Application of Machine Learning and Cervical Impedance Analyses to Preterm Birth Prediction

By

Di Zhang

Supervised by:

Professor Zi-Qiang Lang

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

University of Sheffield

Faculty of Engineering

Department of Automatic Control and Systems Engineering

April 2023

Abstract

Objectives Preterm birth (PTB) is the second leading cause of premature child deaths under 5 years old worldwide and is associated with a series of lifetime disease and disability and substantial long-term healthcare costs. Prediction and prevention of PTB remains limited by the recognition of PTB causes and the modest accuracy of current prediction approaches. This work was based on the Electrical Impedance Prediction of Preterm birth by spectroscopy of the cervix (ECCLIPPx^{TM}) project and the Electrical Impedance Prediction of Preterm Birth by Spectroscopy of the cervix II (ECCLIPPx II) project, which proposed the novel PTB prediction approaches using cervical electrical impedance spectroscopy (EIS) and cervical magnetic impedance spectroscopy (MIS) data. The aim was to apply machine learning and cervical impedance analysis to preterm birth prediction. The objectives include (1) pre-process EIS data and carry out the EIS based PTB prediction, (2) improve the PTB prediction accuracy by combining EIS-based and maternal characteristics-based PTB risk models and (3) carry out a feasibility study on the MIS-based PTB risk prediction.

Methods For the study of EIS data based PTB prediction, the EIS data from 438 recruited women at 19 to 23 weeks of gestation were investigated. A principle of identifying the optimum EIS spectrum among several recorded EIS spectra for each patient was proposed preparing for training the optimal-EIS spectrum filter model. Then the association of the EIS data with the PTB risk was determined by training the logistic regression (LR) model using a series of data modelling and analysis techniques. Moreover, a models-combining approach of combining two probabilities given by EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction model, was proposed to improve accuracy of prediction. For the study on MIS data based PTB prediction, the MIS data from 84 recruited women at 18 to 35 weeks of gestation were used. Two calibration methods were proposed to pre-process the MIS data. Then, the support vector machine (SVM) classifier was applied to determine the association of MIS data with the PTB risk.

Results For the EIS data based PTB prediction, the best EIS-based PTB risk prediction model performed a testing AUC of 0.85 (95% CI: 0.77 – 0.93, sensitivity: 0.70, specificity: 0.87), and a training AUC of 0.81 (95% CI: 0.74 – 0.86, sensitivity: 0.85, specificity: 0.68). The best combined model by an EIS-based PTB risk prediction model and a maternal characteristics-

based PTB risk prediction model had a testing AUC of 0.89 (95% CI: 0.80 – 0.97, sensitivity: 0.60, specificity: 0.90), and a training AUC of 0.82 (95% CI: 0.76 – 0.89, sensitivity: 0.67, specificity: 0.85). For the MIS data based PTB prediction, two proposed calibration methods succeeded in MIS data pre-processing, and outperformed the recommended calibration method on building MIS-based PTB risk prediction models. The training AUCs of MIS-based PTB risk prediction models in this work ranged from 0.72 to 0.97, and the testing AUCs ranged from 0.29 to 0.78.

Conclusions The idea of building optimal-EIS spectrum filter model helped process the heterogeneous EIS records. After that, the EIS-based PTB risk prediction model and the combined model by an EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction model had excellent predicting abilities for both high-risk and low-risk women, within one to 23 weeks. The proposed MIS data calibration methods can reduce the MIS device's effects on MIS data, by removing components unrelated to cervix. MIS-based PTB risk prediction models had good training performances, so that MIS-based PTB risk prediction is worth a further study.

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor Professor Zi-Qiang Lang for the academic supports. I feel lucky to have such an experienced supervisor who showed a great patient and gave me the space to become as an independent researcher while providing great encouragement, enthusiasm, and guidance during my PhD. I also obtained the experience from being a graduate teaching assistant provided by Professor Zi-Qiang Lang.

I also would like to thank my colleagues, Sikai Zhang and David Tian, for their advice on my research, and must also thank the support staff of our University, Keira, for providing such excellent research supports.

Contents

List of Figures

List of Tables

List of Abbreviations

AC: alternating current

ACC: prediction accuracy

ADC: analog-to-digital converter

AHR: asymptomatic high risk

ARX: autoregressive with exogenous input

AUC: area under the receiver operating characteristic curve

BCC: point-biserial correlation coefficient

BIS: bioelectrical impedance spectroscopy

BMI: body-mass index

CCA: canonical correlation analysis

CL: cervical length

CV: coefficient of variation

DC: direct current

DDS: direct digital synthesis

ECCLIPPx II: electrical impedance prediction of preterm birth by spectroscopy of the cervixII

ECCLIPPx^{TM}: electrical impedance prediction of preterm birth by spectroscopy of the cervix

EHG: electrohysterography

EHRs: electronic health records

EIS: electrochemical impedance spectroscopy

EMG: electromyogram

ERR: error reduction ratio

ESR: error-to-signal ratio

fFN: fetal fibronectin

FPGA: field-programmable gate array

FROLS: forward regression with orthogonal least-squares

FROLS: forward regression with orthogonal least-squares

IL-10: interleukin-10

IL-6: interleukin-6

IL-8: interleukin-8

LR: logistic regression

MIS: magnetic impedance spectroscopy

MIT: magnetic induction tomography

NIHR: national institute for health research

NRMSE: normalized root mean square error

OBCC: orthogonal point-biserial correlation coefficient

OFR: orthogonal forward regression

OLS: orthogonal least-squares

PAMG-1: placental alpha macroglobulin-1

PCA: principal component analysis

phIGFBP-1: phosphorylated insulin-like growth factor binding protein-1

PPROM: preterm premature rupture of membranes

PPVs: positive predictive values

PTB: preterm birth

QUiPP: the quantitative innovation in predicting preterm birth

ROC: receiver operating characteristic

SMOTE: synthetic minority oversampling technique

sPTB: spontaneous preterm birth with intact membranes

STD: standard deviation

SVM: support vector machine

SYMP: symptomatic

TNF-α: tumor necrosis factor alpha

TSFRA: two-stage fast recursive algorithm

TSOLS: two-stage orthogonal least-squares

TVUS: transvaginal ultrasound

WHO: World Health Organization

Chapter 1

1. Introduction

1.1 Background

Preterm birth, in general, is defined by the World Health Organization (WHO) as all births at gestational ages less than 37 completed weeks or fewer than 259 days since the first day of a woman's last menstrual period [1]. It is a major cause of newborn death and the second direct cause of premature child death under the age of five years globally [2, 3]. 5% to 18% of births in 184 countries are premature, with 12% occurring on average in lower-income countries compared to 9% in higher-income nations [1]. Recent reports has indicated that global preterm birth rate for 2010 was 11.1%, equating to 14.9 million preterm newborns in 2010 [3, 4], and the estimated global preterm birth rate was 10.6% (14.84 million preterm newborns) in 2014 [3, 5]. Of these, at least one million newborns died as the result of premature and 7.6 million premature children below the age of five years died of PTB complications in 2010 [2].

Preterm birth is the most common cause of mortality, disability, and other issues in neonates. In addition, premature birth can have a severe impact on the entire lives of preterm kids [3] and have a significant impact on the lungs and brain [5]. Specifically, preterm neonates are more likely to develop asthma [6] and have a higher incidence of cerebral palsy, chronic illness, and learning impairment in adulthood than term newborns. The upshot of these disorders is that families and countries with preterm children must pay a substantial amount for their acute care, subsequent healthcare, and education needs [3]. Doctors must establish a preterm labour diagnosis in order to predict preterm birth and provide effective therapy to patients in light of the associated harms and consequences of early birth.

Aside from the physical damage on pregnant women and newborns, the PTB also causes the economic impact on their countries and families. As researchers at the Oxford Centre for Health

Economics announced, PTB cost the UK government an extra 939 million pounds each year, and the baby charity, Tommy's, had research showing that the total cost of all PTBs for one week in hospital is 260 million pounds annually [6]. Although exploring PTB has been a long history worldwide, as the cause of PTB is of individual and environmental factors, the etiology of PTB is still unknown [4]. In theory, the PTB can be prevented efficiently if pregnant women are under the strict medical supervision and care in hospital during their gestation. However, there is no doubt that this action will cause unnecessary interventions and social and economic costs.

1.2 Status and Limitations of existing PTB prediction

The attention on PTB including prediction, prevention and treatment currently has risen worldwide. Of these, the PTB risk prediction process is most essential, since only an efficient and accuracy prediction can guarantee a successful follow-up prevention and treatment. The existing prediction approaches mostly lean on etiology analysis, clinical symptoms and signs observation, clinical diagnostic tests, cervical length screening, and identification of risk factors and biomarkers [8]. Of these, the traditional clinical observation and diagnosis are based on the regular uterine contractions, but the studies have indicated that less than 10% of pregnant women with premature uterine contraction will give a true labour within seven days [9, 10].

Cervical length (CL) by transvaginal ultrasound is usually measured during 20 to 29 weeks of gestation and its predicted threshold generally sets 25 mm. The studies on the CL-based models have indicated that the second-trimester CL is an efficient predictor of PTB risk for higher-risk women with a multiple pregnancy, and cervical length have potential clinical use to predict birth within 48 hours and within seven days [14,15,16], however, the use of CL as a predictor applying in lower-risk pregnant women and predicting a longer-time (duration between test and delivery) delivery, has not been addressed so far and has been still a challenge [17-22].

Fetal fibronectin (fFN) is one of the biomarkers commonly used in clinical diagnosis of PTB. Like cervical length, fFN may be helpful with delivery within 21 days, its results reduce when the duration between test and delivery increases and fFN is most useful for delivery within

seven days [23, 24]. In addition, the fFN is a predictor for specific pregnant women/preterm births, but it can have little or no benefit to maternal hospitalization [23, 24, 25, 28]. A range of studies have also provided evidence to show that neither fFN nor cervical length can achieve the desired bedside demand on high accuracy of prediction of PTB risk [26, 27, 29, 30, 31].

Besides fFN biomarker, phosphorylated insulin-like growth factor binding protein-1 (phIGFBP-1) is another popular biomarker, which is measured in the vagina during 20 to 36 weeks' gestation [32,35]. Among recent studies, the phIGFBP-1 was pointed to have a low predictive ability for symptomatic pregnant women [32], similarly, another study also suggested that the positive phIGFBP-1 results can only contribute to the pregnant women with uterine contractions while the negative results cannot [33]. Furthermore, a few studies all confirmed that the phIGFBP-1 to predict a negative PTB result within 48 hours of testing for symptomatic women had the best performance [34], and the phIGFBP-1 had a better ability to predict for symptomatic women with short cervical length and at higher risk [36-39].

Placental alpha macroglobulin-1 (PAMG-1) was regarded as a standard biomarker test for detection of PTB risk in some institution [42]. In recent studies, all of them indicated that PAMG-1 can provide good performances on predicting PTB risk, with high specificities and positive predictive values (PPVs) for spontaneous premature delivery and pregnant women with a sign of PTB within seven days and 14 days of testing [39-44]. However, PAMG-1, like phIGFBP-1, is limited to the subgroup of PTB such as symptomatic women, since both PAMG-1 and phIGFBP-1 can generally be detected during uterine contraction or inflammation. Additionally, although cytokines typically including interleukin-6 (IL-6), interleukin-8 (IL-8), interleukin-10 (IL-10) and tumor necrosis factor alpha (TNF-α), were considered as potential PTB predictors in recent studies [45-53], the conclusions were not consistent and the association with PTB was equivocal so far [52, 53].

Apart from the above risk factors/biomarkers, cervical electrohysterography (EHG) is a relatively recent non-invasive measurement method that provides information on uterine contractions during pregnancy and labour by measuring the electrical activity responsible for the mechanical response [67, 68, 70, 71, 74]. Because signal amplitude is low, particularly during the early stages of pregnancy, and recordings may be corrupted by movement artefacts and contractions [66, 72]. These unwanted components can lead to erroneous interpretations of results [66, 72, 73]. As a result, obtaining high-quality surface signals is one of the most challenging aspects of the application of cervical EHG [66, 71, 78], and preprocessing and

presegmentation of the information are frequently required prior to analysis. In clinical diagnosis, cervical EHG recordings cannot be employed because of the complexity of the technology, the lack of a physical interpretation, and the modest performance on clinical information.

Recent years have seen a rise in the development of machine learning methods for predicting PTB risk. For example, a logistic regression model [20, 31] trained on the CL and fFN features achieved an area under Receiver Operating Characteristic Curve (AUC) of 0.67 (95% CI: 0.64-0.70). This method outperformed both traditional methods of predicting PTB risk using CL and fFN alone, with an average AUC of 0.59. According to [100], a demographic and obstetric features-based training set was created with feature selection method of linear correlation, and a variety of models were then trained using the reduced training set to estimate the probability of PTB, where AUCs ranged from 0.64 to 0.67 for the models. The QUantitative Innovation in Predicting Preterm birth (QUiPP) app is a predictive method for PTB based on CL, fFN, and other clinical risk variables [113, 114]. An AUC of 0.78 was attained by using the QUiPP app to predict PTB risk in 263 asymptomatic high-risk women in [115]. A logistic regression model with an AUC of 0.83 was then developed using the EIS and QUiPP app estimates of these women. Linear correlation was used to minimise the dimensionality of the training set, which included demographic and obstetric characteristics. The dimensionally reduced training set was used to train prediction models for PTB risk prediction using logistic regression, gradient boosting trees, neural networks, and ensemble approaches [116]. In [117], a validation set of women's demographic and obstetric history features was used to test five prediction models, and their AUCs ranged from 0.54 to 0.67 for the models.

Additionally, electronic health records (EHRs) including maternal demographic characteristics and clinical records of women, were fed into recent neural networks [63, 101] to predict extreme PTB risk (before 28 weeks of pregnancy). Although, it was found that the networks had an AUC of 0.83, this is not a cost-effective solution since it needed a big recruitment and a long-term data collection. From reviewing the current studies, the present approaches to predicting PTB risk are not able to achieve the desired accuracy, and collecting such vast amounts of EHR data is too expensive, as a result, it is not a sensible strategy.

1.3 Challenges of Predicting PTB

Predicting preterm birth is a complex and challenging task due to several factors. Here are some of the challenges involved:

1. Multiple causes: Preterm birth can be caused by several factors, including infections, genetic factors, lifestyle choices, and medical conditions. Identifying the specific cause of PTB can be difficult, and predicting the probability of PTB in individual cases requires consideration of multiple factors.
2. Variability in timing: PTB can occur at any time between 20 and 37 weeks of gestation. However, most PTBs occur between 32 and 37 weeks. Predicting the exact timing of PTB is challenging, and many studies on predicting PTB focused on prediction within 21 days.
3. Lack of reliable factors/biomarkers: Despite extensive research, there are no reliable factors/biomarkers that can accurately predict PTB. Several potential factors/biomarkers have been identified, including cervical length, fetal fibronectin, and biomarkers of inflammation, but none are accurate enough to be used routinely in clinical practice.
4. Data heterogeneity: The quality of PTB data from measurements or tests is modest. For example, one woman would have several measurements or tests at the same visit, however, the results of these measurements or tests would be different. The different results can lead different performances on training PTB prediction models.
5. Weak generalizability of models: Most of existing PTB prediction models focused on specific populations, in terms of age range, race, PTB risk level and so on.

1.4 Aims and Objectives

Many existing studies have shown that identification of a single risk factor/biomarker to predict PTB risk has not been evolved so far, and the discovery of novel risk factors/biomarkers related with PTB risk, and the use of multiple predictors were strongly advised [8, 11, 12]. The

development of multivariate prediction models based on large-scale-data-analysis techniques to improve diagnostic efficiency and prediction accuracy, was also advised in current studies.

The project of ECCLIPPx^{TM} is a novel investigation into the prediction of preterm birth located in Sheffield. Its aim was to determine if the potential Electrical Impedance Spectroscopy test could accurately identify women of preterm birth in order to give them treatments to prolong pregnancy. After that, with the National Institute for Health Research (NIHR) funding, the investigators have also proposed a feasibility study of ECCLIPPx II on a potential approach of Magnetic Impedance Spectroscopy to increase the accuracy of EIS and decrease potential sources of error.

Based on these hypothesises proposed by the project of ECCLIPPx^{TM} and in order to address the challenges of existing studies, the present study aimed to focus on the study of PTB risk prediction based on machine learning techniques using novel potential biomedical signals including cervical EIS signal and cervical MIS signal. The objectives were to develop the data processing techniques for cervical EIS signals and cervical MIS signals as well as the machine learning based data analysis approaches with its strong capability of large-scale-data processing and analysis, for creating EIS-based and MIS-based PTB risk prediction models.

The research investigations include three main works which are:

(1) EIS data processing and EIS data only based PTB risk prediction

(2) EIS and maternal characteristics-based PTB risk production, and

(3) Feasibility study on the MIS-based PTB risk prediction.

1.5 Structure of Main Works

For work (1) in Chapter 4, I implemented data pre-processing for EIS signal including the comprehension of data size, data construction, missing data and statistics, and then evaluated the data quality and proposed a principle of the optimum-EIS spectrum filter model, which was created to select reliable and high-quality EIS data. Then I proposed a machine learning

pipeline for training the EIS-based PTB risk prediction model, including employing polynomial feature structure to extend the feature dimension, and introducing a range of feature selection techniques to evaluate the relation between feature and PTB risk, and then selecting related features to train logistic regression model. In this work, three kinds of EIS-based PTB risk prediction models were trained, and a general machine learning PTB risk prediction model was trained to compare with EIS-based PTB risk prediction models, to assess the improvements given by the optimum-EIS spectrum filter model and the proposed EIS-based PTB risk prediction models.

For work (2) in Chapter 5, based on the previous outcomes of analysis of EIS-based PTB risk prediction, an attempt of combination of the EIS-based PTB risk prediction model and the maternal characteristics-based PTB risk prediction model was made, and the combined model was compared to PTB risk prediction models based on only EIS, maternal characteristics, CL and fFN. First, three machining learning pipelines of training the maternal characteristics-based PTB risk prediction model, the CL/fFN-based PTB risk prediction model, and the EIS-based PTB risk prediction model, were proposed. After these models trained, two approaches of combining probabilities of PTB from different PTB risk prediction models were proposed based on LR and SVM, respectively.

For work (3) in Chapter 6, before applying machine learning to MIS data to build a model for PTB risk prediction, directly obtained MIS data (called real-time MIS data) was pre-processed to calibrate MIS measurements with air and ferrite references, which were subtracted from the MIS spectrum. A new calibration method was proposed in this work, that divided the MIS spectrum by air and ferrite references and used the calibrated MIS spectrum to predict PTB risk. Additionally, unlike EIS signals, the raw MIS signals were collected at the time domain. An Autoregressive with Exogenous Input Model (ARX) was built using raw MIS data to represent the dynamic system, and only features representing cervical magnetic impedance rather than MIS device circuit properties in the ARX model, were used for PTB prediction. After the MIS data calibration, a machine learning pipeline for training the MIS-based PTB risk prediction model was proposed using SVM model, to evaluate the three calibration methods and the potential of PTB prediction using MIS data.

