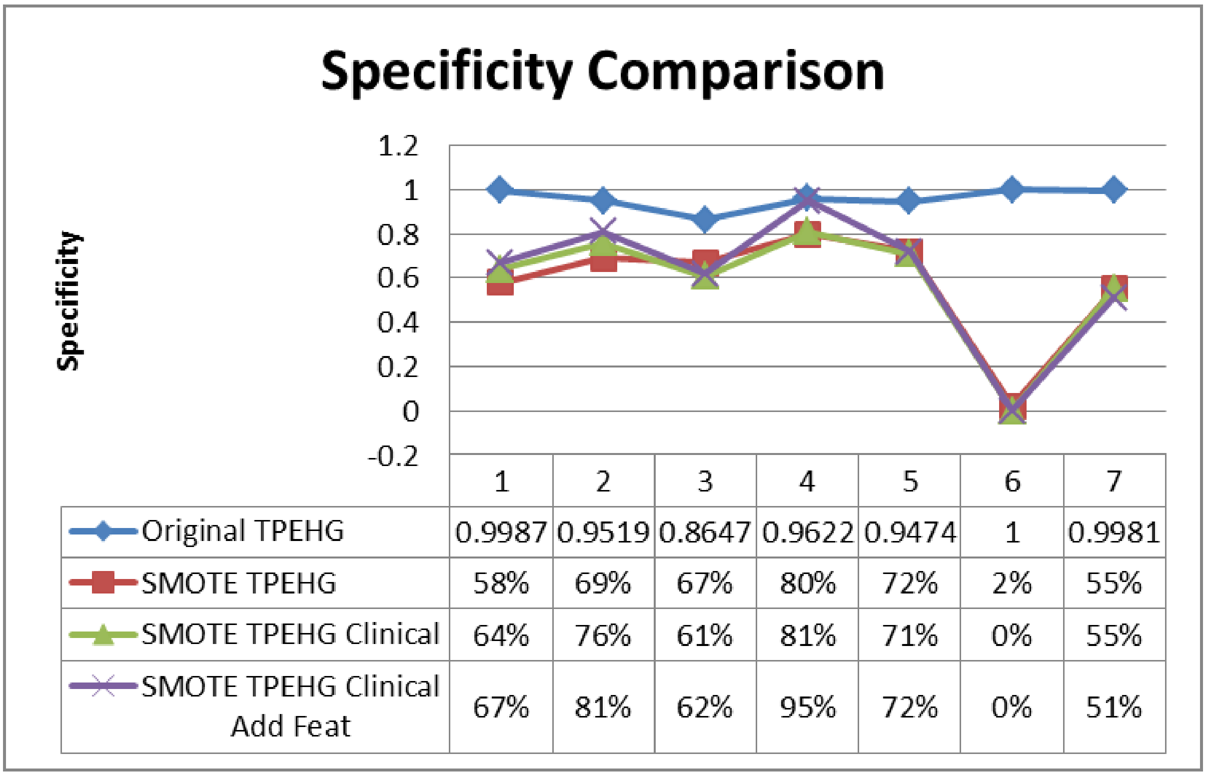


Figure 5.5: Comparison of Specificity values using the Four Strategies

5.1.9 Results for TPEHG on Channel 3 with Additional Classifiers

In this thesis, another set of experiments has been conducted to determine whether the results can be further improved when using several other classifier models. This experiment considers the Random Forest Classifier (RFC), the Support Vector Machine classifier (SVM), Decision Tree Classifier (TREEC), Naïve Bayes Classifier (NAIVBC), and a Functional Link Neural Network (FLNN).

The experiment is performed using the oversampled TPEHG dataset (term records 262 and preterm 262) and all of the 13 features; each experiment has been repeated 30 times with holdout cross validation technique for training and testing. The results illustrated in Table 5.9 indicate that the FLNN classifier produced the best Sensitive, Specificity and AUC values.

Table 5.9: Results for 0.34-1 Hz TPEHG filter on Channel 3 with Additional Classifiers

Model	Sensitivity	Specificity	Precision	F1 Score	Youden's J statistic	Accuracy	AUC

RFC	**84%**	**84%**	**85%**	**84%**	**67%**	**84%**	**88%**
SVM	**64%**	**84%**	**81%**	**71%**	**47%**	**73%**	**78%**
TREEC	**75%**	**73%**	**76%**	**75%**	**48%**	**74%**	**84%**
NAIVEBC	**75%**	**76%**	**77%**	**76%**	**50%**	**75%**	**82%**
FLNN	**84%**	**84%**	**85%**	**84%**	**67%**	**84%**	**89%**

The FLNN algorithm can capture non-linear input, output relationships, provided that they are fed with an adequate set of polynomial inputs, which are constructed out of the original input attributes. In contrast to linear weights of the input patterns produced by the linear links of artificial neural network, the functional link acts on an element of a pattern or on the entire pattern itself by generating a set of linearly independent functions, then evaluating these functions with the pattern as the argument. Thus, class separability is possible in the enhanced feature space.

The RFC shows a strong generalisation towards our dataset with a balanced percentage of 84% for Sensitive and Specificity. The AUC was 88% which is one percent lower than the FLNN at 89%. Random Forests attempts to mitigate the problems of high variance and high bias by averaging to find a natural balance between the two extremes. While the SVM shows that this model is significantly less capable for classifying our data. It is shown that the model does not generalise well from training to testing.

5.1.10 Model Selection on Channel 3 with Additional Classifiers

In this section, we show the ROC curves for each classifier in Figure 5.6.

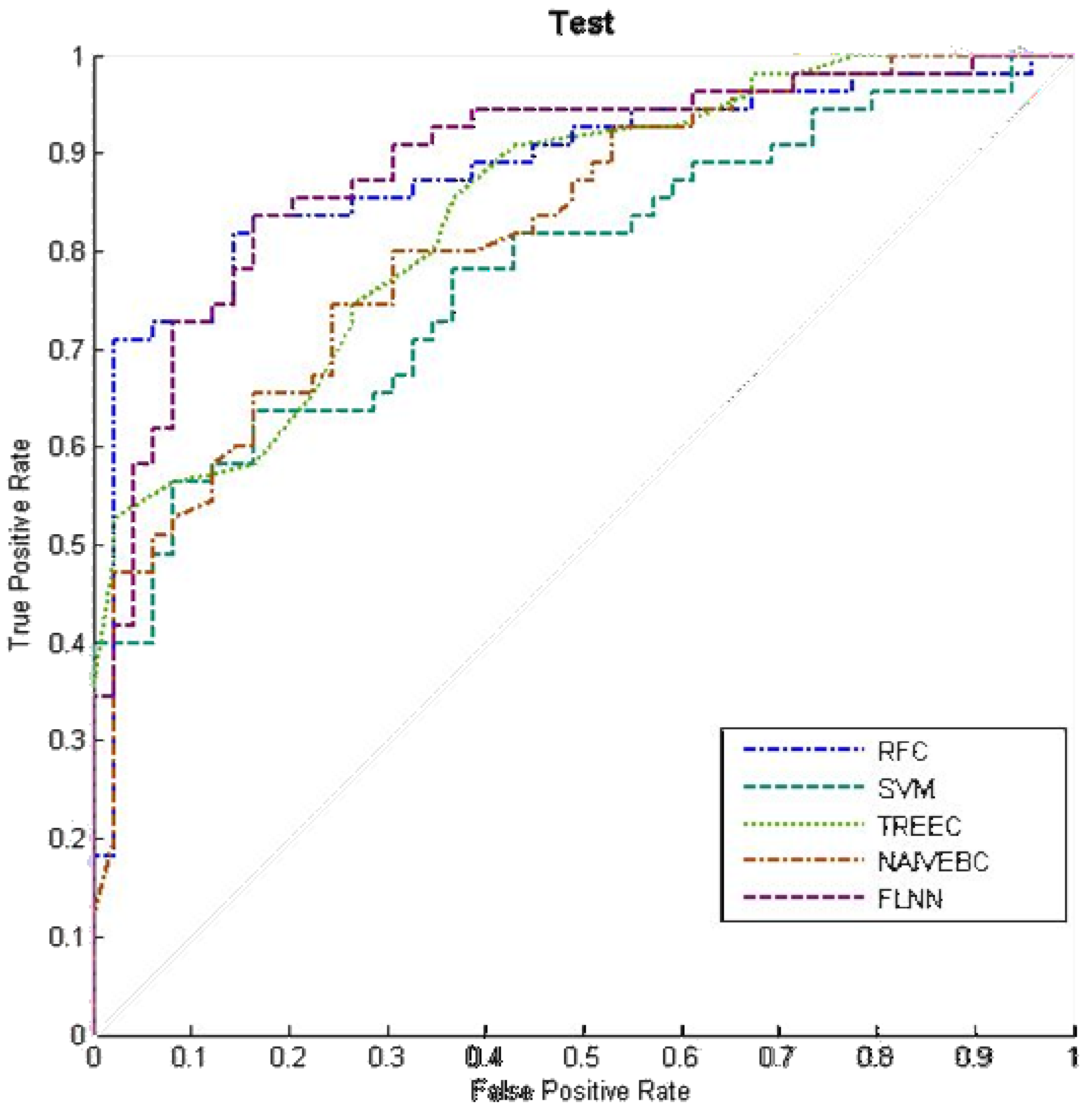


Figure 5.6: Received Operator Curve for the Oversampled 0.34-1Hz Signal TPEHG Dataset Additional Classifiers

In the experiment, classifiers like RFC and FLNN perform well compared to SVM, NAIVEBC and TREEC. The models that produce the best results are considered to be strong non-linear classifiers and are appropriate to act as comparators of high accuracy and performance. The poor results of other classifiers indicate that the algorithms do not learn well from the dataset.