1.6 Contributions of This Thesis

Firstly, as for EIS data based PTB prediction, I noticed the issue of EIS data's heterogeneity, which may lead a challenge to machine learning based PTB risk prediction, if averaging the heterogeneous EIS data to build machine learning models. In this case, I implemented the EIS data statistical analysis and proposed a principle of identifying the optimal EIS spectrum. This principle was to identify the optimal EIS spectrum depending on delivery results. To deal with this problem, I added labels for EIS spectra, which were used to train optimal-EIS spectrum filter model without using delivery results.

Then, a machine learning strategy of training EIS-based PTB risk prediction model was proposed in this work. Borderline-SMOTE1 oversampling [112] was used to balance the EIS data. The logistic regression model was chosen as the EIS-based PTB risk prediction model due to its good interpretability, and the application of polynomial feature structure tried to deal with the problem of non-linear decision boundaries. However, polynomial feature structure can lead a huge number of features. Orthogonal point-biserial correlation coefficient [105], canonical correlation analysis [108] and two-stage canonical correlation analysis were applied to evaluate the correlation between EIS data and PTB risk, and to reduce the dimension. After the EIS-based PTB risk prediction model trained, a models-combining approach was proposed, applying a machine learning model to combine two probabilities obtained by an EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction model, so that the accuracy prediction improved.

As for MIS data based PTB prediction, since the MIS data were measured by different probes respectively, the representations of ARX models built using raw MIS data were also different. To deal with this problem, the common component of ARX models was assumed as the representation of cervix to be remained by pole-zero analysis. This implementation was calibration via modelling. Because real-time MIS data were complex numbers, real-time MIS data divided with references to calibrate corresponded with physical significance of complex numbers. Therefore, calibration via division was proposed in this study. After the MIS data calibrated, the machine learning was applied to evaluate the feasibility of MIS data based PTB prediction.

In conclusion, this study contributed the idea of optimal-EIS spectrum filter model to overcome the problem of EIS data's heterogeneity, contributed the EIS-based PTB risk prediction model and the combined model by combining an EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction, where the two efficient PTB risk prediction models had excellent predictions, were generalized for a wide population, and prolonged the prediction period from 21 days to 12 weeks on average. Finally, the calibration via modelling contributed the idea of modelling for a target component to reduce the device's effects on measurements. The performances of MIS-based PTB risk prediction models indicated that MIS data were related to PTB risk, and a further study is needed to improve the prediction.

Chapter 2

2. Literature Review

2.1 The Knowledge of Preterm Birth

Preterm birth is all births at gestational ages before 37 completed weeks or fewer than 259 days since the first day of a woman's last menstrual period [1]. Furthermore, preterm birth can be further sub-categorized based on the gestational age [1]: extremely preterm (<28 completed weeks of gestation), very preterm (28 to 32 weeks of completed weeks of gestation), moderate preterm (32 to 34 completed weeks of gestation), and late preterm birth (34 to 37 weeks of completed weeks of gestation).

Also, preterm birth can be classified into three subtypes by obstetric precursors [81-83], which are spontaneous preterm birth with intact membranes (sPTB), preterm premature rupture of membranes (PPROM) and medically indicated preterm birth or induced preterm birth. 40-45% of preterm births are sPTB; PPROM occupies 25-30% and 30-35% are induced preterm birth [81, 83]. In general, sPTB and PPROM are together regarded as spontaneous preterm births. Further, there was a study summarizing the currently recognised risk factors of PTB [81, 82], where the sPTB is associated with previous PTB history, low body mass, strenuous physical workload, adverse behaviours, infection, uterine and cervical anomalies and psychosocial stress, while medically induced PTB is led by maternal risk factors including pregnancy hypertension, chronic conditions, obstetrical complication and antepartum bleeding, and by fetal risk factors including pregnancy history, fetal distress, unstable fetal condition and intrauterine growth restriction [81-83].

Furthermore, the high risk of PTB and low risk of PTB are identified by if the woman has one or more risk factors, for example, previous preterm birth, multiple pregnancies, short intervals between pregnancies, infections during pregnancy, short cervical length and smoking, alcohol,

or medicine use during pregnancy [82]. The asymptomatic high risk (AHR) woman is defined if the woman is sPTB and has the previous obstetric history or a short cervical length at 20 weeks of gestation. The symptomatic (SYMP) woman is defined if the woman has potential chances of PTB but without advanced cervical dilation [81-83].

The exploration of PTB including prediction, prevention and treatment currently depends on five key points [11] including:

1) A better comprehension and further research of the etiology of PTB
2) The identification of risk factors and determining the individual risk for pregnant women
3) Medical supervision and interventions for at-risk pregnant women
4) Necessary preventions for pregnant women with early symptoms, and
5) Improvement of treatments through clinical trials for high-risk pregnant women.

Theoretically, the first two points can be considered as the prediction of PTB risk, 3) and 4) are categorized into prevention process. and the last point is the treatment for PTB. Of these, the prediction process is most essential, since only an efficient and accurate prediction can guarantee the subsequent processes of prevention and treatment to be successful.

2.2 The Classification System of Approaches to Predicting Preterm Birth Risk

One current approach to the prediction of PTB risk is based on the application of risk factors/biomarkers, which was proposed in the study by Dr. Harry Georgiou and his colleagues from Department of Obstetrics and Gynaecology in University of Melbourne. They also advised that the current potential risk factors/biomarkers to predict PTB risk can be divided into three sub-categories [11]. First, the maternal risk factors are composed of maternal

demographic characteristics, nutritional status, pregnancy history, present pregnancy characteristics, psychological characteristics, adverse behaviours, infection, uterine contractions, and body-mass index (BMI) and so on. Of these, gynecologic history of pregnant women, especially previous PTB history, is a high-risk factor to predict PTB. Second, cervical length is popular and has been applied in clinical diagnosis of PTB, especially after Dr. Jackson of University of Utah in 1992 proposed to measure cervical length using transvaginal ultrasound (TVUS) [80]. Finally, the biological and genetic markers including amniotic fluid, saliva, urine, blood, fetal fibronectin phosphorylated insulin-like growth factor binding protein-1 and placental alpha macroglobulin-1 and so on, are becoming relatively credible for detection of PTB.

Another popular approach to predicting PTB risk was to use single or multiple risk factors/biomarkers-based prediction models [11, 12]. First, the prediction of PTB is conducted with single risk factor/biomarker such as, for example, CL, fFN, cytokines containing interleukin-6, interleukin-8, interleukin-10 and tumor necrosis factor alpha, phosphorylated phIGFBP-1, and PAMG-1 [8, 12]. Second, multiple risk factors/biomarkers are introduced into prediction models, that combines previous PTB history, low body mass, adverse behaviours, infection, maternal and fetal risk factors/biomarkers, with the data analysis techniques to build appropriate prediction models.

A further approach to predicting of PTB risk involves biomedical signals such as cervical EHG signal which is different from the traditional risk factors/biomarkers. Hence, current prediction models of predicting PTB risk can be mainly classified into three types: i) single-risk factor/biomarker-based prediction models, ii) multiple-risk factors/biomarkers-based prediction models and iii) new biomedical-signal-based prediction models. These models will be introduced in details as follows.

2.3 The Current Prediction Models of Preterm Birth Risk

i) Single-risk factor/biomarker-based prediction models

Many studies have concentrated on predicting PTB with single risk factor/biomarker that is all advised by clinical representations, typically including CL, fFN, phIGFBP-1, PAMG-1, and some cytokines containing IL-6, IL-8, IL-10 and TNF-α [3, 8, 12, 13]. Table 2.1 is the summary of a part of current studies where clinical risk factors/biomarkers are used to predict PTB risk.

Table 2.1: The summary of current studies applying single-risk factor/biomarker-based prediction models to predict PTB within 14 days.

Features	Author/Year		Comments
CL	Honest et al./2003 [14] Lim et al./2011 [15] Boots et al./2014 [16] Vandermolen et al./2016 [17] Melamed et al./2014 [18]	Conde et al./2015 [19] Esplin et al./2017 [20] Hiersch et al./2016 [21] Hiersch et al./2014 [22]	The studies showed that CL alone may not always be an accurate predictor of preterm birth, and the predictive value of CL alone is highest in women with a history of preterm birth or other risk factors, but it may be less accurate in low-risk populations.
fFN	Leitich et al./1999 [23] Deshpande et al./2013 [24] Berghella et al./2008 [25] Zhou et al./2015 [26]	Abbott et al./2012 [27] Jwala et al./ 2016 [28] Hezelgrave et al./2016 [29] Rizzo et al./1996 [30]	The studies showed that the accuracy of the fFN test is highly dependent on testing between 22 and 35 weeks of gestation, similarly fFN is a useful biomarker for predicting preterm birth in high-risk populations.
phIGFBP-1	Fuchs et al./ 2017 [32] Goyal et al./2016 [33] Conde et al./ 2016 [34] Cooper et al./2012 [35]	Riboni et al./2011 [36] Danti et al./2011 [37] Brik et al./2010 [38] Melchor et al./2018 [39]	The studies showed that although phIGFBP-1 and PAMG-1 have shown promising results in some studies, there is still limited data on their efficacy as predictors of preterm birth, especially in low-risk populations. More research is needed to fully evaluate the clinical utility.
PAMG-1	Melchor et al./2018 [39] Wing et al./2017 [40] Nikolova et al./2017 [41]	Melchor et al./2018 [42] Nikolova et al./2014 [43] Ravi et al./2019 [44]	

IL-6, IL-8, IL-10 and TNF-α	Yoneda et al./2011 [45] Yoneda et al./2007 [46] Ng et al./2006 [47] Thomakos et al./2009 [48]	Holst et al./2009 [49] Son et al./ 2016 [50] Huang et al./ 2019 [51]	The studies showed that the levels of cytokines can vary widely among individuals, making it difficult to establish universal cut-off values or models for identifying high-risk individuals, and the positive predictive value of cytokine testing is generally low.

CL: cervical length, fFN: fetal fibronectin, phIGFBP-1: phosphorylated insulin-like growth factor binding protein-1, PAMG-1: placental alpha macroglobulin-1, IL-6: cytokines containing interleukin-6, IL-8: cytokines containing interleukin-8, IL-10: cytokines containing interleukin-10, TNF-α: tumor necrosis factor alpha.

Screening of cervical length by transvaginal ultrasound has long been used as the criterion for the prediction of PTB risk [12, 13], its predicted threshold is the length of cervix generally less than 25 mm under vaginal ultrasonography, and its measurement time is usually recommended at 20 to 29 weeks of gestation [11]. The pregnancy with the length of cervix more than 30 mm is often at low risk of PTB, while the pregnancy with the length of cervix less than 15 mm is most likely at high risk of PTB [14, 17, 19]. Table 2.1 includes three review studies and six articles about CL-based model [14-22]. Of these, a review of 21 studies from 1966 to 2009, among women with a multiple pregnancy, sensitivity and specificity of choosing different second-trimester CLs as criterion for spontaneous PTB before 34 weeks' gestation, were 78% and 66% for 35 mm of CL, 41% and 87% for 30 mm, 36% and 94% for 25 mm and 30% and 94% for 20 mm [14, 15]. Another systematic review between 1966 and 2013 summarized that the sensitivity and specificity of CL for delivery within 48 hours were 0.77 (95% CI, 0.54-0.90) and 0.88 (95% CI, 0.84-0.91), and for delivery within seven days were 0.74 (95% CI, 0.58-0.85) and 0.89 (95% CI, 0.85-0.92) respectively [16].

Fetal fibronectin is one of the biomarker tests commonly applied in clinical prediction of PTB risk, since some studies found that fFN of at PTB risk women is higher than normal pregnant women [23]. When the fFN test is positive and more than 50 ng/ml, the intervention should be taken to prevent from PTB [23, 24, 29]. Among the summary studies about fFN in Table 2.1, the statics given by review research of 27 studies from 1989 to 1997, showed that overall sensitivities and specificities of fFN for delivery <37 and <34 weeks' gestation were 56% and 84%, and 61% and 83% respectively [16], meanwhile it presented that the sensitivities of fFN for delivery within 7, 14, 21, 28 days were 76%, 68%, 61%, and 43%, also the specificities

were 86%, 86%, 88%, and 83% respectively [16]. Another summary of studies between 2000 and 2011 reviewed that the sensitivities and specificities of fFN for delivery <37 and <34 weeks' gestation within 7-10 days of testing were 60.8% and 82.3%, and 69.1% and 84.4% [17]; the sensitivities and specificities of fFN for delivery within 7-10 days were 76.7% and 82.7% [17].

The phIGFBP-1 is regarded as a phosphorylated protein that is generally synthesized in endometrium cells [32]. These cells are decidualized in the uterine during pregnancy [35]. phIGFBP-1 normally cannot be found in the vagina, while it will leak into the vagina if uterine contracts, and it usually can be measured at 22 to 36 weeks of gestation [35]. The summary studies of phIGFBP-1 in Table 2.1 contains two systematic reviews and six articles of phIGFBP-1 analysis [32-39]. A review of forty-three studies from 2000 to 2015 were analysed, of which 15 studies about phIGFBP-1 test for asymptomatic pregnant women performed, that the pooled sensitivities and specificities for delivery <37, <34, and <32 weeks' gestation ranged from 14% to 47% and from 76% to 93%, while 34 of 43 studies about phIGFBP-1 test for pregnant women with a sign of PTB showed that the pooled sensitivities and specificities for delivery <37 and <34 weeks' gestation within 7 and 14 days of testing were 60% and 68%, 77% and 81% respectively [34].

Like phIGFBP-1, PAMG-1 is also generated from decidualized endometrium cells during pregnancy and most amounts of PAMG-1 are in amniotic fluid, additionally once early uterine contraction or inflammation associated with PTB happens, PAMG-1 will appear in the vagina and be detected [40, 43]. The recent study reviewed 14 articles from 2015 to 2017 and found that the lowest and highest sensitivity and specificity of PAMG-1 for predicting sPTB within seven days of testing were 0.33 (95% CI, 0.04-0.78) and 0.91 (95% CI, 0.79-0.97), and 1.00 (95% CI, 0.74-1.00) and 1.00 (95% CI, 0.94-1.00), respectively [39]. Similarly, other studies in Table 2.1 showed that the PAMG-1 performed a high specificity and positive predictive value (PPV) [40-44].

IL-6, IL-8, IL-10 and TNF-α are a range of cytokines which are secreted in amniotic fluid, maternal serum and cervicovaginal fluid, response to the infections and inflammatory [12, 52, 53]. Several studies of relation between cytokines and PTB risk from 2004 to 2016 were collected as a review in Table 2.1 [45-51], whereas their conclusions were various. First, some studies indicated that the amniotic fluid IL-6 is strongly associated with PTB and with a p-value < 0.001 [47, 48, 49, 50], while the review studies found many studies pointed that amniotic fluid IL-6 has no association with PTB [51]. Besides, three studies confirmed that

amniotic fluid IL-8 has association with PTB and stronger than cervical mucus IL-8 [45, 46, 50], whilst another review study denied the correlation with PTB [12], similar to IL-10 and TNF-α [48-53].

ii) Multiple-risk factors/biomarkers-based prediction models

Many studies announced that single risk factor/biomarker to predict PTB risk cannot meet the requirement of accuracy [5, 8, 11], hence, they suggested that the further studies should involve combination of risk factors/biomarkers and large data analysis techniques [11, 12]. For example, some studies implied that the fFN test should be applied with cervical length, where the pregnant women with cervical length less than 25 mm are generally at high risk of PTB and should implement the fFN test [17, 20, 26, 28, 31]. Furthermore, some studies also added phIGFBP-1 and PAMG-1 into the combination of fFN and cervical length, and the combination of cytokines [32, 36, 37, 39, 40, 41, 42, 44], to improve the accuracy of prediction models.

Table 2.2: The summary of current studies applying multiple-risk factors/biomarkers-based prediction models to predict PTB within 21 days; some performances are reported as ranges since the studies used several feature sets or classifiers.

Author/Year	No. of features	Data analysis Techniques	Classifiers	Performances
Amini et al./2017 [54]	14/5	(1) Stepwise variable selection	(1) Decision tree (2) Logistic regression	AUC on clinical data: 0.59-0.85
Alleman et al./2013 [55]	9/11	(1) Continuous net reclassification improvement (2) Integrated discrimination improvement	(1) Logistic regression	AUC on clinical data: 0.54-0.70
Weber et al./2018 [56]	20	(1) Super learning	(1) Original logistic regression	AUC on clinical data: 0.60-0.68

			(2) Random Forest (3) K-nearest neighbors (4) Generalized additive models (5) Lasso regression (6) Ridge regression (7) Elastic net regression	
García et al./2017 [57]	7	(1) Parametric survival model	(1) Logistic distribution	AUC on clinical data: 0.63
Tran et al./2016 [58]	10	(1) Weighted feature standard deviation	(1) Stabilized Sparse Logistic Regression (2) Randomized Gradient Boosting	AUC on clinical data: 0.70-0.77
El-Ardat et al./2014 [59]	5	(1) Mann-Whitney test (2) Chi-Squared test	(1) Logistic regression	AUC on clinical data: 0.88
Zhu et al./2017 [60]	10	(1) T test (2) x^2 test (3) Correction x^2test	(1) Cut-off values	ACC on clinical data: 0.79
He et al./2018 [61]	3-5	(1) Harrell's c statistics (2) Martingale residuals (3) Log survival plots (4) Schoenfeld residuals	(1) Cox proportional hazards regression	AUC on clinical data: 0.51-0.56
Raja et al./2021 [62]	36	(1) Entropy notion (2) Synthetic minority oversampling technique (SMOTE)	(1) Decision tree (2) Logistic regression (3) Support vector machine	ACC on clinical data: 0.78-0.91

AUC: area under the receiver operating characteristic curve, ACC: prediction accuracy.

In Table 2.2, I summarized the recent studies of predicting PTB risk with multiple-risk factors/biomarkers-based models [54-62]. Of these, the summary concentrates on the combination of features techniques and classifiers applied for feature selection and prediction. Furthermore, the performances are displayed with the prediction accuracy (ACC) and area under the receiver operating characteristic (ROC) curve (AUC).

Among the studies, the combination of maternal demographic characteristics only, such as age, history of PTB, amount of vaginal-bleeding, adverse lifestyle and folic acid intake before pregnancy, cannot construct a high-quality predictive model [61]. The various combinations of demographic characteristics and risk factors/biomarkers, which include CL, fFN and IL-1β [60], amine odortest [59], and cortisol and α-amylase level [57], can improve the association with PTB risk. Several review studies summarized a large amount of articles about PTB risk prediction and claimed that the previous PTB, age of pregnancy, BMI, chronic disease and adverse habits are most frequent predictors [63, 65], and the multivariate prediction models are able to improve the accuracy whereas the overall prediction performances are not consistent, the AUC ranging from 0.51 to 0.91 [63-65].

iii) New biomedical-signal-based prediction models

Electrophysiological technique is a measurement of electrical activity in different parts of body, such as brain, eye, heart, muscle, uterus and skin. The electrophysiological technique applied in uterus is named as electrohysterography, also as uterine electromyogram (EMG), which measures the electrical change in uterus where movements happen [66, 71, 74]. In addition, EHG/EMG technique is regarded as a safe measurement that does not hurt pregnant women, and it is convenient and flexible to measure during treatment [64, 79]. Hence, a range of studies focused on developing EHG/EMG technique to predict PTB risk.

In general, the EHG/EMG signal mainly composes of linear features and nonlinear features or time domain features and frequency domain features [64, 66, 75, 76]. Of these, linear features can be the root mean square, peak frequency, median frequency, variance, average amplitude change, mean absolute value, Wavelet length, frequency ratio and maximum fractal length of EHG/EMG signal and so on [64, 66, 69, 70, 72], and nonlinear features include maximal Lyapunov exponent, sample entropy, integrated EHG signal, simple square integral, singular

spectrum analysis, principal component analysis (PCA), correlation dimension, Coupling and multivariate analysis and so on [64, 70, 77, 78]. Then these features will be assessed by kinds of feature selection techniques to determine prediction models.

Table 2.3: The summary of current studies applying new biomedical-signal-based prediction models to predict PTB within 21 days; some performances are reported as ranges since the studies used several feature sets or classifiers.

Feature	Author/Year	No. of features	Data analysis Techniques	Classifiers	Performances
EHG/EMG	Fergus et al./2013 [67]	4	(1) Synthetic minority oversampling technique (SMOTE)	(1) Linear Discriminant Classifier (2) Quadratic Discriminant Classifier (QDC) (3) Uncorrelated Normal Density Classifier (4) Polynomial Classifier (5) Logistic Classifier (6) K-Nearest Neighbor (7) Decision Tree Classifier (8) Parzen Classifier (9) Support vector machine	(1) AUC on open data: 0.50-0.61 (no SMOTE) (2) AUC on open data: 0.66-0.89 (SMOTE) (3) AUC on clinical data: 0.49-0.55 (no SMOTE)
	Fergus et al./2016 [68]	4	(1) Synthetic minority oversampling technique (SMOTE) (2) Statistical significance (3) Linear discriminant analysis using independent search	(1) Back-propagation trained feed-forward neural network classifier (2) Levenberg–Marquardt trained feed-forward neural network classifier (3) Perceptron linear classifier	(1) AUC on open data: 0.50-0.58 (no SMOTE) (2) AUC on open data: 0.50-0.90 (SMOTE) (3) AUC on clinical data:

		(4) Linear discriminant analysis using forward search (5) Linear discriminant analysis using backward search (6) Gram-Schmidt	(4) Radial basis function neural network classifier (5) Random neural network classifier (6) Voted Perceptron classifier (7) Discriminative Restricted Boltzmann Machine classifier	0.50-0.91 (SMOTE)
	Ahmed et al./2017 [69]	<= 8	(1) Huang-hilbert transform (2) Intrinsic mode function (3) Inter-correlation coefficient	(1) Support vector machine classifier	(1) AUC on clinical data: 0.87-0.96
	Ren et al./2015 [70]	180 or 4	(1) Synthetic minority oversampling technique (SMOTE) (2) Empirical Mode Decomposition (EMD) (3) Intrinsic mode function (4) Shannon entropy	(1) Support Vector Machine (2) Random Forests (RF) (3) Multilayer Perception (4) AdaBoost (5) Bayesian Network (BN) (6) Simple Logistic Regression	(1) AUC on open data:0.50-0.91 (non-EMD) (2) AUC on open data: 0.76-0.98 (EMD)
	Idowu et al./2015 [71]	4	(1) Synthetic minority oversampling technique (SMOTE)	(1) Random Forest (2) Penalized Logistic Regression (3) Rule-Based Classifier	(1) AUC on open data: 0.49-0.68 (no SMOTE) (2) AUC on open data: 0.89-0.96 (SMOTE)
	Smrdel et al./2015 [72]	<=6	(1) Synthetic minority oversampling technique (SMOTE)	(1) K-nearest neighbors (2) Linear discriminant analysis	(1) ACC on open data: 0.49-0.88

			(2) Adaptive autoregressive	(3) Quadratic discriminant analysis (4) Support vector machine (5) Decision tree	
	Naeem et al./2014 [73]	N/A	(1) Discrete cosine transform (2) Singular spectrum approach (3) Amplitude adjusted Fourier transform (4) Principal component analysis (PCA)	(1) Kohonen neural network (2) Cascade-forward neural network (3) Feed-forward neural network	(1) ACC on open data: 0.65-0.92
	Hussain et al./2015 [74]	4	(1) Min/max oversampling technique	(1) Multi-Layer Perceptron (2) Self- organized Network Inspired by the Immune Algorithm (SONIA) (3) K-nearest neighbors (4) Decision tree (5) Support vector machine (6) Dynamic Self-Organized Network Inspired by the Immune Algorithm	(1) ACC on open data: 0.65-0.92

EHG: electrohysterography, EMG: electromyogram, AUC: area under the receiver operating characteristic curve, ACC: prediction accuracy.

Table 2.3 summarized a range of current studies concentrating on exploring EHG/EMG signal for predicting PTB risk [67-74]. The summary includes techniques and classifiers applied in those studies, and prediction performance of EHG/EMG. The techniques employed in studies involve over sampling technique, feature extraction technique, feature selection technique and a range of data analysis techniques. The classifiers included the popular machine learning techniques that are widely applied in prediction and classification problems. Also, the

performances were summarized with ACC or AUC using open data or clinical data with consideration of all results via different classifiers.

2.4 Problems with Existing Preterm Birth Predictions

Firstly, prediction of PTB risk with risk factors/biomarkers alone consists of CL, fFN, phIGFBP-1, PAMG-1, and certain inflammatory mediators containing IL-6, IL-8, IL-10 and TNF-α. As for CL and fFN, the use of CL or fFN as a predictor in low-risk pregnant women and the prediction of a longer-time (time from test to delivery) delivery has not yet been addressed and remains a difficulty [14, 15, 16]. Neither fFN nor CL can achieve the desired bedside demand on high accuracy of prediction of PTB risk [26, 27, 29, 30, 31]. As for other risk factors/biomarkers, PAMG-1 and phIGFBP-1 are restricted to the subset of PTB that includes sPTB and symptomatic women, as both are typically found during uterine contraction or inflammation [39, 40, 41, 44]. In addition, although cytokines, such as IL-6, IL-8, IL-10 and TNF- α, were considered as potential PTB predictors in recent studies. the conclusions were inconsistent and the association with PTB remained ambiguous yet [52, 53].

Then, the combination of maternal demographic factors cannot develop a high-quality predictive model, while prediction of PTB with combination of risk factors/biomarkers has the superior performance on clinical data. However, the overall prediction performances vary widely, with AUC ranging from 0.51 to 0.91 [63-65]. This means that, on the one hand, while employing the same methodologies or features to develop prediction models for different training data of pregnant women, these models may exhibit varied performance on the same testing data [57, 59]. Alternatively, the performance of a model may vary depending on the testing data of pregnant women. This is also likely to indicate that certain models or features cannot guarantee predictive universality and generalization [61, 63]. Also, most of PTB risk prediction models from current works can only have the best performance before 7 to 14 days of onset [58, 60, 63, 64, 65].

Cervical EHG is a relatively new non-invasive measurement tool that offers information on uterine contractions during pregnancy and labour by monitoring the electrical activity

responsible for the mechanical response [66]. But obtaining high-quality surface signals is one of the most difficult tasks in the application of cervical EHG. Signal amplitude is modest, especially in the early stages of pregnancy, and recordings might be distorted by movement artefacts and contractions [66, 71, 78]. These unwanted components can result in erroneous interpretations of outcomes. Therefore, the analysis often needs pre-process and pre-segmentation of the records [72, 73]. These complicated implements and lack of a physical interpretation lead the EHG records not to be used on clinical diagnosis [66]. In conclusion, almost none of these methods can consistently define premature birth, resulting in ineffective treatments, missed opportunities to treat preterm infants, and an inability to conduct therapeutic effect analyses.

Chapter 3

3. Electrical Impedance Spectroscopy and Magnetic Impedance Spectroscopy

3.1 Electrical Impedance Spectroscopy Technology

Background Electrical Impedance Spectroscopy is a technology involving applying a sinusoidal test voltage or current to a sample under test to measure its transfer impedance over an appropriate frequency range [84, 85, 86, 87]. Of these, applying a sinusoidal test voltage is called 'potentiostat EIS', which is one of its two versions, and another version called 'galvanostat EIS' is applying a sinusoidal test current. Generally, these two versions are equal and give the same results. The EIS technique has been employed to study the electrical properties of various materials, where the measured electrical impedance spectra show an electrical image of the sample under test, often with an equivalent circuit mode, representing its properties and behaviour [85]. Additionally, this quick and easily-implement technique has been widely applied to find out the characters of solid, liquid, semi-liquid, animal and human organic, such as monitoring of fruits, vegetable oils, dairy products, and especially metal surfaces [85].

EIS device Characterizing of tissue of upper cervix has been ongoing by a study group of ECCLIPPx^{TM}. The ECCLIPPx^{TM} study was carried out in the Jessop Wing Maternity Unit of the Royal Hallamshire Hospital in Sheffield, UK, a tertiary PTB referral institution with about 7500 deliveries annually, between January 2014 and August 2016. The study received approval from the Yorkshire & Humber (Sheffield) Committee of the UK National Research Ethics Service (REC Number 13/YH/0167). All study participants gave written informed consent.

The EIS technique was implemented by an advanced device namely Sheffield Mark V (shown in Figure 3.1), which is the latest version and is developed to reduce the noise and increase the sensitivity of measurement on the deep of tissue [86, 89]. The researchers of ECCLIPPx^{TM} group hypothesised that the impedance changes recorded by Sheffield Mark V device in cervical tissue in the period of pregnancy will associate with PTB risk [86, 89, 91].

In the ECCLIPPx^{TM} study [86, 89], the device has an accelerometer and a pressure sensor, and researchers applied a constant force of 2 Newtons to take measurements of cervical EIS. The probe tip has eight electrodes arranged in two rings of different sizes, and the electrodes are made of 37.5% gold with varying diameters. The overall diameter of the probe tip is 11mm, and the device measures and displays the applied force to the clinician using a load cell and a triaxial accelerometer.

Figure 3.1: The EIS device of Sheffield Mark V, showing the device base, and the device probe piece (a), the tetrapolar probe configuration (b) and the complete coupled device (c).

EIS measurement The transfer impedance was measured (Figure 3.2) with the device by applying sine-wave current (Equation 3.1) at 14 different frequencies ranging from 76.3Hz to 625 kHz in octave increments through two injecting electrodes, and then the device measured the electric voltage change (Equation 3.2) between two sensing electrodes. Each frequency

sweep took 200 milliseconds and was repeated until the standard error of eight measurements is under a threshold for quality control purposes. This means that it took a minimum of 1.6 seconds and 3.2 seconds to record the full-frequency spectrum for the two electrode rings, respectively. The measured transfer impedance spectral data (Equation 3.3) and its variance were transmitted via Bluetooth technology to a custom central database and stored.

$$I(t) = \bar{I} + |I| \times sin(\omega t) = \bar{I} + |I| \times sin(2\pi f t), \quad (3.1)$$

$$V(t) = \bar{V} + |V| \times sin(\omega t + \phi) = \bar{V} + |V| \times sin(2\pi f t + \phi), \quad (3.2)$$

$$Z(j\omega) = \frac{V(j\omega)}{I(j\omega)} = \frac{|V|}{|I|} \times e^{-j\phi} = |Z| \times e^{j \times Arg(Z)} \quad (3.3)$$

$$= Re(Z(j\omega)) + j \times Im(Z(j\omega)),$$

where, $\bar{I}$ and $\bar{V}$ are the direct current (DC) voltage and DC current; $|I|$ and $|V|$ are the amplitude of voltage and current at the angular frequency $\omega = 2\pi f$, where f is the physical frequency (also the signal frequency); ϕ is the phase difference between $I(t)$ and $V(t)$; $I(j\omega)$ and $V(j\omega)$ are the Fourier transform of the $I(t)$ and $V(t)$; $|Z|$ and $Arg(Z)$ is the amplitude and phase of the impedance; and $Re(Z(j\omega))$ and $Im(Z(j\omega))$ are the real and imaginary part.

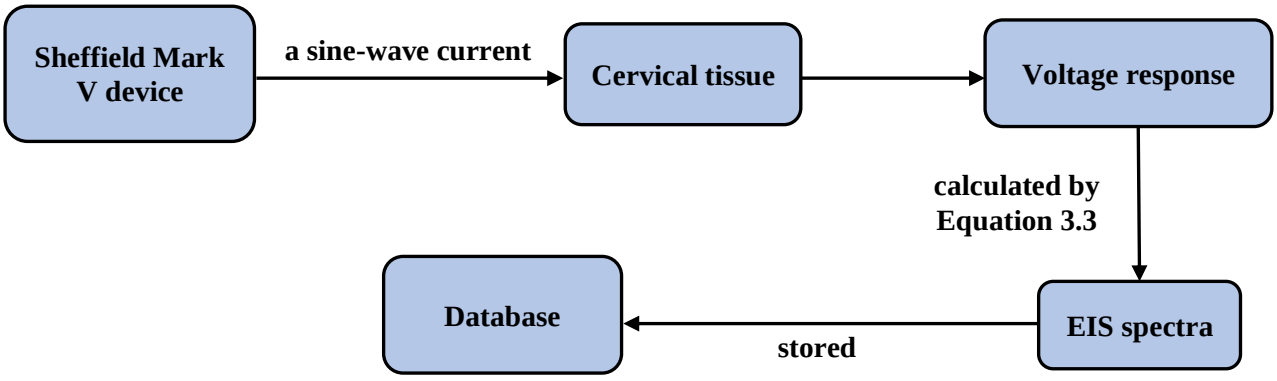

Figure 3.2: The proposed technology for measurement of EIS.

Finally, the measured cervical impendences (also EIS features) from ECCLIPPx^{TM} group, are over 14 frequencies ranging from 76.3 Hz to 625 kHz,

$$Z(j2\pi f_i) = |Z_i| \times e^{j \times Arg(Z_i)} = Re(Z_i) + j \times Im(Z_i), \quad (3.4)$$

where $76.3\ Hz \leq f_i \leq 625\ kHz,\ (i = 1, 2 \ldots 14)$. Thus, 14 pairs (amplitude and phase or real and imaginary part) of measured cervical impendences (also EIS features) were recorded for each pregnant woman, meanwhile, the measurement was carried out 1 to 3 times randomly at one visit. More details about the EIS data are described in Chapter 4.

3.2 Magnetic Impedance Spectroscopy Technology

Background When applying EIS technique, a main drawback is that it requires contact between the sample being measured and the electrodes. Impedance values can vary greatly depending on the amount of pressure applied to the probe during data collection [97]. In order to increase the accuracy of EIS and decrease potential causes of inaccuracy, with the NIHR funding, the ECCLIPPx II study developed a prototype magnetic induction spectroscopy gradiometer probe since impedance spectroscopy can detect precancerous alterations in cervical tissue and reveal changes in cervical tissue that occur during pregnancy. The ECCLIPPx II study was also carried out in the Jessop Wing Maternity Unit of the Royal Hallamshire Hospital in Sheffield, UK, from May 2018 to August 2019. The study was approved by the Yorkshire & the Humber Committee of the UK National Research Ethics Service (Research Ethics Committee 17/YH0179), and all study participants gave written informed consent.

The technique known as Magnetic Impedance Spectroscopy can be used to take contactless bio-impedance of cervical tissue over a frequency range, by using safe magnetic fields and collecting information of the tissue through which the current flows [92, 93, 94]. MIS is the multi-frequency variant of magnetic induction tomography (MIT) and one of the bioelectrical impedance spectroscopy (BIS) techniques [93, 96, 99]. For bio-impedance measurements, MIS approach has been developed and applied in order to eliminate the necessary for electrodes to be in contact with the tissue or sample [94, 95]. Instead of using electrodes to inject electrical current into tissue, MIS uses coils to generate magnetic fields that trigger currents within the sample or tissue. MIS is then able to measure these currents [96, 97, 98]. These induced currents also produce magnetic field disturbances that can be detected by sensor coils since they are

generated by the magnetic field itself. Consequently, the obstacles created by the measurement imperfections induced by electrodes in EIS can be avoided by the employment of magnetic inductive techniques [95, 96]. An additional advantage of MIS techniques with the characteristic of contactless measurement is to avoid electrical hazards when the device causes large currents to flow through electrode contacts by incident [93, 94, 96].

MIS device The MIS2 system (shown in Figure 3.3) was designed to safely measure the cervix conductivity in pregnant women in order to explore the potential association between cervical changes and women at risk of preterm birth. A modest localised magnetic field was generated by the probe head of MIS2 system (two probes was designed – Probe 2a and Probe 2c). When this magnetic field was introduced close to tissue, it caused small currents in the tissue, which were detected by sensing the magnetic field generated by these induced currents. The generated currents in the cervix generated a magnetic field that manifested itself as a minute shift in the applied field [97]. Placing two matching gradiometers on either side of the primary drive coil and spacing them equally reduced the dynamic range of the measurement device. When the gradiometer coils were coupled in antiphase, the drive magnetic field and the field produced by tissue currents cancelled each other out, leaving just a signal generated by the induced currents [97]. If the equipment was calibrated properly, the detected magnetic fields could relate to tissue conductivity. Therefore, air cored and ferrite cored were required to take the calibration.

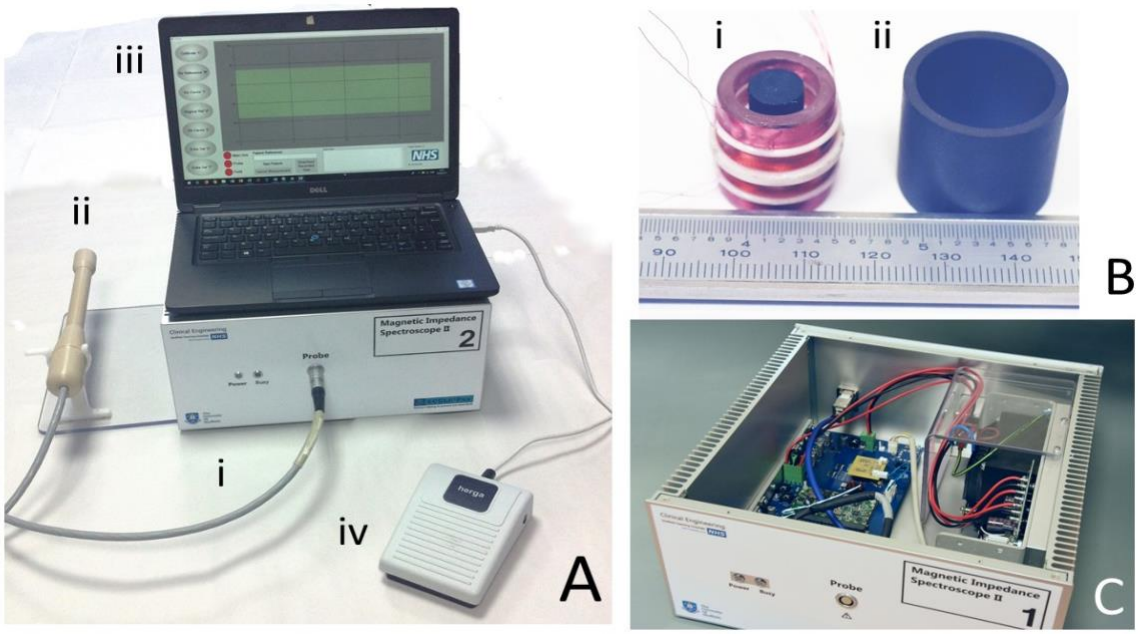

Figure 3.3: The MIS device of MIS2 system, showing the MIS cable (i), MIS probe (ii), MIS software (iii), and MIS footswitch (iv) in (A), air cored (i), and ferrite cored (ii) in (B), and the inside of MIS2 (C).