5.2 Combining Classifiers

The previous set of experiments has shown some very interesting results that warrant further investigation. It is the intension of this experiment to combine the capabilities of the best performing classifiers to see if overall classification accuracy can be improved further.

The combined classifier pattern recognition system was designed using the bootstrap aggregating algorithm to improve the stability and accuracy of the selected classifier algorithms used in our experiments. Our approach selects base-level classifiers and searches for a module that is needed to assemble the classifiers together. These involve combining the best classifiers that produce consistent AUC values. The first experiment is conducted using the neural network classifiers. The study presented in this thesis shows that the classifiers that fulfil this criterion are the LMNC, RBNC and the RNNC.

5.2.1 Combine Neural Network Classifier Performance

In Table 5.10, we show the results from the combination of three different classifiers to see if the results can be improved further. The rationale behind combining these three particular classifiers from our model base pool is due to the robustness of the classifiers. The results in Table 5.10 show the results when the three ANN classifiers are combined. Instead of choosing the training instances randomly using a uniform distribution, it chooses the training instances in such a manner as to favour the instances that have not been accurately learned. After several cycles, the prediction is performed by taking a weighted vote of the predictions of each classifier, with the weights being proportional to each classifier's accuracy on its training set.

Table 5.10: Combined Classifiers

Classifiers	Sensitivity	Specificity	AUC

LMNC, RBNC, and RNNC Combined	**91%**	**84%**	**94%**

Table 5.11 illustrates that there is a 12% error, which is slightly high, but much lower than the expected error rate. The cross validation results demonstrate that the 80% holdout technique produces the better results.

Table 5.11: SMOTE TPEHG signal (Term and Preterm) with Additional Features and Clinical Data cross-validation

	80% Holdout: 30 Repetitions		**Cross Val, 5 Folds, 1 Repetition**
Classifiers	**Mean Err**	**Standard Deviation**	**Mean Err**
LMNC, RBNC, and RNNC	**0.1254**	**0.0521**	**0.1623**

Collectively, these set of experiments show that the use machine learning in the prediction of term and preterm records is encouraging. Within a wider context, this approach may be utilised in real-life clinical practice to predict, with high confidence, whether an expectant mother is likely to have a premature birth or proceed to full term.

5.3 Summary

In this chapter, the results from our evaluation have been demonstrated. This thesis focused on how our dataset was constructed and pre-processed, and how features were extracted. Using this feature space, feature selection, classification and combined classification strategies were presented. The results obtained from the experiments are encouraging and support our proposed solution. The novel idea of combining classifiers resulted in significantly better results than those achieved by a single classifier alone. From the literature, it has been revealed that combining base-level classifiers provide a better chance of getting accurate results. It also helps eliminate those classifiers that are not suitable for particular linear or nonlinear datasets. The following chapter discusses these findings further.

Chapter 6 Discussion

Most uterine EHG signal analysis studies concentrate on predicting true labour, which is based on the last stage of pregnancy. However, this thesis has studied the uterine EHG signals of women in order to classify preterm and term deliveries from the earlier stages of the pregnancy. In this thesis we filtered the TPEHG signals using a 0.34 -1Hz Butterworth filter as compared with other approaches (Fele-Žorž et al. 2008; Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013) that used 0.8-4 Hz, 0.3-4 Hz, and 0.3-3Hz filter configurations. The ideology behind the adoption of 0.34-1Hz is based on the findings in (Maner 2003) who argue that uterine electrical activity only occurs within 1Hz. Furthermore, 95% of the patients measured had respiration rates of 0.33 Hz or less. Hence, this research used the 0.34-1Hz bandpass filter to remove these and other artefacts such as, movement.

In this thesis, thirteen features were extracted from channel 3 signals in the TPEHG database. This has never been done in this area of research. The rational of this approach is to look at linear and nonlinear feature methods that could increase the prediction rates in our chosen classifiers. Our results show that apart from features like mean frequency, peak frequency and sample entropy which were proposed in (Fele-Žorž et al. 2008; B Moslem et al. 2011; Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013), features like waveform length, log detector, and variance can also help algorithms to learn better. While these features have typically be used in EMG (Angkoon Phinyomark 2009; Phinyomark, A. Nuidod, P.Phukpattaranont 2012) they have never been used to monitor electrical activity in the uterus. To the best of our knowledge we are the first to do so.

To validate these findings, the discriminate capabilities, of all features, used in this study were assessed using feature ranking. This was achieved using several measures, including statistical significance, linear discriminant analysis using independent search (LDAi), linear discriminant analysis using forward search (LDAf), and linear discriminant analysis using backward search (LDAb).

It has been suggested that ANN are a better solution for nonlinear medical decision support systems than traditional statistical techniques (Li et al. 2000). Therefore, several experiments in this thesis were used in the classification of term/preterm records. The approach is characterised by the use of a supervised machine learning framework to analyse the information embedded within EHG signals. We pose the EHG analytical problem in the form of a discriminative function search and therefore define a classification workflow for application to the empirical trials undertaken. The machine learning approach has been identified as a promising direction in our literature review, offering advantages over manual forms of analysis by human actors and conventional statistical methodologies.

Statistical methods provide a mathematically grounded analytical process, mediating between raw or pre-processed forms of data and human level interpretation (Paydar et al. 2017). However, the analytical forms permissible under the constraints of statistical validity, that is those accompanied by a mathematical proof, restricts the class of addressable problems to those that can be represented using existing mathematical theorems. Machine learning, in contrast, allows for the analysis of arbitrary problem domains, while minimising the need for human grounded assumptions.

The initial classification with the data set in its original form achieved very low sensitivity, below 20%, while the specificity was higher. This means that the classifiers were classifying most of the cases into the majority class, which are term subjects. The main reason for the ineffective classification was the unequal amount of term records to preterm records. Therefore, in these experiments, using the oversampling SMOTE method has significantly improved the sensitivity and specificity rates, for most of the classifiers used.

The first publication of the TPEHG data set was in 2010. However, additional clinical data became freely available in 2012. In our experiments, when analysing the data set, this supplementary data was used. The experimental results demonstrate that the general performance of most ANN classifiers is significantly improved further by comprising the information from the clinical data set. By combining the features with the clinical data, our simulation results showed further improvements in terms of the average sensitivity, specificity and area under the curve values. In this case, the results show that the Levenberg-Marquardt trained Feed-Forward Neural Network classifier performs better at predicting preterm records. This training algorithm approximates Newton's method of least squares optimization and is an efficient learning algorithm, especially when applied to neural networks that have a few hundred weights. However, the efficiency of the algorithm is compromised by high computational requirements.

The Radial Basis Neural Network classifier is preferential for predicting term records. This is due to the properties of the network as an effective multi-dimensional structure, which can provide an alternative to polynomial values. This clearly indicates that using a single classifier for the prediction of term/preterm data may not generate good results. However, in our experiment, the combination of a number of classifiers has been deemed to produce a more reliable classification output result. This has improved on those previously attained by (Paydar et al. 2017; Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013; Diab 2010; Hussain et al. 2015). These results are encouraging and suggest that the approach posited in our experiments shows a line of enquiry worth pursuing.

The simulation results have also shown that the random neural network's ability to classify term and preterm records is good, with an accuracy of 83%. This is a recurrent neural network model, which is inspired by the spiking behaviour of biological neuronal networks. As the problem domain of this thesis is related to classification, rather than prediction, the use of recurring links has no effect on the decision of the classification. Hence, we believe that the RNNC did not generate the highest classification values. This is despite the fact that random neural networks are universal approximators for bounded continuous functions.