MIS measurement The MIS2 device took measurements of conductivity over 16 discrete frequencies from 21 kHz to 1013 kHz. The actual drive waveforms were created by a Direct Digital Synthesis (DDS) generator and by choosing a proper phase accumulator value of $\Delta\phi$. Therefore, the frequencies are calculated by Equation 3.5, and the 16 phase accumulator values, approximate frequencies and approximate peak drive currents were listed in Table 3.1. The last two frequencies were both 1013 kHz, hence, there were 15 frequencies applied in this work.

$$\Delta\phi = f_{out} \times \frac{2^{24}}{40\times10^{6}} \tag{3.5}$$

Table 3.1: The 16 phase accumulator values, approximate frequencies, and approximate peak drive currents.

Phase accumulator value	Approx freq in kHz	Approx peak drive current
8808	21	0.30 A
17616	42	0.26 A
24327	58	0.31 A
30199	72	0.28 A
36071	86	0.29 A
41943	100	0.27 A
84725	202	0.13 A
126668	302	90 mA
168611	402	67 mA
210554	502	54 mA
253336	604	45 mA
295279	704	39 mA
337222	804	34 mA
379165	904	30 mA
424883	1013	27 mA
424883	1013	27 mA

During the measurement process, the MIS2 system used alternating current (AC) as input, and outputted the resulting induced voltage. Basically, one measurement consisted of seven sweeps, also called events for Ferrite Target, Air, Cervix 1, Vaginal Reference 1, Cervix 2, Vaginal Reference 2 and Cervix 3 respectively (more details shown in Table 3.2). When implementing the measurement, firstly, the nurse put the MIS probe into the ferrite cored and pressed the footswitch. Ferrite target was recorded. Then, the probe was put into the air cored to record the air target. After that, the nurse put the probe into vagina close to cervix tissue to record the

cervix 1 data and left the tissue a short distance to record vaginal 1 data. Finally, repeated the measurement process several times to record cervix 2 and cervix 3, and vaginal 2.

Table 3.2: The descriptions of seven events of Ferrite Target, Air, Cervix 1, Vaginal Reference 1, Cervix 2, Vaginal Reference 2 and Cervix 3.

Events	Description
Ferrite Target	measurement of ferrite solution
Air	measurement of air
Cervix 1	measurement of cervix at first time
Vaginal Reference 1	measurement of vagina for reference at first time
Cervix 2	measurement of cervix at second time
Vaginal Reference 2	measurement of vagina for reference at second time
Cervix 3	measurement of cervix at third time

Thereafter, the MIS data were recorded as two types of data, one of that was the raw data in time domain, where seven events were stored respectively; and the other one of that was the Real-Time MIS data in frequency domain which were converted from the raw data by Field-Programmable Gate Array (FPGA), which applied a matched filter to the analog-to-digital converter (ADC) - collected data, where seven events were stored together. More details about the two types of MIS data were described in Chapter 6.

Chapter 4

4. EIS Data Processing and EIS Data-based PTB Prediction

The ECCLIPPx^{TM} project (mentioned in Chapter 3) aimed to determine if a potential test of electrical impedance spectroscopy measured by the latest version of the cervical EIS device (The Sheffield Mark V) can accurately identify women at risk of PTB, so that who can be offered the treatment to prolong pregnancy [89]. The group has revealed that EIS may be beneficial in predicting the onset of PTB, and revealed that women who are more likely to deliver prematurely have lower cervical impedance in the middle of their pregnancies than those who deliver normally.

Based on this hypothesis, in order to analyse and evaluate the feasibility of EIS for prediction of PTB risk, in this study, a strategy of machine learning was given for the EIS-based PTB risk prediction. The proposed method involved an optimum-EIS spectrum filter, a range of feature selection techniques and a ML-based PTB risk prediction model, in such a way that the filter selected the optimum EIS spectrum most relevant to the cervix condition of women from among all of the repeatedly measured EIS spectra, and then feature selection techniques were used to choose most associate predictors with PTB risk, finally these PTB predictors were trained into a ML-based PTB risk prediction model. Both the optimal-EIS spectrum filter and the PTB risk prediction model were constructed using machine learning approaches, which adopted a range of techniques including oversampling, polynomial representation, Orthogonal point-biserial correlation coefficient (OBCC), Canonical Correlation Analysis (CCA) and Two-stage canonical correlation analysis (TSCCA) based feature selection, as well as support vector machine for optimal-EIS spectrum filter and logistic regression for EIS-based PTB risk prediction model.

4.1 Methodology

4.1.1 EIS Dataset Analysis

The dataset contains 438 recruited women (Table 4.1) with their EIS spectra over 14 frequencies ranging from 76.3 Hz to 625 kHz measured at between 19 to 23 weeks of gestation, where 197 women (45%) are at higher risk of PTB and 241 women are at lower risk of PTB, and 374 women (85.4%) deliver on term and 64 women (14.6%) give a preterm birth. Of 374 on-term birth women (Table 4.2), 143 women (38.2%) and 231 women (61.8%) are at higher risk of PTB and at lower risk of PTB respectively. The on-term birth women give the delivery from 37 to 42 weeks of gestation. The period between their test and delivery is 15 to 23 weeks of gestation and average period is 21 weeks of gestation. 64 PTB women (Table 4.3) consist of 54 higher risk PTBs (84.4%) and 10 lower risk PTBs (15.6%). Their premature births happen at from 20 to 36 weeks of gestation after 1 to 17 weeks of test (average: after 12 weeks of test), where 10.9% of premature births are extreme PTB (<28 completed weeks of gestation), 20.3% are very PTB (28 to 32 weeks), 18.8% are moderate PTB (32 to 34 weeks) and 50.0% are late PTB (34 to 37 weeks).

Table 4.1: The information of 438 recruited women, PTB: preterm birth.

Content	Description
Recruits:	438 women
Age range:	17 to 45
Gestation of visit:	19 to 23 weeks of gestation
Higher risk of PTB:	197 of 438 women (45%)
Lower risk of PTB:	241 of 438 women (55%)
On-term birth:	374 women (85.4%)
Preterm birth:	64 women (14.6%)

Table 4.2: The information of 374 on-term birth women, PTB: preterm birth.

Content	Description
On-term birth:	374 women
Higher risk of PTB:	143 of 374 women (38.2%)
Lower risk of PTB:	231 of 374 women (61.8%)
Gestation of on-term birth:	37 to 42 weeks of gestation
Delivery after testing:	15 to 23 weeks
Average:	21 weeks

Table 4.3: The information of 64 PTB women, PTB: preterm birth.

Content	Description
Preterm birth:	64 women
Higher risk of PTB:	54 of 64 women (84.4%)
Lower risk of PTB:	10 of 64 women (15.6%)
Gestation of PTB:	20 to 36 weeks of gestation
Extreme PTB (<28 weeks):	7 of 64 women (10.9%)
Very PTB (28 to 32 weeks):	13 of 64 women (20.3%)
Moderate PTB (32 to 34 weeks):	12 of 64 women (18.8%)
Late PTB (34 to 37 weeks):	32 of 64 women (50.0%)
Delivery after test:	1 to 17 weeks
Average:	12 weeks

The EIS spectra are complex structure consisting of amplitude and phase or real and imaginary part, also there are 28 EIS features measured once for a recruited woman and the measurement was carried out 1 to 3 times randomly at their visit. Hence, 438 women contain 1022 spectrum readings in total, where 41 women have one spectrum reading only, 210 women have two spectrum readings only, and 187 women have three spectrum readings. For women with more

than one spectrum reading, the coefficient of variation (CV) of amplitude of spectrum readings at the lowest frequency (at 76.3 Hz) is calculated by,

$$CV = \frac{\sigma}{\mu} \tag{4.1}$$

to evaluate reliability of spectrum readings, where σ is standard variation and μ is mean. Generally, the CV-based evaluation of reliability is categorized into four levels: "excellent" if the percentage is less than 0.05, "good" if it falls between 0.05 – 0.1, "acceptable" if it ranges from 0.1 – 0.15, and "unacceptable" if it exceeds 0.15 [102]. 210 coefficients of variation of women with two spectrum readings only and 187 coefficients of variation of women with three spectrum readings are demonstrated in histograms respectively (Figure 4.1 and Figure 4.2).

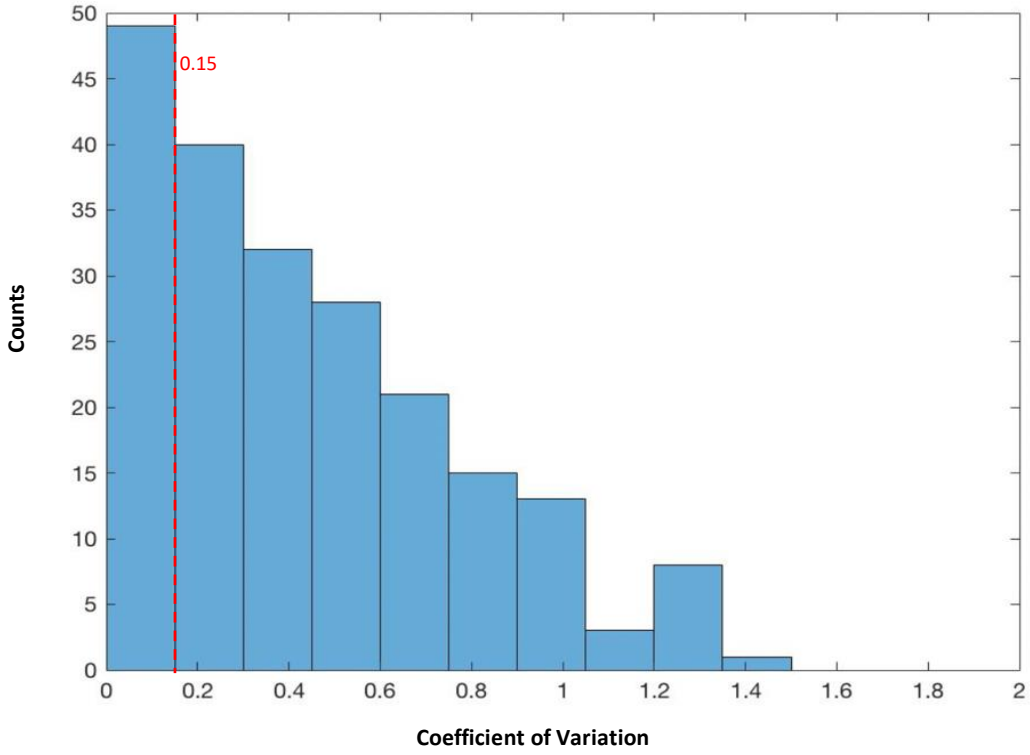

Figure 4.1: The histogram plot of 210 CVs of women of having two spectrum readings only at the lowest frequency (at 76.3 Hz) (161 CVs above 0.15), CV: coefficient of variation.

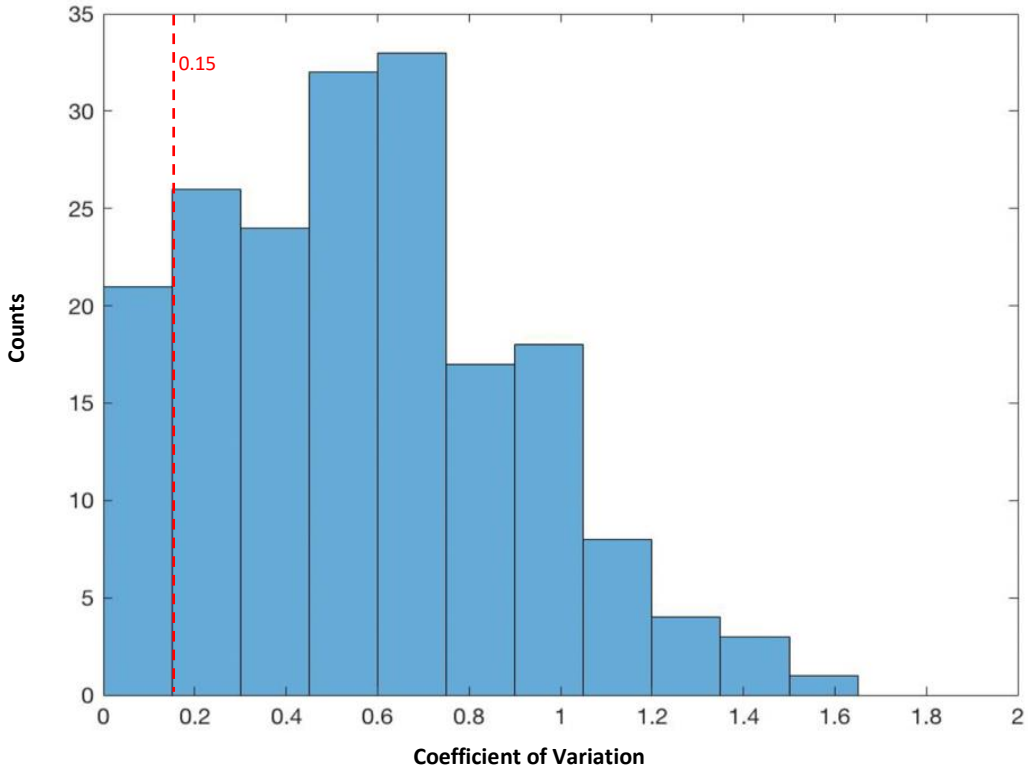

Figure 4.2: The histogram plot of 187 CVs of women of having three spectrum readings at the lowest frequency (at 76.3 Hz) (166 CVs above 0.15), CV: coefficient of variation.

From the statistics, totally, 327 of 397 women's coefficients of variation are more than 0.15 and the ratio is around 82.4%, thus the reliability of readings of the 327 women are regarded as unacceptable according to the principle in [102] and the optimal readings need to be selected in this work. Besides, Figure 4.3 illustrates what two spectrum readings over 14 frequencies look like, from six women whose CVs are around 0.1, 0.3, 0.5, 0.7, 0.9 and 1.1 respectively. Meanwhile, the amplitudes of three spectrum readings of 6 women are also illustrated in Figure 4.4 respectively. In Figure 4.3 and Figure 4.4, it is obvious to observe the differences among spectrum readings with coefficients of variation 0.5 are equal to and more than 0.5.

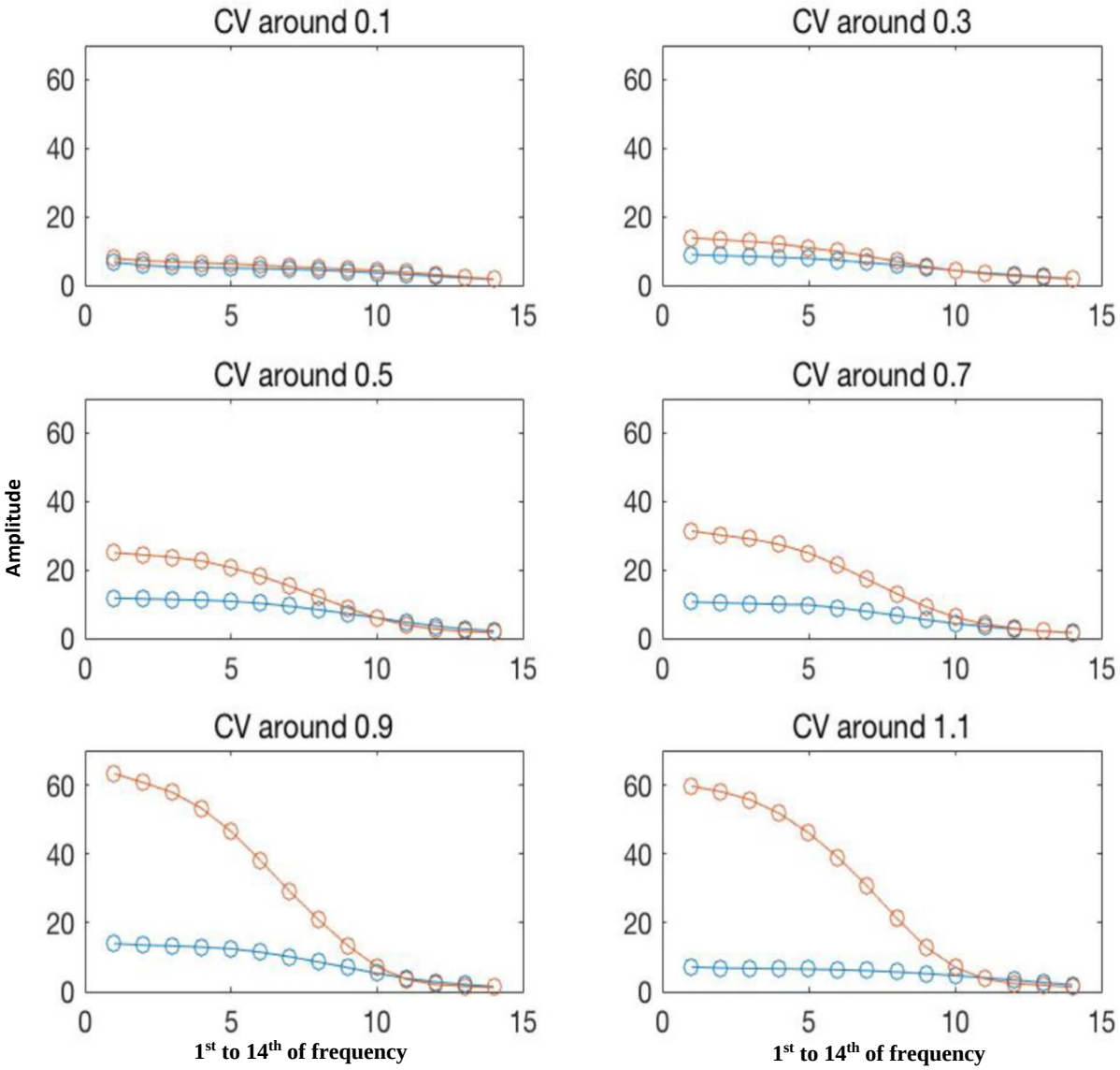

Figure 4.3: The illustration of what two spectrum readings (red and blue) over 14 frequencies look like, from six women whose CVs are around 0.1, 0.3, 0.5, 0.7, 0.9 and 1.1 respectively, CV: coefficient of variation.

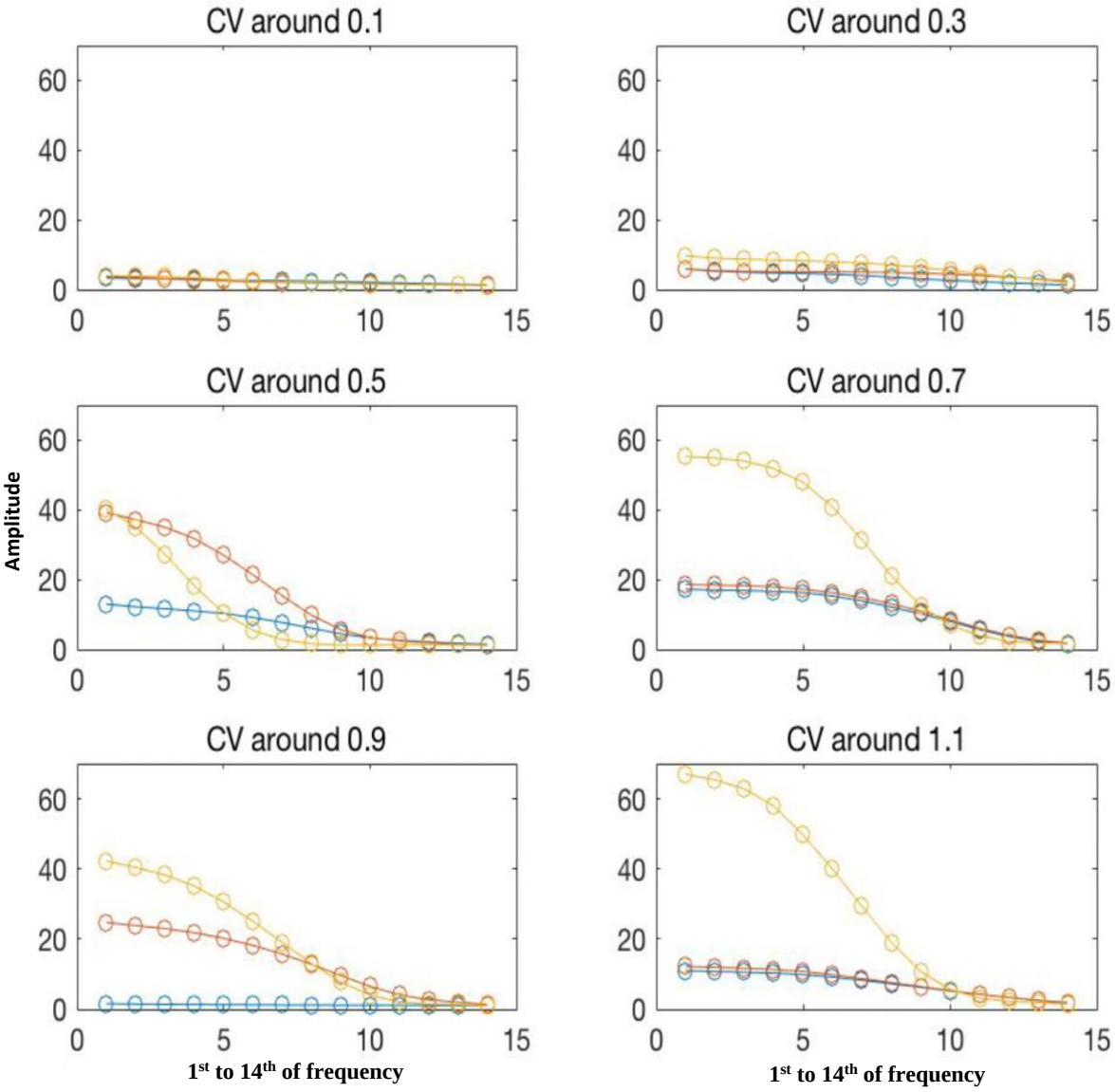

Figure 4.4: The illustration of what three spectrum readings (red, blue and yellow) over 14 frequencies look like, from six women whose CVs are around 0.1, 0.3, 0.5, 0.7, 0.9 and 1.1 respectively, CV: coefficient of variation.

4.1.2 The Principle for Determination of Labels for EIS Spectrum Readings

The data pre-processing is carried out to deal with the differences among spectrum readings and select the optimum spectrum readings. Firstly, the 1022 EIS spectrum readings of 438 women are visualized using 2-D plots (Figure 4.5) in respect of the means of amplitudes of PTB women's spectrum readings and on-term birth women's spectrum readings, and the means of phases of PTB women's spectrum readings and on-term birth women's spectrum readings, respectively. Observing from Figure 4.5, the means of amplitudes of PTB women's spectrum readings over lower frequencies are larger than on-term birth women's and the differences are smaller and almost equal to 0 with the frequencies increasing. The means of phases of on-term birth women's spectrum readings over first 9 frequencies are greater than PTB women's and the differences are consistent, while they are suddenly smaller than PTB women's over the last 4 frequencies. There is almost no difference between means of phases of on-term birth and PTB at 10^{th} frequency.

The difference between the means of amplitudes and phases of on-term birth and PTB women's spectrum readings is hypothesized that cervical tissues of on-term birth women and PTB women have different compositions and remodelling processes during pregnancy, so that the EIS impedances of on-term birth and PTB women's cervical tissues present different. Based on this hypothesis, the pseudo-code of the proposed principle for labelling is described in Principle 4.1, to emphasize the differences. Eventually, all spectrum readings labelled with '1' (selected to be optimal spectrum readings) of 438 women are visualized in respect of the means of amplitudes and the means of phases over 14 frequencies (Figure 4.6).

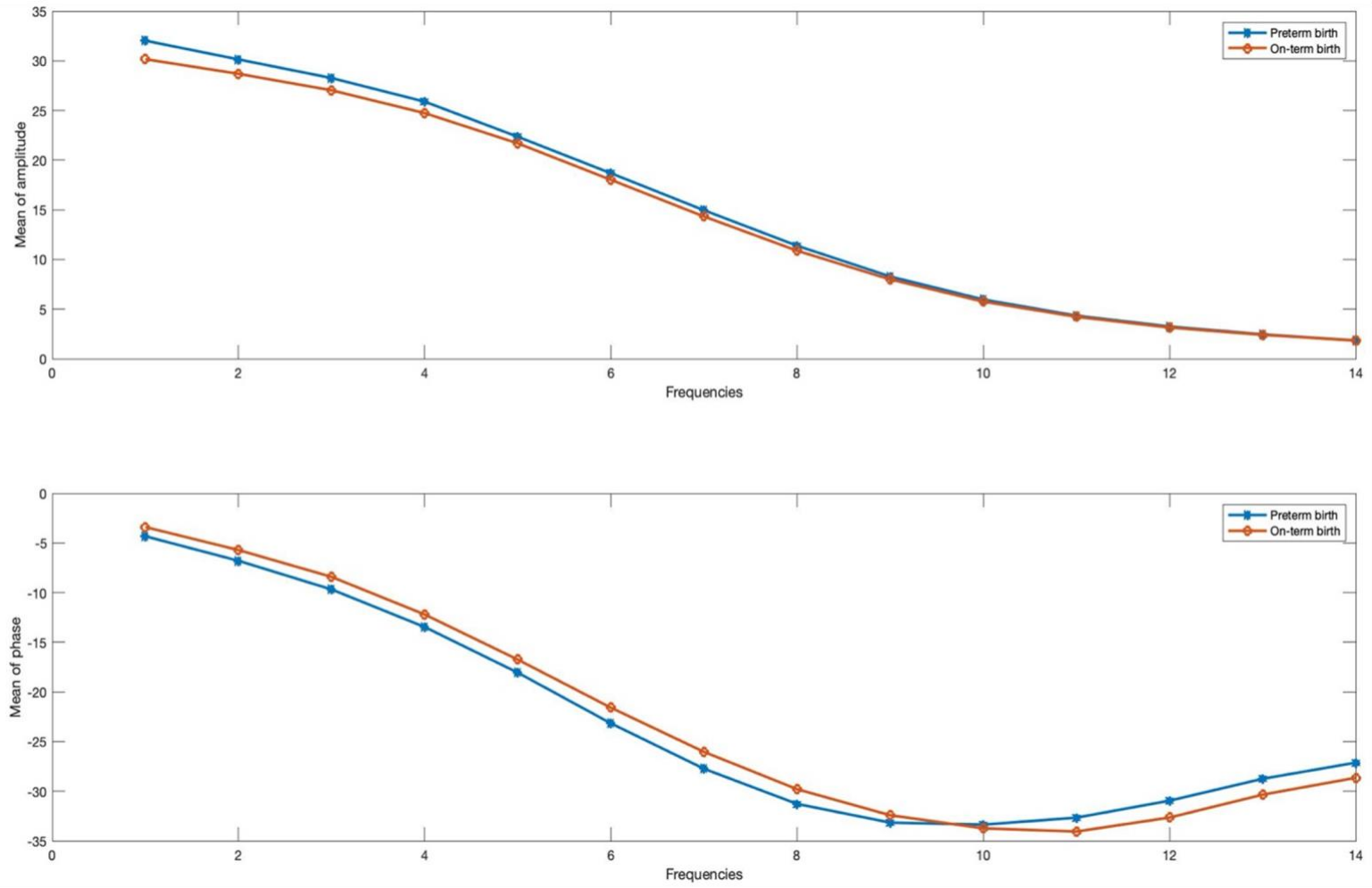

Figure 4.5: The visualization of the means of amplitudes of all EIS spectrum readings (above) and the visualization of the means of phases of all EIS spectrum readings (below) over 1st to 14th of frequency.

Principle 4.1 Determination of labels for EIS spectrum readings

Input: The recruited women of $W = \{w_1, w_2, \dots, w_i\}, for\ i = 1, 2, \dots, 438$, and i^{th} woman's EIS spectrum readings of $w_i = R^i_{j\times 14}, for\ j = \{1, 2\ or\ 3\}$.

Output: The EIS dataset with labelled spectrum readings.

1. **for** all w_i in W do
2. **if** $j = 1$ of $R^i_{j\times 14}$
   define $p_1 = R^i_{j\times 14}$ and label $p_i$with '1' (selected EIS spectrum reading)
   store p_1 and the label of p_1 in P
3. **else if** $j = 2$
   computer CV_i of w_i
   if $CV_i < 0.15$
   define $p_1 = mean(R^i_{j\times 14}, 1)$ and label $p_i$with '1'
   store p_1 and the label of p_1 in P
   else
   define $p_1 = R^i_{j\times 14}(n, :)$, where find $n = max(sum(abs(R^i_{j\times 14}), 2))$
   remove p_1 from $R^i_{j\times 14}$
   label p_1 with '1' or '0' depending on w_i is PTB or on-term birth
   define $p_2 = R^i_{j\times 14}$with the opposite label
   store p_1 and p_2, and the labels of p_1 and p_2 in P
   end
4. **else if** $j = 3$
   computer CV_i of w_i
   if $CV_i < 0.15$
   define $p_1 = mean(R^i_{j\times 14}, 1)$ and label $p_1$with '1'
   store p_1 and the label of p_1 in P
   else
   computer CVs of any two rows of $R^i_{j\times 14}$
   if there is a CV < 0.15

mean two rows (that produced CV<0.15) by columns, and replace two rows with the mean in $R^i_{j\times 14}$, then go to step 3

else

if w_i is PTB

define $p_1 = R^i_{j\times 14}(n,:)$, and label with '1'

where find $n = max(sum(abs(R^i_{j\times 14}),2))$

else

define $p_1 = R^i_{j\times 14}(n,:)$, and label with '1'

where find $n = min(sum(abs(R^i_{j\times 14}),2))$

end

remove p_1 from $R^i_{j\times 14}$

define $p_2 = R^i_{j\times 14}$with the labels of '0'

store p_1 and p_2, and the labels of p_1 and p_2 in P

end

end

end

end

5. **Return** the EIS dataset P with labelled spectrum readings.

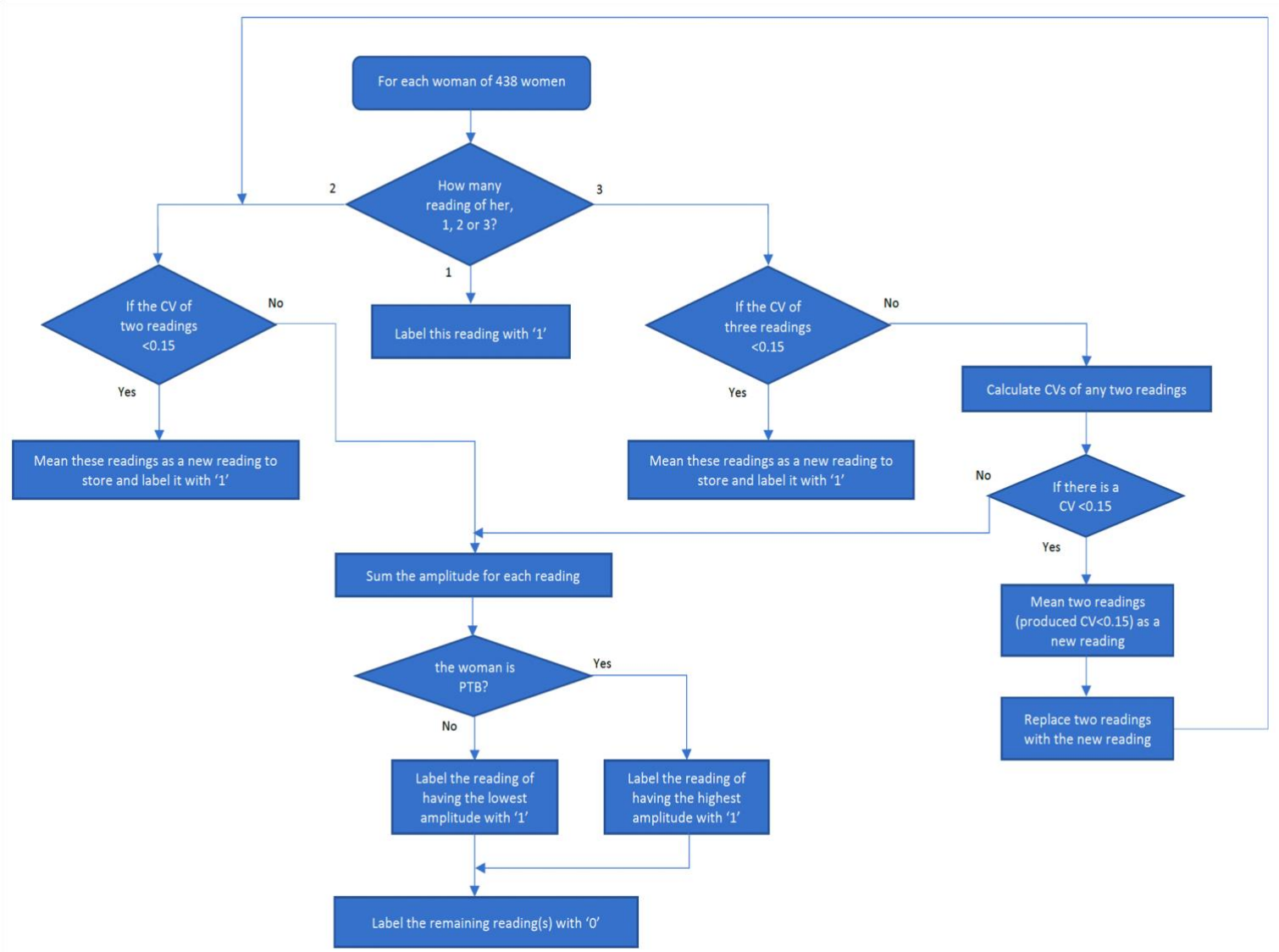

Figure 4.6: The flowchart of determination of labels for EIS spectrum readings, CV: coefficient of variation, PTB: preterm birth.

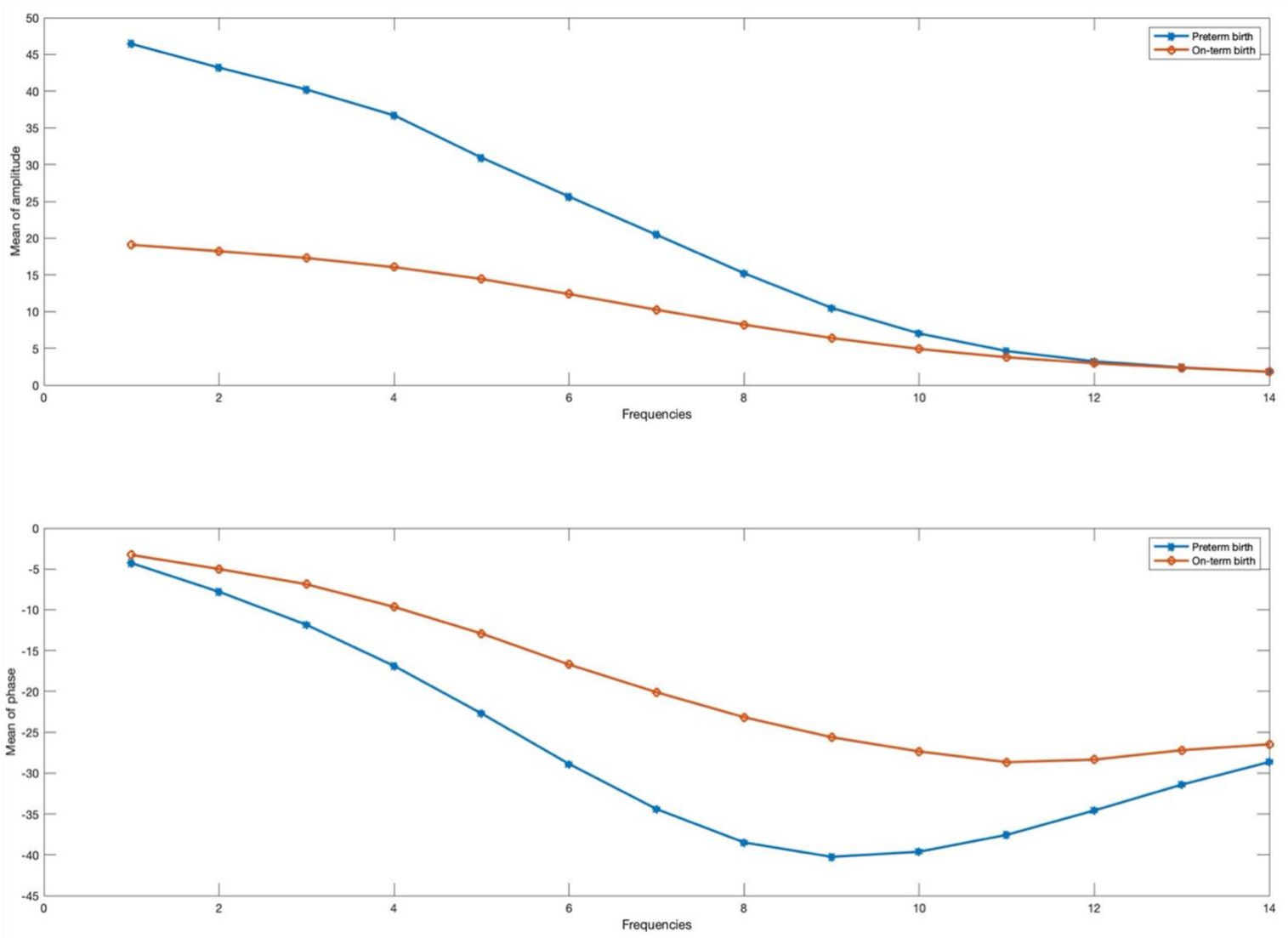

Figure 4.7: The visualization of the means of amplitudes of optimal EIS spectrum readings (above) and the visualization of the means of phases of optimal EIS spectrum readings (below) over 1st to 14th of frequency.

4.1.3 Polynomial Feature Construction

First of all, logistic regression model can describe the generalized linear relationship between predictors (independent variables) and class posterior probability (dependent variable) in terms of the logit function [103, 104],

$$p = \frac{1}{1+e^{-z}} \tag{4.2}$$

$$z = \alpha + \beta x = \alpha + \beta_1 x_1 + \cdots + \beta_m x_m \tag{4.3}$$

p is the probability, $\frac{1}{1+e^{-z}}$ is the logit function, α and β are parameters to be determined, and x is a range of predictors.