It was discovered that the FLNN model was able to yield an average AUC of 89%. The reason behind this is because the model learnt all the signals very quickly and generalised well compared to other network models such as SVC, NAIVEBC and TREEC which suffer from overfitting leading to poor generalisation and performance. The results indicate that the Random Forest AUC of 88% was more effective and yielded better outcomes in comparison to other classification algorithms. Random Forest can ensemble the classifier using many decision tree models, thus making the parameters easy to set and prevent the problem of overfitting.

The results also indicate that the SMOTE oversampling algorithm did not significantly affect the accuracy of the DRBMC or VPC classifiers. This is reasonable since DRBMCs are usually used for feature extraction and initialization procedures for other neural network architectures rather than standalone classifiers. Moreover, the combination of different types of algorithm through model search in our framework produced further increases in Sensitivity, Specificity and AUC values

Our framework ranks the model on how each model class performs against each other, this provides information about how conducive the data for each model class is i.e. if a less powerful model outperforms a complex or less restricted class of model, the test model pulls away from the comparators in terms of upside performance, this reinforces the significance of the achievement, avoiding those classifiers that are not suitable for a particular type of linear or nonlinear dataset. This is also encouraging given that sensitives are important in this research than specificities.

6.1 Summary

In this chapter, the findings from the results obtained are discussed. This includes a discussion on each of the experiments conducted and a justification for the performance metrics achieved. The different evaluation strategies utilised in this thesis are summarised and arguments presented on the usefulness of each. The novel idea of combining classifiers resulted was discussed to highlight that better results were possible than those achieved by a single classifier alone.

Chapter 7 Conclusions and Future Work

7.1 Conclusion

The development of medical information systems has played an important role in the biomedical domain. This has led to the extensive use of Artificial Intelligence (AI) techniques for extracting biological patterns in data. Furthermore, data pre-processing and validation techniques have also been used extensively to analyse such datasets for classification problems. In this thesis, the main aim was to classify between term and preterm records through the use of different types of classification algorithm and their combination to classify term and preterm records contained in the TPEHG dataset. A more conservative filter was used in comparison to many other studies (between 0.34 and 1Hz) to focus only on the electrical activity generated in the myometrium.

7.2 Contribution

The results demonstrate that combining classifiers with high sensitivity, specificity and AUC values can lead to better classification. In this instance, combining the Levenberg-Marquardt trained Feed-Forward Neural Network, Radial Basis Neural Network classifier and the Random Neural Network classifier produced 91% sensitivity, 84% specificity, 94% AUC and a 12% mean error rate. These results are encouraging and suggest that the approach, posited in this thesis, is a line of enquiry worth pursuing. The results of this thesis also encourage more extensive use of Artificial Neural Networks given that models produce more accurate results compare to other machine learning algorithms currently used. This study has shown the benefits of using EHG classification to determine whether delivery will be preterm or term through the process decision based diagnostic tools to support mid-wife nurses, gynaecologists in obstetric care to make the correct decision when treating patients.

7.3 Future Work

Perhaps one negative aspect of the work is the need to utilize oversampling to increase the number of preterm samples. It would have been better to have a balance dataset using actual recordings obtained from pregnant women who delivered prematurely. This will be the focus of future research, alongside a more extensive investigation into different machine learning algorithms and techniques. There is also a need to create new kinds of classifiers and evaluate them using the TPEHG dataset and others, rather than relying on existing algorithms and their combination. This will be the focus of our future work. There are also parameters such as RMS and peak frequencies that have had conflicting results in the classification of term versus preterm records. Further work is needed to identify whether these features really help to distinguish between term and preterm records or not. Since the recording times in the TPEHG dataset were not constant throughout, this means that any differences between term and preterm records in the TPEHG database will also be affected by the gestational age at recording. This will need to be taken into account when designing a classifier, to ensure that these differences are built into classification models. The classifier needs some way of recognising that the changes in some parameters may be due to changes in the gestational period, and not just because they are term or preterm records. In order to do this, the recording weeks will always be included as a feature in the feature set. A regression model could also have been used to test against the classifier.

An effective approach for the future work towards our research is the generation of a set of base classifiers for ensemble feature selection. Ensemble feature selection is used to find feature subsets for generation of the base classifiers for an ensemble with one learning algorithm. This idea was proposed by Ho et al. (Ho 1998) and it shows that simple random selection of feature subsets may be an effective technique for ensemble feature selection. This technique is called the random subspace method (RSM). In the RSM, one randomly selects

$N* < N$ features from the N-dimensional dataset. By this, one obtains the N-dimensional random subspace of the original N-dimensional feature space. This is repeated S times so as to get S feature subsets for constructing the base classifiers. Then, one constructs classifiers in the random subspaces and aggregates them in the final integration procedure. This approach seems to be particularly computationally advantageous and robust against overfitting and it could improve the results returned by feature selection techniques.

Another future work that will be considered is the use of deep learning. Deep learning is a term associated with machine learning approaches that embrace the use of a succession of intermediate feature representations, of increasing abstraction, which jointly give rise to a final solution (Bengio et al. 2013). The motivation and development of the deep learning paradigm was inspired significantly by the layered architecture of neurons present in the visual and auditory mechanisms present in biological systems, especially given the efficacy of biological systems to respond to such sensory pathways. Since its popularisation around 2006, influenced in no small part by the rise of computational power, deep learning has played a transformative role in the scope and scale of tasks addressable by machine based systems, giving rise to advances in Object Recognition, Natural Language Processing, Multi-Task and Transfer Learning, among others. Such advances have enabled accelerated progress in many scientific and commercial settings, for example Drug discovery and toxicology, Bioinformatics, Customer relationship management, the financial sector, and AI assisted research. The critical advantage of deep learning has been described as the provision of a single kind of algorithm, which can be applied to any problem domain with minimal tuning. In contrast, shallow model classes depend heavily on tailored data representations, thereby demanding increased human assumptions and intervention. In terms of empirical success, deep learning models now dominate all of the significant state of the art benchmarks, providing substantial improvements over their predecessors and perhaps representing a future pathway towards artificial intelligence.

The framework we proposed in this thesis begins with the collection of EHG signals. After the signals were collected, we then pre-processed according to broadly distinct phases: cleaning and evaluation, followed by feature pre-processing and extraction. The cleaning and validation phases are potentially iterative, ensuring that subsequent analytical stages are not contaminated by systematic errors or the effect of unintended artefacts. Upon the output of a clean set of data, feature engineering is undertaken in our approach to provide a problem representation suitable for input to the machine learning elements in subsequent analysis. Such human mediated mapping from raw signal forms to alternative representations is typically necessary to ensure that external prior knowledge about the data form is applied, where computational learning elements may fail to derive such high level mappings. The product of feature extraction is a matrix of explanatory variables, which may be presented directly to computational learning elements. The primary analytical phase follows feature derivation, comprising model building and model evaluation. Following the completion of the data transcription, we make use of exploratory data analysis in the form of principal component analysis (PCA) and t-distributed stochastic neighbourhood embedding (tSNE) (Laurens van der Maaten 2008) in order to reveal any intuitive forms of structure prior to the configuration of subsequent analytical methods. The model building process involves the choice and specification of model architectures, in addition to the optimisation of the free parameters of the models using appropriate learning algorithms. The data used during this learning process is referred to as the training set (in-sample data). Following the modelling process, the generalisation performance of such models is evaluated, a procedure that aims to establish the degree to which the previous process of model development succeeded in producing a general model, capable of explaining future unseen data. The portion of data used to test generalisation performance is referred to as the test set (out of sample data), comprising data not used to influence the training procedure when forming the models. We applied a series of test models to the EHG signal dataset for the classification of term/preterm births classification tasks, whose performance was evaluated using graphical forms of analysis, including the ROC, and scalar summary indices including sensitivity, specificity, the F1 score, Youden's J statistic, and overall accuracy.

One of the major problems is the parameter optimization of different machine learning algorithms used in the classifiers and feature selection methods which require a lot of parameter setting to optimize the training and test set. Optimization can be a tedious task especially when these parameters are continuous variables. A general practice for parameter optimization is to use a validation set on which the impact of tuning parameters can be judged and optimized. One of the underlying assumptions of this process is that the validation set closely mirrors the test set, which is often found to be unreliable. On the experimental design, some of the critical problems relate to the amount of data that is necessary for building a reliable and robust machine learning system, how to sample for a validation set, how to choose classifiers (open vs. closed boundary), how to describe a cost matrix and a rejection threshold, how to develop machine learning systems that can automatically determine the optimal parameters (Bhaskar et al. 2006). In this thesis we have shown that parameters can be better optimized using a set that works well on a cross-validation task, where for N-fold cross validation, the data set is split as $N - 1$ folds for training and 1 fold for validation set. This is possible when an extra set of test set is to be used. However, if all that one has is one large data set, then it should be split as training, validation and test for different fold (70%, 20% and 10%, respectively). Ignoring all test folds, parameters can be optimized on the average results across training data and their respective validation sets.