Previous study of applying EIS features in prediction of PTB risk focused on linear EIS feature structure and used the structure in logistic regression model [86]. However, the relation between features and PTB risk could be nonlinear and more complex in actual. Hence, the complicated feature structure should be considered to extend the potential relationships, so that increase the performance of prediction of PTB risk. Considering of Equation 4.2, the employment of complicated feature structure should remain the linear function between z and parameters of α and β, meanwhile, represent the complicated relation between z and predictors of x. To address this challenge, the polynomial feature structure is employed into logistic regression to replace the linear structure, and the Equation 4.2 can be transformed with polynomial feature structure [105] as,

$$z = \alpha + \sum_{i_1=1}^{n} \beta_{i_1} x_{i_1} + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \beta_{i_1 i_2} x_{i_1} x_{i_2} + \cdots + \sum_{i_1=1}^{n} \ldots \sum_{i_l=i_l-1}^{n} \beta_{i_1 i_2 \ldots i_l} x_{i_1} x_{i_2} \ldots x_{i_l} \tag{4.4}$$

where n is the number of predictors and l is the polynomial degree of structure, which means that the order of each term in structure is not higher than l. The total number of terms in structure can be calculated in terms of n and l by,

$$N = \binom{n+l}{l} = \frac{(n+l)!}{n! \cdot l!} \tag{4.5}$$

4.1.4 Feature Selection Approaches

This work employs three feature selection approaches including **1)** orthogonal point-biserial correlation coefficient, **2)** canonical correlation analysis and **3)** two-stage canonical correlation analysis. These approaches and their advantages will be introduced in details as follows.

1) Orthogonal point-biserial correlation coefficient: The point-biserial correlation coefficient (BCC) $r_{pb}(X,Y)$ can be defined as a specific kind of the Pearson correlation coefficient [106], and it can be computed by,

$$r_{pb}(X,Y) = \frac{\bar{X}_1 - \bar{X}_0}{\sigma_X}\sqrt{\frac{n_0 n_1}{N^2}} \tag{4.6}$$

where X_1 is a set of variables whose corresponding Y is class of '1' and X_0 is another set of variables whose corresponding Y is class of '0'; $\bar{X}_1$ is the mean of variables of X_1 and $\bar{X}_0$ is the mean value of variables of X_0; σ_X is the standard deviation of variables of X; n_0 is the number of Y that belong to class of '0', n_1 is the number of Y that belong to class of '1', and N is the total number of variables. The BCC generally represents the linear correlation between continuous variables and the dichotomous variable [106], so that BCC is an appropriate correlation coefficient used to deal with binary classification problem.

Professor Solares and his colleague from University of Sheffield in 2017 proposed this approach and suggested to replace the error reduction ratio (ERR) of orthogonal forward regression (OFR) or orthogonal least square (OLS) algorithm with BCC [105, 107]. Then, the processes of applied OBCC-based feature selection algorithm [105] are presented in Algorithm 4.1.

- First, select the maximum correlation as the first candidate, by calculating BCCs for all normalized features. Calculate the 3-fold cross validation AUC with the selected feature and y and store it.

- The second step is to orthonormalize the remaining features in terms of the chosen feature at first step.

- Then, calculate the BCCs of orthonormalized features and select the feature with maximum value of BCC for next candidate till reach the maximum number that has been set previously. Calculate the 3-fold cross validation AUCs of target ML (Target ML is determined in term of the ML model served by OBCC-based feature selection. For example, if OBCC-based feature selection serves LR, the target ML is LR.) with the selected feature and y and store them.

- Finally, determine the final feature set in terms of 3-fold cross validation AUCs of target ML.

Algorithm 4.1 OBCC-based feature selection algorithm

Input: The feature library of $F = \{f_1, f_2, \dots, f_M\}$, the binary-class output y and the maximum number of features to be selected n_{max}

Output: the final selected feature set

1. **for** all f_i in F do
2. Define $w_i = \frac{f_i}{|f_i|}$
3. Calculate $r_{pb}(w_i, y)$ and store it
4. **end**
5. Define $q_1 = w_j$ and $s_1 = f_j$,
   where find $j = max\{[r_{pb}(w_1, y), \dots, r_{pb}(w_M, y)]_{1\times M}\}$
6. Calculate 3-fold cross validation AUC of target ML with y and s_1, and store it in auc
7. Remove f_j from F
8. **for** $k = 2$ to n_{max} do
9. **for** all f_i in F
10. Orthonormalize f_i in terms of $[q_1, \dots, q_{k-1}]_{1\times(k-1)}$ to get w_i
11. Calculate $r_{pb}(w_i, y)$ and store it
12. **end**
13. Define $q_k = w_j$ and $s_k = f_j$,
    where find $j = max\{[r_{pb}(w_1, y), \dots, r_{pb}(w_M, y)]_{1\times(M-k+1)}\}$

---

14. Calculate 3-fold cross validation AUC of target ML with y and $[s_1, \dots, s_k]$, and store it in auc
15. Remove f_j from F
16. **end**
17. **Return** the final selected feature set $S = [s_1, \dots, s_m]$, where find $m = max\{auc\}$

2) Canonical correlation analysis: In statistical analysis, canonical correlation analysis is a comprehension of the cross-covariance matrix, and it can be used to express the linear relation between two vectors of variables [108]. The principle is to extract two typical variables that are two linear combinations of the variables from two vectors respectively [108]. Then, use the correlation between two typical variables to reflect the overall relationship between two vectors of variables.

In detail, $X = \{x_1, x_2, x_3, \cdots x_n\}$ and $Y = \{y_1, y_2, y_3, \cdots y_m\}$ are two vectors of variables of $x_1, x_2, x_3, \cdots x_n$ and $y_1, y_2, y_3, \cdots y_m$. The aim of CCA is to adjust the values of a and b to make the correlation $p = corr(aX, bY)$ maximum. Then, the maximum correlation is regarded as the value of CCA. Compared with BCC, the significant character is that CCA can present a global optimization of the combination of features [109], while BCC is the correlation of single feature. The proposed CCA-based feature selection algorithm is demonstrated as follows:

- Calculate CCAs for each candidate feature.
- Choose the feature with the highest value of CCA as the initial feature, calculate the 3-fold cross validation AUC of target ML and store it.
- Then calculate the CCAs for combinations of the initial feature and each remaining feature, choose the remaining feature that achieves the highest value of CCA with the initial feature, calculate the 3-fold cross validation AUC of target ML and store it.
- Repeat the above step to choose the features with the highest value of CCA, till reaching the setup number of features.
- Finally, determine the final feature set in terms of the stored highest values of AUCs.

Theoretically, CCA can achieve the global optimization if considering overall combinations of features. However, the candidate features are generally large-scale, so that there are huge combinations of candidate features. In this case, the CCA-based feature selection algorithm uses the forward selection approach, thereby achieving the local optimization only.

3) Two-stage canonical correlation analysis: The forward feature selection approach is used widely with the properties of rapid and simple computation, where the principle of forward feature selection approach is to choose one best-performed feature each time in terms of a specific criterion and this process of choosing is repeated till reaching the desired number of features. In this case, the forward feature selection approach possibly causes a problem of local optimization. The combination of features, however, is hardly to achieve globally optimal. To deal with this problem, there has been a two-stage fast recursive algorithm (TSFRA) [111] with a novel idea of backward model refinement, besides forward selection methods. Based on the TSFRA, a two-stage orthogonal least square method (TSOLS) [110] has been proposed to improve the TSFRA's problem of complicated and time-consuming process, using orthogonal least square method to replace fast recursive algorithm in the section of forward selection [110]. The TSOLS method uses the error reduction ratio as a criterion. An improvement of replacing ERR with CCA is carried out in present work to combine two-stage strategy and CCA.

The TSCCA approach, as the name implies, contains two stages to adjust the feature structure. The first stage is forward subset selection that selects the well-performed features as an initial feature structure from all candidate features in terms of the criterion of CCA. The second stage is backward feature structure refinement that reviews the initial structure determined on first stage, to justify if there are features in initial structure able to be replaced by the remaining features that can achieve a better value of CCA. The proposed TSCCA-based feature selection algorithm is shown in Algorithm 4.2.

- To be specific, $F = \{f_1, f_2, f_3, f_4, \cdots \cdots f_n, f_{n+1}, \cdots f_M\}$ is the candidate feature library, where f_j is the jth candidate feature and M is the number of candidate features. Firstly, forward subset selection approach is carried out in terms of the CCA-based feature selection algorithm, to determine the initial feature structure of $I = \{I_1, I_2, I_3, I_4, \cdots I_{Ms}\}$ where MS is the optimal number of features.

- Then, start to review features in initial structure backward from $(MS-1)th$ to $1st$, and replace I_{Ms-1} with each term in remaining library $\{f_{Ms+1}, f_{Ms+2}, \cdots f_M\}$. Replace I_{Ms-1} with the feature that achieves the higher value of CCA and place it on the final position in I, then carry on reviewing from $(MS-2)th$. If there is no feature in remaining library that achieves the higher value of CCA, place I_{Ms-1} on the final position in I and continue to review from $(MS-2)th$.

- Stop reviewing until all of features in I cannot be replaced with features in remaining library and then obtain the refined feature set I.

Algorithm 4.2 TSCCA-based feature selection algorithm

Input: The feature library of $F = \{f_1, f_2, f_3, f_4, \cdots \cdots f_n, f_{n+1}, \cdots f_M\}$ and the binary-class output y

Output: The selected feature set I

Stage 1: forward subset selection

Determine the initial feature structure of $I = \{I_1, I_2, I_3, I_4, \cdots I_{Ms}\}$ in terms of the CCA-based feature selection algorithm, and the remaining feature set is $R = \{f_{Ms+1}, f_{Ms+2}, \cdots f_M\}$.

Return the initial feature structure of I, the remaining feature set of R and the optimal number of features of MS.

Stage 2: backward feature structure refinement

1. **for** $j = 1:(MS-1)$
   define $R = \{I_{MS-j}, R\} = \{R_1, R_2, \cdots R_{M-MS+1}\}$
   remove I_{MS-j} from I
   calculate the $CCA_i([I, R_i], y)$
   define $I = [I, R_k]$, where find $k = \max(CCA_i([I, R_i], y))$
   remove R_k from R
   end
2. **repeat** the first step till I does not change
3. **return** the selected feature set I

4.2 Machine Learning Pipelines

The EIS dataset of 438 women was processed based on the principle of determination of labels for EIS spectrum readings, so that each woman contains a result of selecting EIS spectrum readings (label of '1' for selected spectrum and label of '0' for unselected spectrum), and another result of PTB (label of '1' for PTB and label of '0' for on-term birth). The 438 women were randomly split into 100 subsets of 66% women and 100 subsets of 34% women, and each subset of 66% women consisted of 66% PTB women and 66% on-term birth women, similarly, each subset of 34% women also consisted of 34% PTB women and 34% on-term birth women. More details are shown as follows and Figure 4.8.

438 recruited women at PTB risk are,

- of 64 preterm births (14.6%),
- of 374 on-term births (85.4%),
- split into a subset of 290 women including 43 preterm births and 247 on-term births,
- split into a subset of 148 women including 21 preterm births and 127 on-term births.

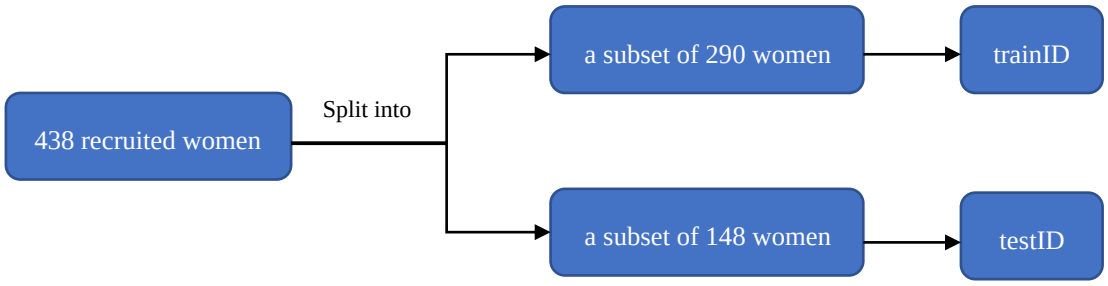

Figure 4.8: The splitting process showing how 438 women were split into trainIDs and testIDs. The testID was unseen to the trainID. The splitting process was repeated 100 times to obtain 100 trainIDs and 100 testIDs.

In terms of the information of 438 of recruited women, the ratio of PTB women and on-term birth women was around 1:5.8, so that the subsets were unbalanced, which may lead to a biased

machine learning model predicting towards the class of majority. Additionally, the EIS features would be extended by polynomial feature structure, causing a high dimension problem, which may result in poor accuracy on unseen data because of overfitting issue of machine learning model to training sets. To deal with the problems, a machine learning pipeline for training optimal-EIS spectrum filter and a machine learning pipeline for training EIS-based PTB risk prediction model were proposed in detail as follow.

4.2.1 A Machine Learning Pipeline for Training Optimal-EIS Spectrum Filter Models

The proposed machine learning pipeline for training optimal-EIS spectrum filter model is as demonstrated in Figure 4.9. Firstly, for training optimal-EIS spectrum filter model, a training set (trainsetI) included all EIS readings of a trainID and the EIS readings' labels (label of '1' for selected spectrum and label of '0' for unselected spectrum), in total, 100 trainsetIs were generated and each trainsetI consisted of all readings of 290 women and readings' labels.

After that, for each trainsetI, the mean, standard deviation (STD) and centroid of amplitudes and phases of EIS spectra were calculated respectively, therefore, six statistics features were obtained and combined with 28 EIS features. Then, the polynomial feature structure with three nonlinear degree was used to the 34 features, and the trainsetI was extended to 7770 features ($\frac{(n+l)!}{n!\cdot l!} = \frac{(34+3)!}{34!\cdot 3!} = 7770$). The OBCC-based feature selection approach with a maximum number of features $n_{max} = 16$ (according to experiment) was used to choose 16 features, feature by feature. Once a feature selected into the feature set, a SVM model was trained between the feature set and readings' labels, and a 3-fold cross validation AUC of SVM was computed and saved. Then, the optimal number of features was determined according to the highest 3-fold cross validation AUC of SVM. Finally, a SVM model as the optimal-EIS spectrum filter model was trained on the trainsetI with the optimal combination of features. The optimal-EIS spectrum filter model trained on each corresponding trainsetI was saved to select optimum EIS spectra for the corresponding testing set. Therefore, total 100 optimal-EIS spectrum filter models were trained and saved. The description of trained optimal-EIS spectrum filter model is shown in Table 4.4.

Algorithm 4.3 Machine learning pipeline for training optimal-EIS spectrum filter models

Input: The 100 trainsetIs, and the maximum number of features to be selected n_{max}

Output: SVM models and optimal numbers of features for all SVM models

1. **for** all $trainsetI_i$ of $trainsetIs$ do
2. Calculate $mean$, STD and $centroid$ of $trainsetI_i$ and add them into $trainsetI_i$
3. Transform $trainsetI_i$ to polynomial features with l nonlinear degree, as $Poly\_trainsetI_i$
4. Implement OBCC-based feature selection approach with 3-fold cross validation of SVM
5. Define $F = \{f_1, \cdots f_{n_{max}}\}$, the selected features
6. Define $AUC = \{AUC_1, \cdots AUC_{n_{max}}\}$, the 3-fold cross validation AUCs of SVM
7. The optimal combination of features is defined $P = \{f_1, \cdots f_j\}$, where find $j = max\{\{AUC_1, \cdots AUC_{n_{max}}\}\}$
8. Train a SVM model with the optimal combination of features, $P = \{f_1, \cdots f_j\}$
9. Save the SVM model and optimal number of features, j
10. **end**
11. **Return** SVM models and optimal numbers of features for all SVM models

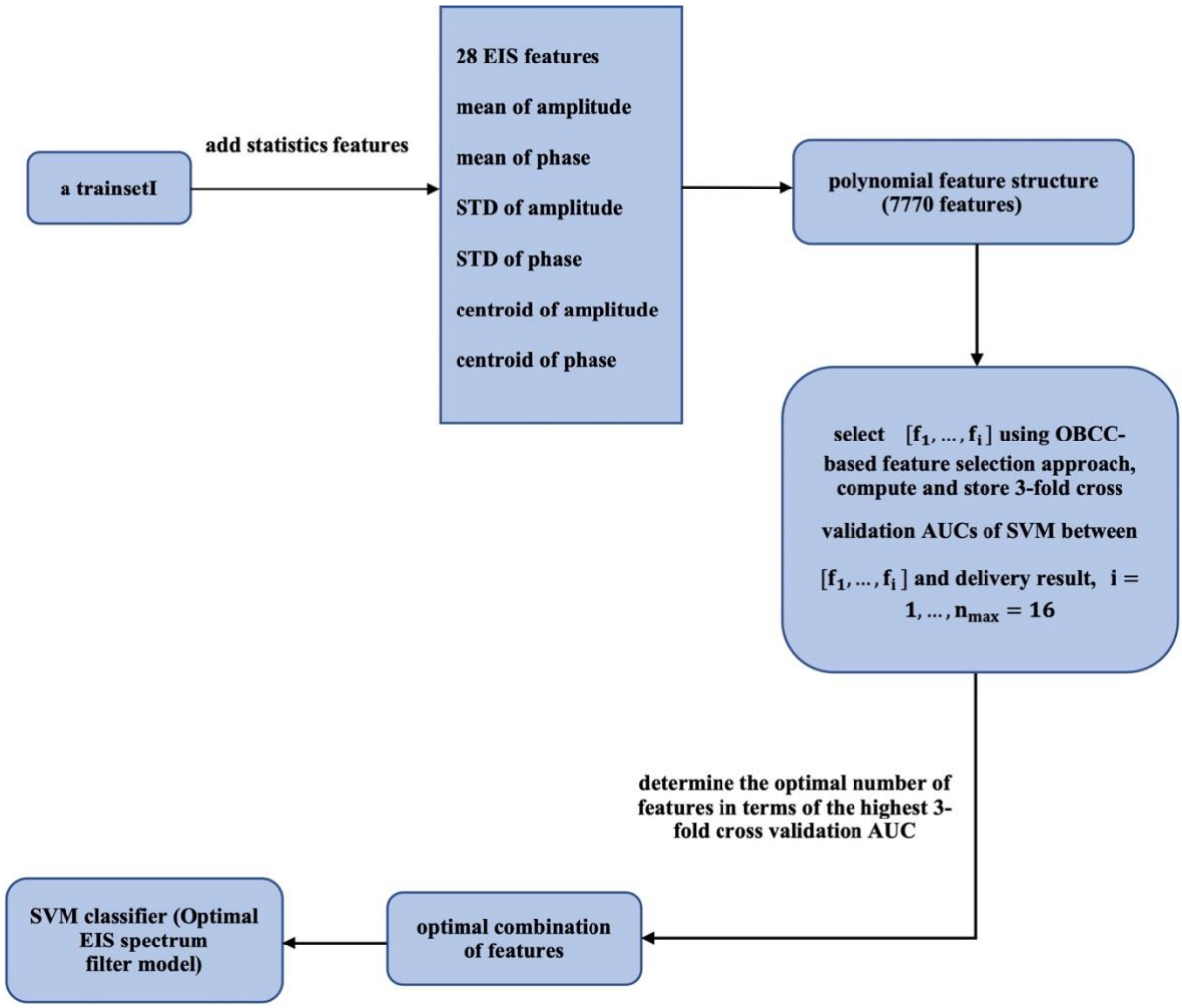

Figure 4.9: The proposed machine learning pipeline for training optimal-EIS spectrum filter model. EIS: electrical impedance spectroscopy, STD: standard deviation, AUC: area under the receiver operating characteristic curve, SVM: support vector machine, OBCC: orthogonal point-biserial correlation coefficient.

Table 4.4: The description of Optimal-EIS spectrum filter model, which used the SVM as a classifier and OBCC-based feature selection approach to select features in advance. SVM: support vector machine, OBCC: orthogonal point-biserial correlation coefficient.

Optimal-EIS spectrum filter model	
Classifier	Feature selection approach
SVM	OBCC-based

4.2.2 A Machine Learning Pipeline for Training EIS-based PTB Risk Prediction Models

The proposed machine learning pipeline for training EIS-based PTB risk prediction model is as demonstrated in Figure 4.10. Firstly, for training EIS-based PTB risk prediction model, a training set (trainsetII) included optimal readings (determined according to the proposed optimal reading principle) of a trainID and their results of delivery (label of '1' for PTB and label of '0' for on-term birth), in total, 100 trainsetIIs were generated and each trainsetII consisted of optimal readings of 290 women and their delivery results.

Then, each trainsetII was oversampled using the Borderline-SMOTE1 [112], which randomly generated 204 new synthetic data of PTB and added them into the trainsetII to balance the classes of '0' and '1'. After that, the polynomial feature structure was used to represent the nonlinear correlation of features. The OBCC-based, CCA-based and TSCCA-based feature selection approaches were applied to select the optimum combination of features, where 3-fold cross validation of LR was used to determine the number of features. Finally, three LR models as EIS-based PTB risk prediction models were trained on the balanced trainsetII with the three selected combinations of features. More details and setups of the process are described as follows.

1) Oversampling: To balance the trainsetII, the Borderline-SMOTE1 [112] is an oversampling approach which can over-sample the minority examples and majority examples near the borderline only, since classifying the examples near borderline is a common challenge of most of classifiers. The parameters of Borderline-SMOTE1 approach include m, $pnum$ and $nnum$, where m is a number to create the borderline, for example, if a minority example has n nearest neighbours of majority examples and $\frac{m}{2} \leq n < m$, this minority example is regarded as 'DANGER' and used to generate the new data; $pnum$ and $nnum$ are the numbers of new minority examples and new majority examples. In this work, m, $pnum$ and $nnum$ were set up as 10, 204 and 0, so that 204 new PTB data were generated to balance the trainsetII.

2) Polynomial Feature Structure: The polynomial feature structure with a given nonlinear degree of four was used to the 28 EIS features, after that, the balanced trainsetII was extended

to 35960 features ($\frac{(n+l)!}{n!\cdot l!} = \frac{(28+4)!}{28!\cdot 4!} = 35960$). The polynomial features have a high dimensionality, which can lead to overfitting and time-consuming problems in training process. To deal with the problems, feature selection approaches were employed in next step.

3) Feature Selection: In this step, OBCC-based, CCA-based and TSCCA-based feature selection approaches were used to select three optimum combinations of features from 35960 polynomial features. The maximum numbers of features to be selected of OBCC-based and CCA-based feature selection approaches were both set up as $n_{max} = 16$ (according to experiment). The OBCC-based feature selection approach and the CCA-based feature selection approach chose 16 features, feature by feature, respectively. Once a feature selected into the feature set, a general LR model was trained between the feature set and delivery results, and a 3-fold cross validation AUC of LR was computed and saved. Then, the optimal number of features was determined according to the highest 3-fold cross validation AUC of SVM. For TSCCA-based feature selection approach, since its first stage was to implement CCA-based feature selection approach, its optimum number of features was determined by the CCA-based feature selection approach at first stage. Then, TSCCA-based feature selection approach started to refine the temporary optimum combination of features at second stage to obtain the final optimum combination of features.

4) Train Logistic Regression: In this step, for each trainsetII, three logistic regression models (shown in Table 4.5) were trained on the balanced trainsetII with three selected combinations of polynomial features respectively. The models of each corresponding trainsetII were saved to test the corresponding testing set.

Table 4.5: The description of three kinds of EIS-based PTB risk prediction models. OBCC: orthogonal point-biserial correlation coefficient, CCA: canonical correlation analysis, TSCCA: two-stage canonical correlation analysis.

EIS-based PTB risk prediction models	
Classifier	Feature selection approach
logistic regression	OBCC-based
logistic regression	CCA-based
logistic regression	TSCCA-based

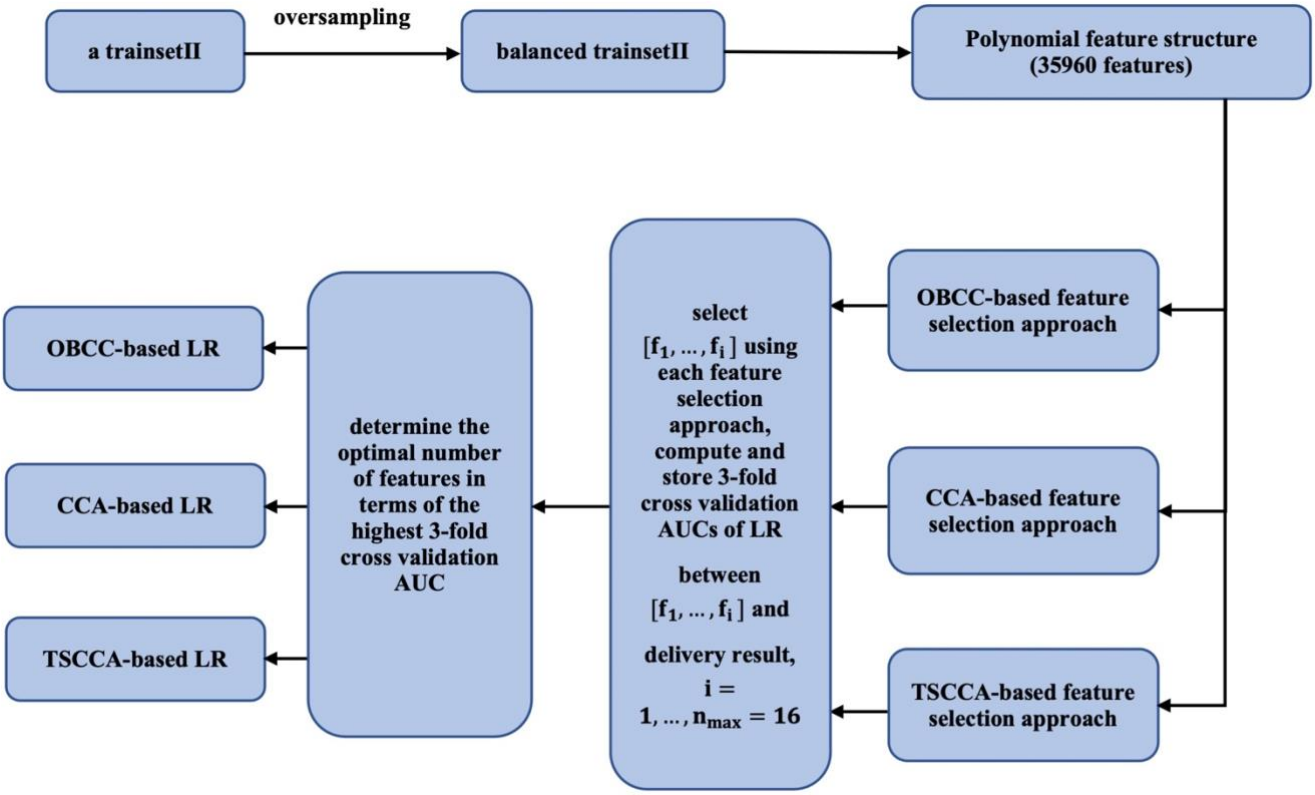

Figure 4.10: The proposed machine learning pipeline for training EIS-based PTB risk prediction model, and the training process repeated on 100 trainsetII, so 100 OBCC-based LR, CCA-based LR and TSCCA-based LR models were trained and saved. AUC: area under the receiver operating characteristic curve, LR: logistic regression, OBCC: orthogonal point-biserial correlation coefficient, CCA: canonical correlation analysis, TSCCA: two-stage canonical correlation analysis.

4.3 Results of EIS-based PTB Risk Prediction

In this section, performances of training and testing optimal-EIS spectrum filter models and EIS-based PTB risk prediction models are displayed as follows, in terms of median, mean, minimum and maximum of training AUCs on 100 trainsetIs and 100 trainsetIIs, respectively, and of testing AUCs on 100 testIDs (generated in Chapter 4.2 in advance).

For the testing process, 100 trained optimal-EIS spectrum filter models were used to select the optimum EIS spectra for the 100 corresponding testIDs, and the selection results were used to test optimal-EIS spectrum filter models. The selected EIS spectra were then input the corresponding EIS-based PTB risk prediction models, which output probabilities of PTB risk, and the probabilities were used to test EIS-based PTB risk prediction models. Also, best performances of the optimal-EIS spectrum filter and the EIS-based PTB risk prediction model are illustrated including number of features, training and testing AUCs, sensitivity and specificity.

4.3.1 Results of Optimal-EIS Spectrum Filter Models

The training and testing performances of 100 optimal-EIS spectrum filter models (OBCC-based SVM classifiers) are illustrated in Table 4.6 below. Training AUCs of 100 optimal-EIS spectrum filters ranged from 0.61 to 0.78 with the numbers of features from 6 to 16, and testing AUCs were from 0.57 to 0.71. Additionally, Table 4.7 showed that an optimal-EIS spectrum filter model with the highest testing AUC of 0.71 (95% CI: 0.64 – 0.78) was trained with 10 features and performed a training AUC of 0.67 (95% CI: 0.61 – 0.73).

Table 4.6: The training and testing performances of 100 optimal-EIS spectrum filter models. OBCC: orthogonal point-biserial correlation coefficient, SVM: support vector machine, AUC: area under the receiver operating characteristic curve.

100 optimal-EIS spectrum filter models			
Classifier		Training AUC	Testing AUC
OBCC-based SVM	mean	0.71	0.63
	minimum	0.61	0.57
	maximum	0.78	0.71

Table 4.7: The best performance of the optimal-EIS spectrum filter model. OBCC: orthogonal point-biserial correlation coefficient, SVM: support vector machine, AUC: area under the receiver operating characteristic curve, CI: confidence interval.

The best optimal-EIS spectrum filter model			
Classifier	Features	Training AUC (95% CI)	Testing AUC (95% CI)
OBCC-based SVM	10	0.67 (0.61 – 0.73)	0.71 (0.64 – 0.78)

4.3.2 Results of EIS-based PTB Risk Prediction Models

The training and testing performances of 300 EIS-based PTB risk prediction models (100 OBCC-based LR models, 100 CCA-based LR models and 100 TSCCA-based LR models) are illustrated in Table 4.8 and Table 4.9 below. Training AUCs of 100 OBCC-based LR models ranged from 0.76 to 0.88, training AUCs of 100 CCA-based LR models ranged from 0.75 to 0.88, and training AUCs of 100 TSCCA-based LR models ranged from 0.77 to 0.89 (all models were trained with four to eight features). Testing AUCs of 100 OBCC-based LR models were between 0.23 and 0.85, testing AUCs of 100 CCA-based LR models were between 0.23 and 0.83, and testing AUCs of 100 TSCCA-based LR models were between 0.23 to 0.81. In testing performances, there were a few extreme testing AUC values ranging from 0.23 to 0.3, which would affect the overall performance. Therefore, the median AUC was employed comparing to the mean AUC, to present the distribution of AUCs, reveal potential limitations of the proposed machine learning strategy and propose improvements.

Additionally, Table 4.10 and Table 4.11 showed that an OBCC-based LR model with the highest testing AUC of 0.85 (95% CI: 0.77 – 0.93) was trained with six features and performed a training AUC of 0.81 (95% CI: 0.74 – 0.86); a CCA-based LR model with the highest testing AUC of 0.83 (95% CI: 0.75 – 0.91) was trained with eight features and performed a training AUC of 0.80 (95% CI: 0.73 – 0.85); and a TSCCA-based LR model with the highest testing AUC of 0.81 (95% CI: 0.74 – 0.88) was trained with seven features and performed a training

AUC of 0.80 (95% CI: 0.73 – 0.85), respectively. Also, the three Receiver Operating Characteristics (ROC) curves were plotted in Figure 4.11(left) to present training performances of the best OBCC-based LR model, the best CCA-based LR model and the best TSCCA-based LR model, and the three ROC curves were plotted in Figure 4.11(right) to present their testing performances.

Table 4.8: The training performances of 300 EIS-based PTB risk prediction models (100 OBCC-based LR models, 100 CCA-based LR models and 100 TSCCA-based LR models). Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. OBCC: orthogonal point-biserial correlation coefficient, CCA: canonical correlation analysis, TSCCA: two-stage canonical correlation analysis, LR: logistic regression, AUC: area under the receiver operating characteristic curve.

300 EIS-based PTB risk prediction models			
PTB models	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 OBCC-based LR models	0.82, 0.82 (0.76, 0.88)	0.80, 0.79 (0.61, 0.93)	0.75, 0.75 (0.59, 0.92)
100 CCA-based LR models	0.82, 0.82 (0.75, 0.88)	0.79, 0.78 (0.59, 0.95)	0.77, 0.76 (0.60, 0.92)
100 TSCCA-based LR models	0.83, 0.83 (0.77, 0.89)	0.79, 0.78 (0.59, 0.93)	0.76, 0.76 (0.60, 0.96)

Table 4.9: The testing performances of 300 EIS-based PTB risk prediction models (100 OBCC-based LR models, 100 CCA-based LR models and 100 TSCCA-based LR models). Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set.

300 EIS-based PTB risk prediction models			
PTB models	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 OBCC-based LR models	0.72, 0.68 (0.23, 0.85)	0.67, 0.64 (0.32, 0.95)	0.72, 0.65 (0.01, 0.90)
100 CCA-based LR models	0.70, 0.66 (0.23, 0.83)	0.64, 0.62 (0.32, 0.89)	0.73, 0.66 (0.01, 0.96)
100 TSCCA-based LR models	0.68, 0.63 (0.23, 0.81)	0.63, 0.62 (0.27, 0.91)	0.73, 0.65 (0.04, 0.96)

Table 4.10: The best training performances of EIS-based PTB risk prediction models (OBCC-based LR model, CCA-based LR model, and TSCCA-based LR model). Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set (optimal operating point of OBCC-based LR model = 0.13, optimal operating point of CCA-based LR model = 0.17, and optimal operating point of TSCCA-based LR model = 0.18), p-values were obtained using the Student's t-test.

The best EIS-based PTB risk prediction models					
PTB models	Features	Training AUC (95% CI)	Sensitivity	Specificity	p-value
OBCC-based LR	6	0.81 (0.74 – 0.86)	0.85	0.68	<0.00001
CCA-based LR	8	0.80 (0.73 – 0.85)	0.83	0.66	<0.00001
TSCCA-based LR	7	0.80 (0.73 – 0.85)	0.83	0.66	<0.0001

Table 4.11: The best testing performances of EIS-based PTB risk prediction models (OBCC-based LR model, CCA-based LR model, and TSCCA-based LR model). Sensitivity and specificity were obtained using the optimal operating point; p-values were obtained using the Two-Sample t-test.

The best EIS-based PTB risk prediction models				
PTB models	Testing AUC (95% CI)	Sensitivity	Specificity	p-value
OBCC-based LR	0.85 (0.77 – 0.93)	0.70	0.87	<0.0001
CCA-based LR	0.83 (0.75 – 0.91)	0.67	0.83	<0.0001
TSCCA-based LR	0.81 (0.74 – 0.88)	0.65	0.80	<0.0001

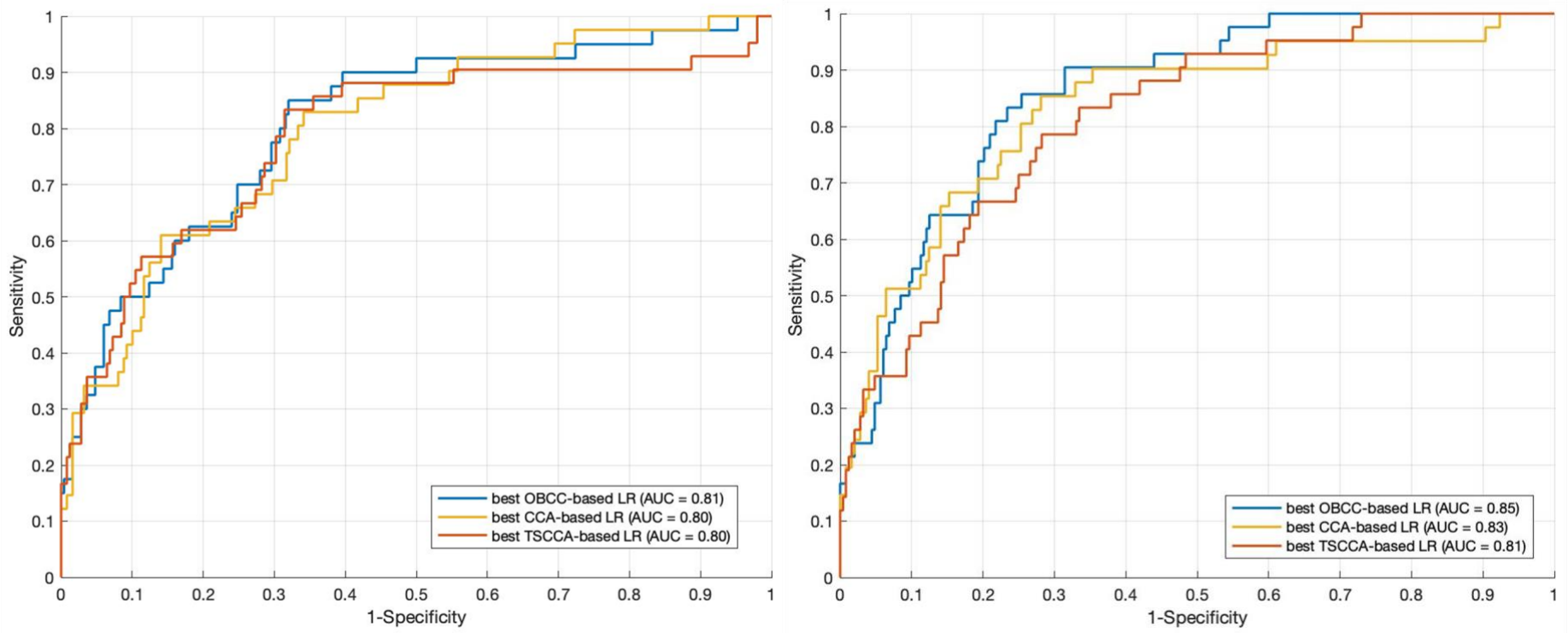

Figure 4.11: The plot (left) of Receiver Operating Characteristics (ROC) curves depicting the training performances of the best OBCC-based LR model, the best CCA-based LR model and the best TSCCA-based LR model. The plot (right) of Receiver Operating Characteristics (ROC) curves depicting the testing performances. OBCC: orthogonal point-biserial correlation coefficient, CCA: canonical correlation analysis, TSCCA: two-stage canonical correlation analysis, LR: logistic regression.