The reason what makes our proposed work different from other research such as (SMS Baghamoradi 2011; Al-askar Haya, Dhiya Jumeily, Abir Hussain 2013; Lange et al. 2014; Nader et al. 2016) is that, it goes far beyond previous attempts of utilizing Electrohysterography for true labour detection during the final seven days, before labour. It offers analysis of Electrohysterography signal using adaptive techniques and utilising advanced computer analysis methods adaptive predictor structures based on different machine learning algorithm architectures that has never been attempted before in the classification of term and preterm records. Within the national healthcare system for many developed countries, there is an immediate need to move from a system focused on treating medical conditions to one that can predict the onset of such conditions. In an increasingly aging population, healthcare is becoming increasingly more unsustainable – understanding the early signs of a condition and implementing countermeasures are seen as a real viable way of reducing spiralling national healthcare costs. The approach posited in this thesis has the potential to achieve this within the field of premature deliveries.

REFERENCES

37steps, 2013. Pattern Recognition Tools. *Version 5*, pp.1–2. Available at: http://www.37steps.com/prtools/prtools5-intro/ [Accessed June 3, 2013].

Agrafiotis, D.K., 2003. Stochastic proximity embedding. *Journal of Computational Chemistry*, 24(September 2002), pp.1215–1221.

Al-Ani, Ahmed and Deriche, M., 2002. A New Technique for Combining Multiple Classifiers using The Dempster-Shafer Theory of Evidence. *Journal of Artificial Intelligence Research*, 17, pp.333–361.

Al-askar Haya, Dhiya Jumeily, Abir Hussain, P.F., 2013. The Application of Self-Organised Network Inspired by Immune Algorithm for Prediction of Preterm Deliveries from EHG Signals. In *PgNet*. Liverpool John Moores University Byroom Street Liverpool L3 3AF, pp. 1–6. Available at: http://www.cms.livjm.ac.uk/pgnet2013/Proceedings/papers/1569764261.pdf [Accessed August 27, 2013].

Alamedine, D., Marque, C. & Khalil, M., 2013. Binary particle swarm optimization for feature Selection on uterine electrohysterogram signal. *2013 2nd International Conference on Advances in Biomedical Engineering*, pp.125–128. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6648863.

Alexander, G.R. et al., 2003. 1995 – 1997 Rates for Whites , Hispanics , and Blacks. *Pediatrics*, 111(1), pp.1995–1997.

Ananth, C. V., Ananth, C. V. & Vintzileos, A.M., 2006. Epidemiology of preterm birth and its clinical subtypes. *Journal of Maternal-Fetal and Neonatal Medicine*, 19(October), pp.773–782. Available at: http://www.tandfonline.com/doi/full/10.1080/14767050600965882.

Andrew Ng, 1998. On Feature selection: Learning with Exponentially many Irrelevant Features as Training Examples. *Proc. 15th Interntional Conference on Machine Learning*, pp.404–412.

Angkoon Phinyomark, C.L. and P.P., 2009. A Novel Feature Extraction for Robust EMG Pattern Recognition. *Journal of Computing*, 1(1), pp.71–79.

Anil K. Jain, J.M., 1996. Artificial neural networks: A tutorial. *IBM Almaden Research Centre*, 1, pp.31–44. Available at: http://web.iitd.ac.in/~sumeet/Jain.pdf [Accessed April 9, 2014].

Bassam, M. et al., 2011. Classification of multichannel uterine EMG signals. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2011, pp.2602–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22254874.

Baxter, M.J., Beardah, C.C. & Westwood, S., 2000. Sample size and related issues in the analysis of lead isotope data. *Journal of Archaeological Science*, 27, pp.973–980. Available at: [PDF].

Bengio, Y., Courville, A. & Vincent, P., 2013. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), pp.1798–1828.

Bhaskar, H., Hoyle, D.C. & Singh, S., 2006. Machine learning in bioinformatics: A brief survey and recommendations for practitioners. *Computers in Biology and Medicine*, 36, pp.1104–1125.

Biau, G., 2012. Analysis of a Random Forests model. *The Journal of Machine Learning Research*, 13, pp.1063–1095. Available at: http://dl.acm.org/citation.cfm?id=2343682.

Blencowe, H. et al., 2013a. Born Too Soon: The global epidemiology of 15 million preterm births. *Reproductive Health*, 10(Suppl 1), p.S2. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3828585&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.reproductive-health-journal.com/content/10/S1/S2.

Blencowe, H. et al., 2013b. Born too soon: the global epidemiology of 15 million preterm births. *Reproductive health*, 10 Suppl 1(Suppl 1), p.S2. Available at: /pmc/articles/PMC3828585/?report=abstract.

Blencowe, H. et al., 2012. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *Lancet*, 379, pp.2162–2172. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22682464.

Blondel, B. et al., 2006. General obstetrics: Preterm birth and multiple pregnancy in European countries participating in the PERISTAT project. *BJOG: An International Journal of Obstetrics & Gynaecology*, 113, pp.528–535. Available at: http://doi.wiley.com/10.1111/j.1471-0528.2006.00923.x.

Blondel, B., 1996. The length of the cervix and the risk of spontaneous premature delivery. *Revue d'epidemiologie et de sante publique*, 44(9), pp.292–294.

Blondel, B. & Kaminski, M., 2002. Trends in the occurrence, determinants, and consequences of multiple births. *Seminars in perinatology*, 26(4), pp.239–249.

Breiman, L., 1996. Bagging predictors. *Machine Learning*, 24(421), pp.123–140.

Breiman, L., 2001. Random Forrest. *Machine Learning*, pp.1–33.

Buhimschi, C., Boyle, M.B. & Garfield, R.E., 1997. Electrical activity of the human uterus during pregnancy as recorded from the abdominal surface. *Obstetrics and Gynecology*, 90(1), pp.102–111.

Bulletin, S., 2011. Statistical Bulletin Gestation-specific Infant Mortality in England and Wales. *National Office for Statistics*, (October), pp.1–10. Available at: http://www.ons.gov.uk/ons/dcp171778_237668.pdf [Accessed February 5, 2014].

Cao, L., 2003. Support vector machines experts for time series forecasting. *Neurocomputing*, 51, pp.321–339.

Carré, P. et al., 1998. Denoising of the uterine EHG by an undecimated wavelet transform. *IEEE transactions on bio-medical engineering*, 45(9), pp.1104–13. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/9735560.

Cass, R. & Radl, B., 1996. Adaptive Process Optimization Using Functional-Link Networks And Evolutioary Optimization. *Control Engineering Practice*, 4(11), pp.1579–1584.

Catalin, Buhimschi, R.G., 1996. Uterine contractility as assessed by abdominal surface recording of electromyographic activity in rats during pregnancy. *American journal of obstetrics and gynecology*, 174(2), pp.744–53.

Catley, C. et al., 2006. Predicting high-risk preterm birth using artificial neural networks. *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 10(3), pp.540–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16871723.

Chen, X., 2003. An Improved Branch and Bound Algorithm for Feature Selection. *Pattern Recognition Letters*, 24, pp.1925–1933.

Chiswick, M.L., 1985. Birth Counts. Statistics of Pregnancy and Childbirth. *Archives of Disease in Childhood*, 60, pp.400–400. Available at: http://adc.bmj.com/cgi/doi/10.1136/adc.60.4.400.

Cookson Victoria, N.C., 2010. NF- κ B function in the human myometrium during pregnancy and parturition Histology and. , pp.945–956.

Cutler, D.R. et al., 2007. Random forests for classification in ecology. *Ecology*, 88(11), pp.2783–2792.

Dalal Snehal, M.L., 2008. A Survey of Methods and Strategies for Feature Extraction in Handwritten Script Identification. In *2008 First International Conference on Emerging Trends in Engineering and Technology*. IEEE, pp. 1164–1169. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4580080 [Accessed November 25, 2014].

Dehuri, S. & Cho, S.-B., 2010. Evolutionarily optimized features in functional link neural network for classification. *Expert Systems with Applications*, 37(6), pp.4379–4391. Available at: http://linkinghub.elsevier.com/retrieve/pii/S095741740901032X [Accessed August 9, 2014].

Diab, A. et al., 2013. Effect of decimation on the classification rate of nonlinear analysis methods applied to uterine EMG signals. *utc.fr*, pp.12–14. Available at: http://www.utc.fr/~umr6600/doc/ABS/RITS-2013_ Final version.pdf [Accessed February 9, 2014].

Diab, M.O., 2010. Classification of uterine EMG signals using supervised classification method. *Journal of Biomedical Science and Engineering*, 3(9), pp.837–842. Available at: http://www.scirp.org/journal/PaperDownload.aspx?DOI=10.4236/jbise.2010.39113 [Accessed August 20, 2013].

Diab, M.O., Marque, C. & Khalil, M., 2009. An unsupervised classification method of uterine electromyography signals: classification for detection of preterm deliveries. *The journal of obstetrics and gynaecology research*, 35(1), pp.9–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19215542 [Accessed January 15, 2014].

Dietterich, T.G., 1997. Machine-Learning Research. *AI Magazine*, 18(4), p.97. Available at:

http://www.aaai.org/ojs/index.php/aimagazine/article/view/1324.

Domingos, P., 2012. A few useful things to know about machine learning. *Communications of the ACM*, 55, p.78.

Doret, M. et al., 2005. Uterine electromyography characteristics for early diagnosis of mifepristone-induced preterm labor. *Obstetrics and gynecology*, 105(4), pp.822–30.

Elkan, C., 2010. Predictive analytics and data mining. , pp.1–164. Available at: http://cseweb.ucsd.edu/users/elkan/255/dm.pdf.

Escobar, G.J. et al., 1999. Rehospitalization in the first two weeks after discharge from the neonatal intensive care unit. *Pediatrics*, 104(1), p.e2.

Farl, D.T., I, M.B. & Shahbakhti, M., 2015. Prediction of Preterm Labor from EHG signals using Statistical and Non-linear Features. , 1.

Fele-Zorz, G. et al., 2008. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Medical & biological engineering & computing*, 46(9), pp.911–22. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18437439 [Accessed September 10, 2013].

Fele-Žorž, G. et al., 2008. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Medical & biological engineering & computing*, 46(9), pp.911–22.

Fergus, P. et al., 2015. Automatic Epileptic Seizure Detection Using Scalp EEG and Advanced Artificial Intelligence Techniques. *BioMed Research International*, 2015, pp.1–18. Available at: http://www.hindawi.com/journals/bmri/2015/986736/abs/.

Fergus, P. et al., 2013. Prediction of Preterm Deliveries from EHG Signals Using Machine Learning. *PLoS ONE*, 8. Available at: http://www.cms.livjm.ac.uk/pgnet2013/Proceedings/papers/1569764261.pdf [Accessed July 24, 2013].

Fergus, P., Idowu, Ibrahim, Hussain, Abir Jaffar, Dobbins, C. & Haya, A.-A., 2014. Evaluation of Advanced Artificial Neural Network Classification and Feature Extraction Techniques for Detecting Preterm Births Using EHG Records. In *Intelligent Computing in Bioinformatics*. pp. 309–314.

Fergus, P., Idowu, I. & Hussain, A., 2014. *Evaluation of Advanced Artificial Neural Network Classification and Feature Extraction Techniques for Detecting Preterm Births Using EHG Records* 10th Inter., Taiyuan, China: Springer. Available at: http://link.springer.com/chapter/10.1007/978-3-319-09330-7_37 [Accessed July 10, 2014].

Freedman, David, Robert Pisani, R.P., 1997. *Statistics Third Edition*, W W Norton & Co Inc (Np).

Freund, Y. & Schapire, R.R.E., 1996. Experiments with a New Boosting Algorithm. *International Conference on Machine Learning*, pp.148–156. Available at: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.51.6252.

Frigge, M., Hoaglin, D.C. & Iglewicz, B., 1989. Some Implementations of the Boxplot. *The*

American Statistician, 43, pp.50–54. Available at: http://links.jstor.org/sici?sici=0003-1305(198902)43:1%3C50:SIOTB%3E2.0.CO;2-E.

Fuchs, A.-R. et al., 1984. Oxytocin receptors in the human uterus during pregnancy and parturition. *American Journal of Obstetrics and Gynecology*, 150, pp.734–41.

Garcia-Gonzalez, M.T. et al., 2013. Characterization of EHG contractions at term labor by nonlinear analysis. *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, 2013, pp.7432–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24111463.

Garfield, R.E. et al., 1998. Control and assessment of the uterus and cervix during pregnancy and labour. *Human Reproduction Update*, 4(5), pp.673–695.

Gelenbe, E., 1991. Theory of the Random Neural Network. *Elsevier Science Publisher B.V*, 1(1), pp.1–20.

Ghaffari, A. et al., 2006a. Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *International journal of pharmaceutics*, 327(1–2), pp.126–38.

Ghaffari, A. et al., 2006b. Performance comparison of neural network training algorithms in modeling of bimodal drug delivery. *International journal of pharmaceutics*, 327(1–2), pp.126–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16959449 [Accessed January 14, 2014].

Giles Lee, M.T., 1987. Learning, invariance, and generalization in high-order neural networks. *Applied Optics*, 26.

Goldenberg, R.L. et al., 2008. Epidemiology and causes of preterm birth. *The Lancet*, 371, pp.75–84.

Greenough, A., 2012. Long term respiratory outcomes of very premature birth (<32 weeks). *Seminars in fetal & neonatal medicine*, 17(2), pp.73–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22300711 [Accessed September 11, 2013].

Gyamfi-Bannerman, C. et al., 2011. Nonspontaneous late preterm birth: etiology and outcomes. *American Journal of Obstetrics and Gynecology*, 205(5), p.456.e1-456.e6. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0002937811010404.

Ham, J. et al., 2005. Investigation of the Random Forest Framework for Classification of Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3), pp.492–501.

Hinton Geoffrey, R.S., 2002. Stochastic neighbor embedding. *Advances in neural information processing systems*, pp.833–840.

Ho, T.K., 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), pp.832–844.

Hongxing Li, C.L. Philip Chen, H.-P.H., 2000. Chapter 4 Functional-link Neural Networks and Visualization Means of Some Mathematical Methods. In Florida: CRC Press LLC, pp. 72–79.

Hornik Kurt, Stinchcombe Maxwell, W.H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), pp.359–366.

Huddy, C.L., Johnson, a & Hope, P.L., 2001. Educational and behavioural problems in babies of 32-35 weeks gestation. *Archives of disease in childhood. Fetal and neonatal edition*, 85, pp.F23–F28.

Hussain, A.J. et al., 2015. Dynamic neural network architecture inspired by the immune algorithm to predict preterm deliveries in pregnant women. *Neurocomputing*, 151, pp.963–974. Available at: http://dx.doi.org/10.1016/j.neucom.2014.03.087.

Iams, J.D., 2003. Prediction and early detection of preterm labor. *Obstetrics and gynecology*, 101(2), pp.402–12. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12576267.

Ian Witten, E.F., 2005. *Data Mining: Practical Machine Learning Tools and Techniques* Second Edi., San Francisco: Elsevier Ltd.

Isabelle Guyon, A.E., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp.1157–1182.

Ivan Verdenik, Marjan Pajntar, B.L., 2001. Uterine electrical activity as predictor of preterm birth in women with preterm contractions. *European journal of obstetrics, gynecology, and reproductive biology*, 95, pp.149–53. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11301159.

Jain, A., Mao, J. & Mohiuddin, K., 1996. Artificial neural networks: A tutorial. *Computer*, 1(1), pp.31–44. Available at: http://www.computer.org/csdl/mags/co/1996/03/r3031.pdf [Accessed July 17, 2014].

Jansson, P. a., 1991. Neural Networks: An Overview. *Analytical Chemistry*, 63(6), p.357A–362A. Available at: http://pubs.acs.org/doi/abs/10.1021/ac00006a739.

John, K.R.K. and G.H., 1997. Wrappers for Feature Subset Selection. *Artificial Intelligence*, 97(1–2), pp.273–324.

Johnson, S. et al., 2015. The long-term consequences of preterm birth: what do teachers know? *Developmental Medicine & Child Neurology*, 57, pp.571–577. Available at: http://doi.wiley.com/10.1111/dmcn.12683.

Kent, A.L. et al., 2011. Mortality and Adverse Neurologic Outcomes Are Greater in Preterm Male Infants TL - 129. *Pediatrics*, 129 VN-, p.124131. Available at: /Users/alexanderweichert/Desktop/Neuer Ordner/Kent et al-2011-PEDIATRICS.pdf%5Cnhttp://dx.doi.org/10.1542/peds.2011-1578.

Kirk Riemer, M.H., 1998. Regulation of uterine smooth muscle function during gestation. *Pediatric research*, 44, pp.615–627.

Kirsopp, C., Shepperd, M.J. & Hart, J., 2002. Search heuristics, case-based reasoning and software project effort prediction. Available at: http://v-scheiner.brunel.ac.uk/handle/2438/1554.

Kittler, J., Hater, M. & Duin, R.P.W., 1996. Combining classifiers. *Proceedings - International Conference on Pattern Recognition*, 2(3), pp.897–901.

Knopff, K.H. et al., 2009. Evaluating Global Positioning System Telemetry Techniques for Estimating Cougar Predation Parameters. *Journal of Wildlife Management*, 73, pp.586–597. Available at: http://www.bioone.org/doi/abs/10.2193/2008-294%5CnWOS:000266133900016%5Cnfile://c/Users/ETOLOGIA/Dropbox/Bibliograf?a VPS/GPS predation/evaluatingglobal.pdf.

Kohavi, R., 1995. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, 14, pp.1137–1143.

Kononenko, I., 2001. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23, pp.89–109. Available at: http://www.sciencedirect.com/science/article/pii/S093336570100077X.

Kotsiantis, S., 2007. Supervised Machine Learning : A Review of Classification Techniques. *Informatical*, 31(3), pp.249–268.

Kuan, C.-M., 2008. *Artificial Neural Networks*, Taipei, Taiwan.

Kuhrt, K. et al., 2015. Development and validation of a predictive tool for spontaneous preterm birth, incorporating quantitative fetal fibronectin, in symptomatic women. *Ultrasound in Obstetrics & Gynecology*, (May 2015), p.n/a-n/a. Available at: http://doi.wiley.com/10.1002/uog.14894.

Kuldip Vora, S.Y., 2014. A Survey on Backpropagation Algorithms for Feedforward Neural Networks. *International Journal of Engineering Development and Research (IJEDR)*, 1(3), pp.193–197.

Kuriyama Hiroshi, C.A., 1961. A STUDY OF THE PARTURIENT UTERUS WITH THE MICROELECTRODE TECHNIQUE 1. *Endocrinology*, 68(6), pp.1010–1025. Available at: http://press.endocrine.org/doi/abs/10.1210/endo-68-6-1010.

Lange, L. et al., 2014. Velocity and directionality of the electrohysterographic signal propagation. *PloS one*, 9(1), p.e86775. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3897754&tool=pmcentrez&rendertype=abstract [Accessed February 14, 2014].

Lasko, T. a. et al., 2005. The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, 38, pp.404–415.

de Lau, H. et al., 2014. Study protocol: PoPE-Prediction of Preterm delivery by Electrohysterography. *BMC pregnancy and childbirth*, 14(1), p.192. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4057931&tool=pmcentrez&rendertype=abstract [Accessed September 24, 2014].

Laurens van der Maaten, G.H., 2008. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9, pp.2579–2605. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=7911431479148734548related:VOiAgwMNy20J.

Lawn, J.E. et al., 2010. Global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. *BMC pregnancy and childbirth*, 10 Suppl 1(Suppl 1), p.S1.

Laws Christian, Dunis Jason, Georgios, S., 2008. *Higher Order and Recurrent Neural Architectures for Trading the EUR / USD Exchange Rate by*, liverpool. Available at: http://www.livjm.ac.uk/automatic/docs/cibef_0208.pdf.

Lee, S.E. et al., 2008. The frequency and significance of intraamniotic inflammation in patients with cervical insufficiency. *American Journal of Obstetrics and Gynecology*, 198(June), p.633.e1-633.e8. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0002937807022314.

Lee Giles, Steve Lawrence, A.C.T., 2001. *Noisy Time Series Prediction using a Recurrent Neural Network and Grammatical Inference*, Princeton, Australia. Available at: http://clgiles.ist.psu.edu/papers/MLJ-finance.pdf.

Leitich, H. et al., 1999. Cervical length and dilatation of the internal cervical os detected by vaginal ultrasonography as markers for preterm delivery: A systematic review. *American Journal of Obstetrics and Gynecology*, 181, pp.1465–1472. Available at: http://www.sciencedirect.com/science/article/pii/S0002937899704072.

Léman, H., Marque, C. & Gondry, J., 1999. Use of the electrohysterogram signal for characterization of contractions during pregnancy. *IEEE Transactions on Biomedical Engineering*, 46, pp.1222–1229.

Li, Y.C. et al., 2000. Neural network modeling for surgical decisions on traumatic brain injury patients. *International journal of medical informatics*, 57, pp.1–9.

Lim, J.W., 2011. The changing trends in live birth statistics in Korea, 1970 to 2010. *Korean journal of pediatrics*, 54(11), pp.429–35. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3254888&tool=pmcentrez&rendertype=abstract.

Liu, H. et al., 2008. A Survey of Feature Extraction Approaches in Analog Circuit Fault Diagnosis. *2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*, pp.676–680. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4756861 [Accessed November 25, 2014].

Liu, L. et al., 2012. Global, regional, and national causes of child mortality: An updated systematic analysis for 2010 with time trends since 2000. *The Lancet*, 379(9832), pp.2151–2161. Available at: http://dx.doi.org/10.1016/S0140-6736(12)60560-1.

Lucovnik, M., Maner, W.L., et al., 2011. Noninvasive uterine electromyography for prediction of preterm delivery. *American journal of obstetrics and gynecology*, 204(3), p.228.e1-10. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3090039&tool=pmcentrez&rendertype=abstract [Accessed December 20, 2013].

Lucovnik, M., Kuon, R.J., et al., 2011. Use of uterine electromyography to diagnose term and preterm labor. *Acta obstetricia et gynecologica Scandinavica*, 90(2), pp.150–7. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3151256&tool=pmcentrez&rendertype=abstract [Accessed September 11, 2013].

Lucovnik, M., Maner, W.L. & Garfield, R.E., 2010. Accuracy of frequency-related parameters of the electrohysterogram for predicting preterm delivery. *Obstetrical & gynecological survey*, 65(2), p.141. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20112463.

M.Guler, E.S., 1994. A Binary-input Supervised Neural Units that Forms Input Dependent Higher-order Synaptic Correlations. *pric. Of World Congress on Neural Network*, 3, pp.730–735.

Maner, W., 2003. Predicting term and preterm delivery with transabdominal uterine electromyography. *Obstetrics & Gynecology*, 101(6), pp.1254–1260. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0029784403003417 [Accessed September 11, 2013].

Maner, W.L. et al., 2006. Characterization of abdominally acquired uterine electrical signals in humans, using a non-linear analytic method. *Medical and Biological Engineering and Computing*, 44, pp.117–123.

Mangham, L.J. et al., 2009. The cost of preterm birth throughout childhood in England and Wales. *Pediatrics*, 123(2), pp.e312-27. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19171583 [Accessed August 12, 2013].

Marque, C.K. et al., 2007. Preterm labour detection by use of a biophysical marker: the uterine electrical activity. *BMC pregnancy and childbirth*, 7 Suppl 1, p.S5. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1892062&tool=pmcentrez&rendertype=abstract [Accessed February 14, 2014].

Martin, J. a et al., 2009. National Vital Statistics Reports Births : Final Data for 2006. *Statistics*, 57(1), pp.1–104.

Marttila, J.. K. and R.., 2005. *Demographic Trends in Nordic Countries* and P. O. Epidemiology, Gestation & E. by I. B. and L. G. . Keith, eds., CRC Press. Available at: http://www.crcnetbase.com/isbn/9780203089927.

Marx, D., 2008. Parameterized complexity and approximation algorithms. *The Computer Journal*, 51(0), pp.60–78. Available at: http://comjnl.oxfordjournals.org/content/51/1/60.short.

Mathworks, 2016. Matlab.

McPheeters, M. et al., 2005. The epidemiology of threatened preterm labor: a prospective cohort study. *American journal of obstetrics and gynecology*, 192(4), pp.1325-9-30. Available at: http://www.ncbi.nlm.nih.gov/pubmed/15846230 [Accessed November 25, 2014].

Misra, D., 2007. Functional Link Artificial Neural Network for Classification Task in Data Mining. *Journal of Computer Science*, 3(12), pp.948–955. Available at: http://dx.doi.org/10.3844/jcssp.2007.948.955.

Moody, 2009. The WFDB Toolbox For Matlab. *PhysioToolkit open source software for biomedical science and engineering*, p.1. Available at: http://physioforge.csail.mit.edu/wfdbtools/wfdb-swig-matlab/.

Moslem, B. et al., 2011. Classification performance of the frequency-related parameters derived from uterine EMG signals. *Conference proceedings : ... Annual International*

Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2011, pp.3371–4. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22255062.

Moslem, B. et al., 2011. Combining multiple support vector machines for boosting the classification accuracy of uterine EMG signals. *2011 18th IEEE International Conference on Electronics, Circuits, and Systems*, pp.631–634. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6122354.

Moslem Bassam, Diab Mohamad, Khalil Mohamad, M.C., 2012. Combining data fusion with multiresolution analysis for improving the classification accuracy of uterine EMG signals. *EURASIP Journal on Advances in Signal Processing*, 2012(1), p.167. Available at: http://asp.eurasipjournals.com/content/2012/1/167 [Accessed February 7, 2014].

Moslem Bassam, Khalil Mohamad, Diab Mohamad, Chkeir Aly, M.C., 2011. A multisensor data fusion approach for improving the classification accuracy of uterine EMG signals. In *2011 18th IEEE International Conference on Electronics, Circuits, and Systems*. IEEE, pp. 93–96. Available at: http://ieeexplore.ieee.org/document/6122222/.

Muglia Louis, K.M., 2010. The enigma of spontaneous preterm birth. *The New England journal of medicine*, 362(6), pp.529–35.

Nader, N. et al., 2016. A node - wise analysis of the uterine muscle networks for pregnancy monitoring. , pp.712–715.

Naik Arti, P., 2012. Weather Classification and Forecasting using Back Propagation Feed-forward Neural Network. *International Journal of Scientific and Research Publications,*, 2(12), pp.1–3.

Nations, U., Steering, N.S. & Lanka, S., 2005. Why are 4 million newborn babies dying each year ? , 4, pp.12–14.

Neelanjana, Mukhopadhaya, A.S., 2007. Reproductive outcomes after in-vitro fertilization. *Current Opinion in Obstetrics and Gynecology*, 19, pp.113–119. Available at: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=00001703-200704000-00004.

Nicholson, W.K., Frick, K.D. & Powe, N.R., 2000. Economic burden of hospitalizations for preterm labor in the United States. *Obstetrics and gynecology*, 96(1), pp.95–101. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10862850.

Offiah, I., O'Donoghue, K. & Kenny, L., 2012. Clinical Risk Factors for Preterm Birth. *Preterm Birth–Mother and Child*. Available at: http://www.intechopen.com/source/pdfs/27121/InTech-Clinical_risk_factors_for_preterm_birth.pdf [Accessed November 18, 2014].

Pace, L., 2012. Correlation and Regression. *Beginning R: An Introduction to Statistical Programming*, (i), pp.215–242.

Pajntar, M., Roskar, E. & Rudel, D., 1987. Electromyographic observations on the human cervix during labor. *American journal of obstetrics and gynecology*, 156, pp.691–697. Available at: http://www.ncbi.nlm.nih.gov/pubmed/3826220.

Pal, M., 2005. Random forest classifier for remote sensing classification. *International*

Journal of Remote Sensing, 26(June 2015), pp.217–222.

Pao, T., 1992. Functional-link net computing: theory, system architecture, and functionalities. *Computer*, 25(5), pp.76–79. Available at: http://www.neuro.sfc.keio.ac.jp/publications/pdf/paocomputer.pdf [Accessed August 9, 2014].

Park, S., Smith, M.J.T. & Mersereau, R.M., 2000. Target Recognition Based on Directional Filter Banks and Higher-Order Neural Networks. *Digital Signal Processing*, 10(4), pp.297–308. Available at: http://linkinghub.elsevier.com/retrieve/pii/S1051200400903762 [Accessed August 8, 2014].

Patra, J. & Bornand, C., 2010. Nonlinear dynamic system identification using Legendre neural network. *Neural Networks (IJCNN), The 2010*, pp.1–7. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5596904 [Accessed August 9, 2014].

Paydar, K. et al., 2017. A clinical decision support system for prediction of pregnancy outcome in pregnant women with systemic lupus erythematosus. *International Journal of Medical Informatics*, 97, pp.239–246. Available at: http://dx.doi.org/10.1016/j.ijmedinf.2016.10.018.

Petrou, S., 2003. Economic consequences of preterm birth and low birthweight. *BJOG: An International Journal of Obstetrics and Gynaecology*, 110(April), pp.17–23.

Philip Baker, L.K., 2006. *Obstetrics by Ten Teachers* 17th Editi.,

Philip Farrell, R.W., 1975. Epidemiology of Hyaline Membrane Disease in the United States: Analysis of National Mortality Statistics. *official Journal American Academy of Pediatrics*, 58(2), pp.167–176.

Phinyomark, A. Nuidod, P.Phukpattaranont, C.L., 2012. Feature Extraction and Reduction of Wavelet Transform Coefficients for EMG Pattern Classification. *Electronics & Electrical Engineering*, 6(6). Available at: http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=13921215&AN=78022500&h=LtDOFPhYqCXHJ9CQcj4YtJsLP3dTgS2aq34bJAC1WBx2Ci6xCesdsOYofUKwmf8cZAvZfOJvscpAQWQHMNV9fA%253D%253D&crl=c [Accessed February 18, 2014].

Phinyomark, A., Limsakul, C. & Phukpattaranont, P., 2009. A Novel Feature Extraction for Robust EMG Pattern Recognition. , 1(1), pp.71–80. Available at: http://arxiv.org/abs/0912.3973.

PhysioNet, 2012b. The Term -Preterm EHG Database (TPEHG- DB). *physionet.org*, p.1. Available at: http://www.physionet.org/physiobank/database/tpehgdb [Accessed June 3, 2013].

Plunkett, J. & Muglia, L.J., 2008. Genetic contributions to preterm birth: Implications from epidemiological and genetic association studies. *Annals of Medicine*, 40(October), pp.167–179. Available at: http://www.tandfonline.com/doi/full/10.1080/07853890701806181.

Population, N., 2012. Quantile-Quantile Plot (QQ-plot) and the Normal Probability Plot Section 6-6 : Normal Probability Plot. , 2377(Winter), pp.1–8.

Potter, K., 2006. Methods for Presenting Statistical Information: The Box Plot. *Visualization of Large and Unstructured Data Sets*, 4, pp.97–106.

Powell Victor, L.L., 2016. Principal Component Analysis. *Setosa*. Available at: http://setosa.io/ev/principal-component-analysis/ [Accessed January 30, 2016].

Rabotti, C. et al., 2010. Noninvasive estimation of the electrohysterographic action-potential conduction velocity. *IEEE transactions on bio-medical engineering*, 57(9), pp.2178–87. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20460202 [Accessed October 2, 2013].

Rattihalli, R., Smith, L. & Field, D., 2012. Prevention of preterm births: are we looking in the wrong place? The case for primary prevention. *Archives of disease in childhood. Fetal and neonatal edition*, 97(3), pp.F160-1. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22247417 [Accessed September 11, 2013].

Ren, P. et al., 2015. Improved Prediction of Preterm Delivery Using Empirical Mode Decomposition Analysis of Uterine Electromyography Signals. *Plos One*, 10, p.e0132116. Available at: http://dx.plos.org/10.1371/journal.pone.0132116.

Richman, J. & Moorman, J., 2000. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology- ...*, 278(6), pp.H2039-49. Available at: http://ajpheart.physiology.org/content/278/6/H2039.short [Accessed January 17, 2014].

Russell, S. & Norvig, P., 2010. *Artificial Intelligence - A Modern Approach* 3rd ed. M. Hirsch, ed., New Jersey: Pearson Education, Inc.

Saigal, Saroj, Doyle, L.W., 2008. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, 371, pp.261–269.

Saigal, S. et al., 2001. Physical growth and current health status of infants who were of extremely low birth weight and controls at adolescence. *Pediatrics*, 108, pp.407–415.

Salkind, N.., 2008. *Statistics for People Who (Think They) Hate Statistics* 3rd editio., Sage Publication.

Sassano, M., 2008. An Experimental Comparison of the Voted Perceptron and SUpport Vector Machines in Japanese Analysis Tasks. *International Joint Conference on Natural Language Processing*, pp.829–834.

Savoy, J., 2013. Feature selections for authorship attribution. *Proceedings of the 28th Annual ACM Symposium on ...*, pp.939–941. Available at: http://dl.acm.org/citation.cfm?id=2480541.

Shardlow, M., 2007. An Analysis of Feature Selection Techniques. *Studentnet.Cs.Manchester.Ac.Uk*, pp.1–7. Available at: https://studentnet.cs.manchester.ac.uk/pgt/COMP61011/goodProjects/Shardlow.pdf.

Sheather, S.J., 2004. Density estimation. *Statistical Science*, 19(4), pp.588–597.

Sim, S. et al., 2014. Evaluation of Electrohysterogram Feature Extraction to Classify the Preterm and Term Delivery Groups. In *Find out how to access preview-only content The 15th International Conference on Biomedical Engineering IFMBE Proceedings*. pp. 675–678.

Smith, L.K. et al., 2009. Socioeconomic inequalities in survival and provision of neonatal care: population based study of very preterm infants. *BMJ (Clinical research ed.)*, 339, p.b4702.

Smrdel, J., 2015. Separating sets of term and pre-term uterine EMG records. *Physiological Measurement*, 36(2), pp.341–355. Available at: http://stacks.iop.org/0967-3334/36/i=2/a=341?key=crossref.61bb646ab0feaa9b3f47946cfc101a5e.

SMS Baghamoradi, M.N.H.A., 2011. Evaluation of cepstral analysis of EHG signals to prediction of preterm labor. In *18th Iranian Conference on Biomedical Engineering*. Tehran, Iran: 978-1-4673-1005-5/11/$26.00 ©2011 IEEE, pp. 14–16.

Sokolova, M., Japkowicz, N. & Szpakowicz, S., 2006. Beyond accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation. *Advances in Artificial Intelligence (Lecture Notes in Computer Science)*, 4304(c), pp.1015–1021. Available at: http://dx.doi.org/10.1007/11941439_114.

Sotiris, K., 2007. Supervised Machine Learning: A Review of Classification Techniques. *An International Journal of Computing and Informatics*, 31(3), p.3. Available at: http://books.google.com/books?hl=en&lr=&id=vLiTXDHr_sYC&oi=fnd&pg=PA3&dq=Supervised+Machine+Learning:+A+Review+of+Classification+Techniques&ots=CVsxxx_Jgr&sig=-rL-r28GqfN7WTQzX7SoFqYwsT4.

Srinivas, M., 2012. Using Rule Based Classifiers for the Predictive Analysis of Breast Cancer Recurrence. *Journal of Information Engineering & ...*, 2(2), pp.12–20. Available at: http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=22250506&AN=82584075&h=6coJrYSuvMFlhtRzqL9Gqk2SbvtwyVFirCMlbj9+qTM6TGZ8DiABGVC+OvmcIM7Z/cNQQheLV5HG4zzTbXwvIw==&crl=c.

Steer, P., 2005. The epidemiology of preterm labour. *BJOG : an international journal of obstetrics and gynaecology*, 112 Suppl(March), pp.1–3. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17206958.

Strasburger, V., 2006. Clueless: Why do pediatricians underestimate the media's influence on children and adolescents? *Pediatrics*, 117, pp.1427–1431. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16585342.

Stuart Russell, P.N., 1995. *Artificial Intelligence - A Modern Approach* 3rd ed., New Jersey: Pearson Education, Inc.

Svetnik, V. et al., 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences*, 43, pp.1947–1958.

Tang, J., Alelyani, S. & Liu, H., 2014. Feature Selection for Classification: A Review. *Data Classification: Algorithms and Applications*, pp.37–64.

Tanwani, A.K. et al., 2009. Guidelines to select machine learning scheme for classification of biomedical datasets. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5483 LNCS, pp.128–139.

Tezuka, N. et al., 1995. Changes in transcripts encoding calcium channel subunits of rat myometrium during pregnancy. *Am J Physiol*, 269, pp.C1008-17. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7485440.

Tho Dang, X. et al., 2013. A Novel Over-Sampling Method and its Application to Cancer Classification from Gene Expression Data. *Chem-Bio Informatics Journal*, 13, pp.19–29. Available at: http://japanlinkcenter.org/DN/JST.JSTAGE/cbij/13.19?lang=en&from=CrossRef&type=abstract.

Tong, L.I., Chang, Y.C. & Lin, S.H., 2011. Determining the optimal re-sampling strategy for a classification model with imbalanced data using design of experiments and response surface methodologies. *Expert Systems with Applications*, 38(4), pp.4222–4227. Available at: http://dx.doi.org/10.1016/j.eswa.2010.09.087.

Tortosa-Ausina, E., 2002. Financial costs, operating costs, and specialization of Spanish banking firms as distribution dynamics. *Applied Economics*, 34(February), pp.2165–2176. Available at: http://www.tandf.co.uk/journals.

Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. *The Nature of Statistical Learning Theory*.

Vasak, B. et al., 2013. Uterine electromyography for identification of first-stage labor arrest in term nulliparous women with spontaneous onset of labor. *American journal of obstetrics and gynecology*, 209(3), p.232.e1-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23727524 [Accessed January 21, 2014].

Venna Jarkko, K.S., 2007. Nonlinear Dimensionality Reduction as Information Retrieval. *Web*, pp.1–8. Available at: papers2://publication/uuid/E5AE72D2-AC36-42B9-BCA4-2B6E58BCF3F0.

Venugopal, G., Navaneethakrishna, M. & Ramakrishnan, S., 2014. Extraction and analysis of multiple time window features associated with muscle fatigue conditions using sEMG signals. *Expert Systems with Applications*, 41(6), pp.2652–2659. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0957417413009202 [Accessed February 5, 2014].

Volker Tresp, M.T., 1995. Combining estimators using non-constant weighting functions. *Advances in Neural Information Processing Systems 7*, pp.419--426. Available at: http://books.google.com/books?hl=en&lr=&id=M9WuI6tiqRcC&oi=fnd&pg=PA419&dq=Combining+Estimators+Using+Non-Constant+Weighting+Functions&ots=GtcthU2BVA&sig=ouCmCArf7BkcL5ZrieN5L2aVmM0.

WHO, 2012. *Born Too Soon: The Global Action Report on Preterm Birth*, Geneva. Available at: http://www.who.int/pmnch/knowledge/publications/2012_progress_report.pdf.

Włodzisław Duch, Tomasz Winiarski, Jacek Biesiada, A.K., 2003. Feature Ranking, Selection and Discretization. In *Int. Conf. on Artificial Neural Networks (ICANN) and Int. Conf. on Neural Information Processing (ICONIP)*. pp. 251 – 254.

Wold, S., Esbensen, K. & Geladi, P., 1987. Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, pp.37–52.

Wu, Chun-hao, Tseng, Y., 2011. Data Compression by Temporal and Spatial Correlations in a Body-Area Sensor Network: A Case Study in Pilates Motion Recognition. *IEEE Transactions on Mobile Computing*, 10(10), pp.1459–1472. Available at: http://ieeexplore.ieee.org/document/5677553/.

Xu, Xin and Tan, Hongzhuan and Zhou, Shujin and He, Yue and Shen, Lin and Liu, Yi and Hu, Li and Wang, Xiaojuan and Li, X., 2014. Study on the application of Back-Propagation Artificial Neural Network used the model in predicting preterm birth. *PMID*, 35(9), p.1028—1031. Available at: http://europepmc.org/abstract/MED/2549214.

Ye-Lin, Y. et al., 2014. Automatic Identification of Motion Artifacts in EHG Recording for Robust Analysis of Uterine Contractions. *Computational and mathematical methods in medicine*, 2014, p.470786. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24523828.

Yoh-Han, P., 1989. *Adaptive pattern recognition and neural networks*, Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.

Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer*, 3, pp.32–35.

Zardoshti Wheeler, B.H., 1993. Evaluation of emg features for movement control of prostheses. In *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Societ*. IEEE, pp. 1141–1142. Available at: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=979061 [Accessed November 25, 2014].

Zeitlin, J., 2002. Fetal sex and preterm birth: are males at greater risk? *Human Reproduction*, 17(10), pp.2762–2768. Available at: http://humrep.oxfordjournals.org/content/17/10/2762.abstract.

Zhang, G.P., 2000. Neural network for classification: A Survey. *IEEE transaction on system, man and cybernetics - part c: applications and reviewa*, 30(4), pp.451–456.

Zhao, J. et al., 2005. Levenberg-Marquardt based neural network control for a five-fingered prosthetic hand. *... and Automation, 2005 ...*, (50435040), pp.4482–4487. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1570810 [Accessed December 17, 2013].

Zhen-Guo, Tzu-An Chiang, Z.-H.C., 2011. Feed-Forward Neural Networks Training: A Comparison between Genetic Algorithm and Back-Propagation Learning Algorithm. *International Journal of Innovative Computing Information and Control*, 7(10), pp.5839–5843.