4.4 Discussions

The performances based on the proposed machine learning strategy in this work were compared with a general machine learning strategy, which excluded training the optimal-EIS spectrum model, oversampling the imbalanced training set, or determining optimal feature counts using k-fold cross validation. According to the general machine learning strategy, each woman's EIS spectra were averaged firstly, and the result was used to train a PTB risk prediction model. OBCC-based, CCA-based and TSCCA-based feature selection approaches were also used to determine the optimum combination of features with a fixed number of features (10 features). The training sets and testing sets were also generated from 100 trainIDs and 100 testIDs (split in advance). Then, a logistic regression was employed as a classifier, so that 100 OBCC-based LR models, 100 CCA-based LR models, and 100 TSCCA-based LR models were trained based on this strategy. Their training performances and testing performances were evaluated on 100 training sets and 100 testing sets respectively, and the performances were presented as median, mean, minimum, and maximum AUC, sensitivity, and specificity. Table 4.12 and Table 4.13 present the training and testing performance comparison between the proposed strategy in this work and the general strategy for the EIS-based PTB risk prediction.

From the Table 4.12 and Table 4.13, either training performances or testing performances of the proposed EIS-based PTB risk prediction models on EIS data for 438 recruited women were much better than that of the general PTB risk prediction models. The highest median and mean training AUCs were 0.83 and 0.83 (given by 100 proposed TSCCA-based LR models), which were better than the highest median and mean training AUCs of 0.76 and 0.76 (given by 100 general TSCCA-based LR models). Also, the highest median and mean testing AUCs were 0.72 and 0.68 (given by 100 proposed OBCC-based LR models), which were better than the highest median and mean testing AUCs of 0.53 and 0.52 (given by 100 general OBCC-based LR models).

Table 4.12: Training performance comparison between the proposed machine learning strategy in this work and the general machine learning strategy for the EIS-based PTB risk prediction. Compared to the proposed machine learning strategy, the general machine learning strategy excluded training the optimal-EIS spectrum model, oversampling the imbalanced training set, or determining optimal feature counts using k-fold cross validation. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. OBCC: orthogonal point-biserial correlation coefficient, CCA: canonical correlation analysis, TSCCA: two-stage canonical correlation analysis, LR: logistic regression, AUC: area under the receiver operating characteristic curve.

600 EIS-based PTB risk prediction models			
PTB models	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 OBCC-based LR models (proposed)	0.82, 0.82 (0.76, 0.88)	0.80, 0.79 (0.61, 0.93)	0.75, 0.75 (0.59, 0.92)
100 CCA-based LR models (proposed)	0.82, 0.82 (0.75, 0.88)	0.79, 0.78 (0.59, 0.95)	0.77, 0.76 (0.60, 0.92)
100 TSCCA-based LR models (proposed)	0.83, 0.83 (0.77, 0.89)	0.79, 0.78 (0.59, 0.93)	0.76, 0.76 (0.60, 0.96)
100 OBCC-based LR models (general)	0.72, 0.73 (0.66, 0.81)	0.64, 0.65 (0.36, 1)	0.75, 0.72 (0.31, 0.97)
100 CCA-based LR models (general)	0.73, 0.73 (0.68, 0.81)	0.68, 0.67 (0.39, 0.98)	0.72, 0.70 (0.35, 0.93)
100 TSCCA-based LR models (general)	0.76, 0.76 (0.68, 0.84)	0.69, 0.70 (0.35, 0.98)	0.73, 0.71 (0.36, 0.97)

Table 4.13: Testing performance comparison between the proposed machine learning strategy in this work and the general machine learning strategy for the EIS-based PTB risk prediction. Compared to the proposed machine learning strategy, the general machine learning strategy excluded training the optimal-EIS spectrum model, oversampling the imbalanced training set, or determining optimal feature counts using k-fold cross validation. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set.

600 EIS-based PTB risk prediction models			
PTB models	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 OBCC-based LR models (proposed)	0.72, 0.68 (0.23, 0.85)	0.67, 0.64 (0.32, 0.95)	0.72, 0.65 (0.01, 0.90)
100 CCA-based LR models (proposed)	0.70, 0.66 (0.23, 0.83)	0.64, 0.62 (0.32, 0.89)	0.73, 0.66 (0.01, 0.96)
100 TSCCA-based LR models (proposed)	0.68, 0.63 (0.23, 0.81)	0.63, 0.62 (0.27, 0.91)	0.73, 0.65 (0.04, 0.96)
100 OBCC-based LR models (general)	0.53, 0.52 (0.38, 0.69)	0.41, 0.42 (0, 0.82)	0.63, 0.60 (0.23, 0.88)
100 CCA-based LR models (general)	0.52, 0.52 (0.39, 0.63)	0.40, 0.42 (0.05, 0.83)	0.60, 0.59 (0.19, 0.93)
100 TSCCA-based LR models (general)	0.51, 0.52 (0.40, 0.65)	0.41, 0.43 (0.04, 1)	0.61, 0.58 (0.01, 0.86)

4.5 Conclusion

This chapter has applied machine learning techniques to analyse the cervical EIS data measured at 19-23 weeks' pregnancy duration for PTB prediction risk. Two machining learning pipelines were proposed to train the optimal-EIS spectrum filter model and the EIS-based PTB risk prediction model, so that 100 OBCC-based SVM models (optimal-EIS spectrum filter models) were created, and 300 EIS-based PTB risk prediction models ((100 OBCC-based LR models, 100 CCA-based LR models and 100 TSCCA-based LR models) were obtained. The best optimal-EIS spectrum filter model had the highest testing AUC of 0.71 (95% CI: 0.64 – 0.78). The best EIS-based PTB risk prediction model (from an OBCC-based LR model) performed the highest testing AUC of 0.85 (95% CI: 0.77 – 0.93, sensitivity: 0.70, specificity: 0.87).

Furthermore, comparing the proposed PTB risk prediction models in this work to PTB risk prediction models based on a general machine learning strategy, performances of the proposed PTB risk prediction models had obvious improvements on either training or testing EIS data. So, the optimal-EIS spectrum selection principle, training optimal-EIS spectrum filter model and determining optimal number of features have seemed to achieve the improvement on accuracy of PTB risk prediction, and met the challenge of EIS data heterogeneity.

Chapter 5

5. EIS and Maternal Characteristics-based PTB Risk Prediction

This chapter was a continuation of the prior work in Chapter 4, with the objective of proposing a combining approach that combined a cervical EIS-based PTB risk prediction model with a maternal characteristics-based PTB risk prediction model for a more accurate PTB risk prediction. The present work tried to improve the accuracy of PTB risk prediction based on maternal characteristics data alone or cervical EIS data alone, since the combined model involved integrating the predictions of numerous models [118-120], each of which used different types of input information to generate a prediction which often achieved a better performance of prediction.

Based on this hypothesis, the work proposed a models-combining approach to combine a cervical EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction model. Cervical EIS data were measured from 438 women at 19 to 23 weeks' gestation and their maternal characteristics data were also recorded including demographics and obstetric data. Besides, CL and fFN data of the 438 women were also applied as a comparison for PTB risk prediction. Three machining learning pipelines were proposed to train the maternal characteristics-based PTB risk prediction model, the CL/fFN-based PTB risk prediction model, and the EIS-based PTB risk prediction model. After that, a SVM model and a LR model were employed to combine two probabilities of a maternal characteristics-based PTB risk prediction model and an EIS-based PTB risk prediction model. Finally, training and testing performances of each model were displayed and analysed, and the results were also compared with the existing PTB risk prediction models.

5.1 Data Analysis

In this chapter, a total of 438 women recruited by the ECCLIPPx^{TM} project were also applied, whose ages ranged from 17 to 45. 197 of women had a higher risk of premature birth (45%), 241 of women had a lower risk of preterm birth, 374 of women delivered on term (85.4%), and 64 of women delivered prematurely (14.6%). These women visited at between 19 to 23 weeks of gestation, and the ECCLIPPx^{TM} group collected their maternal characteristics including demographics and obestric history, measured their CL, tested and recorded their fFN values, additionally measured EIS spectra. These data were described and analysed as follows.

1) Maternal characteristics: Maternal characteristics (Table 5.1) consist of demographics and obstetric history, where demographics are 438 women's age, smoking behaviour, alcohol-taking behaviour in their pregnancy, non-prescribed medicine behaviour in pregnancy and BMI, and obstetric history recorded the numbers of previous preterm births and on-term births, the number of previous pregnancies, the number of previous early miscarriages and mid-trimester (13 to 28 weeks of gestation) pregnancy loss. Of these, women's age and BMI, and all records of obstetric history are numerical data type, while smoking behaviour, alcohol-taking behaviour and non-prescribed drugs-taking behaviour in pregnancy are Yes/No data type. The Yes/No data were processed by labelling 'Yes' as '1' and labelling 'No' as '0'.

More details are illustrated in Table 5.2. 368 of 438 women (84%) had no smoking history, while 70 women smoke. 399 of 438 pregnant women (91%) did not take alcohol, while 39 of women had alcohol in pregnancy. 350 of 438 women (80%) did not take non-prescribed medicines during pregnancy, while 88 women took in pregnancy. In addition, 438 women's BMI ranged from 17.6 to 49.7. In terms of obstetric history, 62.8% of the women (275 of 438 women) had no previous preterm births, while 30.6% of them experienced one premature delivery and 6.6% of them had two to four premature births. 58.2% of the women (255 of 438 women) had no previous on-term births, while 27.4% had one on-term delivery only and 14.4% had more than one on-term births. 25.8% of women (113 of 438 women) had never been pregnant before, while 26% of women had one pregnancy, 20.3% of them experienced two prior pregnancies and 27.9% of them had three or more prior pregnancies. 69.5% of women (304 of 438 women) did not previously experience an early miscarriage, while 22% of the women had a prior early miscarriage and 8.5% of them experienced two or more previous early miscarriages. 88.4% of women (387 of 438 women) did not experience a miscarriage during

the second trimester. The remaining 11.6% of women experienced at least one miscarriage during the second trimester.

Table 5.1: Maternal characteristics of 438 recruited women.

Type	Characteristics
Demographics	Age (numeric)
	Smoker (Yes/No)
	Alcohol in pregnancy (Yes/No)
	Non-prescribed medicine in pregnancy (Yes/No)
	Body Mass Index (BMI) (numeric)
Obstetric History	Number of preterm births (numeric)
	Number of on-term births (numeric)
	Number of previous pregnancies (numeric)
	Number of previous early miscarriages (numeric)
	Number of mid-trimester pregnancy loss (numeric)

Table 5.2: The statistics of maternal characteristics of 438 recruited women.

Characteristics	Description
Age (numeric)	17 to 45
Smoker (Yes/No)	70 / 368
Alcohol in pregnancy (Yes/No)	39 / 399
Non-prescribed drugs in pregnancy (Yes/No)	88 / 350
Body Mass Index (BMI) (numeric)	17.6 to 49.7
Previous preterm births (Yes/No)	163 / 275
Previous on-term births (Yes/No)	180 / 255
Previous pregnancies (Yes/No)	325 / 113
Previous early miscarriages (Yes/No)	134 / 304
Previous mid-trimester pregnancy loss (Yes/No)	51 / 387

2) Cervical length and fetal fibronectin: CL and fFN were measured and tested for 438 women (Figure 5.1) at their visits between 19 to 23 weeks of gestation. Cervical lengths of 438 women

ranged from 9 mm to 66 mm, with average of 37.6 mm, 374 of that were from on-term birth women ranging from 18 mm to 66 mm and with average of 38.7 mm, while 64 of that were from preterm birth women with the range from 9 mm to 53 mm and a shorter average cervical length of 31.3 mm. Additionally, fFN test values of 438 women ranged from 1 to 497 ng/ml and the average was 27.4 ng/ml, that of on-term birth women were from 1 to 497 ng/ml and with average of 20.6 ng/ml, while that of preterm birth women were from 1 to 453 ng/ml and with a much higher average of 66.9 ng/ml.

More detailed statistics of CL and fFN test values were concluded in Figure 5.2. For on-term birth women, 9.6% of them (36 of 374 women) were with cervical lengths less than 33 mm, 3.2% of whose cervical lengths were less than 25 mm. Whereas, for preterm term birth women, 42.2% of them (27 of 64 women) were with cervical lengths less than 33 mm, 25% and 3.1% of whose cervical lengths were less than 25 mm and 15 mm respectively. Regarding fFN test values, 8.8% of on-term birth women (33 of 374 women) had values more that 50 ng/ml and 1.9% of values were more than 200 ng/ml. However, 34.4% of preterm birth women had fFN test values greater than 50 ng/ml and 15.6% of values are greater than 200 ng/ml.

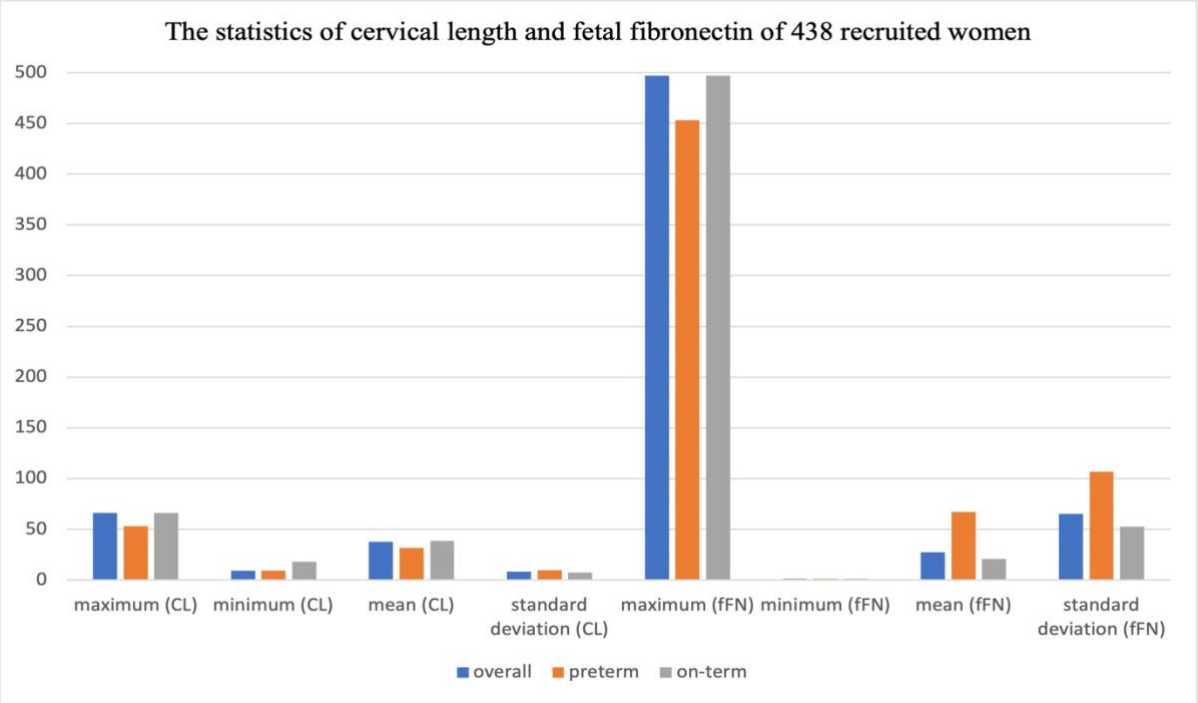

Figure 5.1: The histogram showing maximums, minimums, means and standard deviations of cervical length (CL) and fetal fibronectin (fFN), overall 438 women, 64 preterm women, and 374 on-term women.

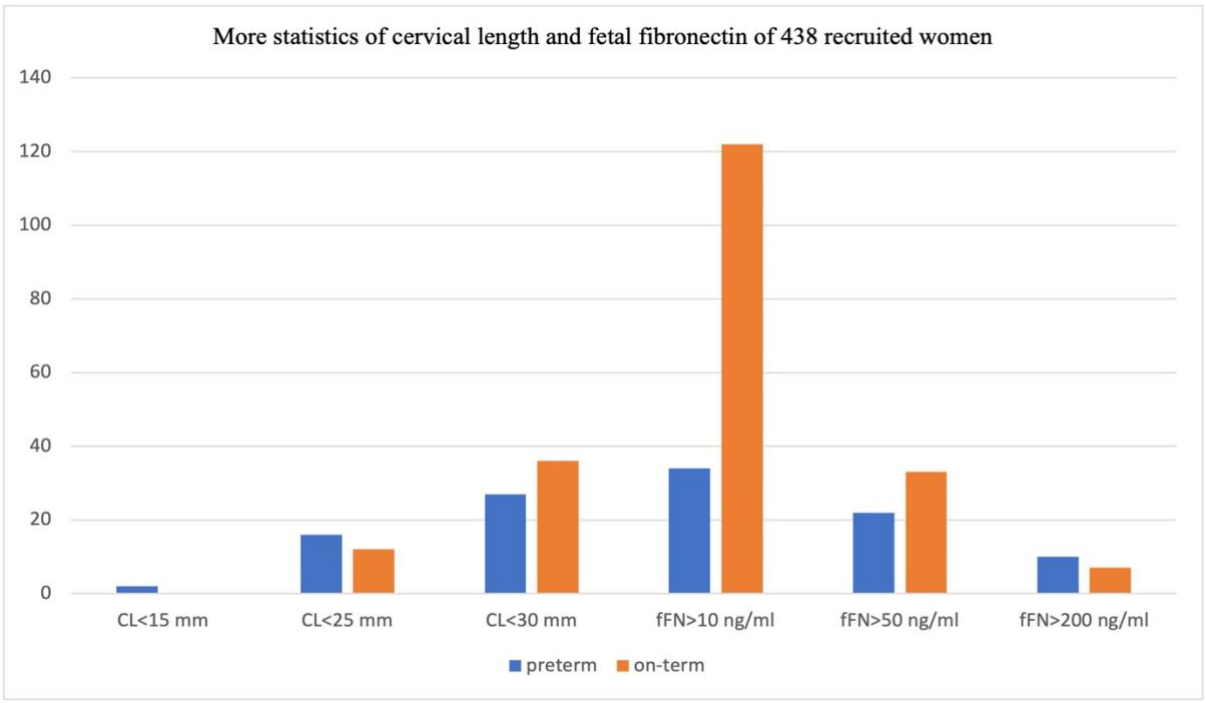

Figure 5.2: The histogram showing the distributions of cervical length (CL)'s ranges and fetal fibronectin (fFN)'s ranges of preterm women and on-term women.

3) Electrical impedance spectroscopy data: 438 recruited women measured their EIS spectra over 14 frequencies ranging from 76.3 Hz to 625 kHz at between 19 to 23 weeks of gestation. The EIS spectra were a complicated structure consisting of amplitude and phase or real and imaginary parts. In addition, there were 28 EIS features measured once for each woman, and the measurements were performed between one and three times at random throughout their visits. Therefore, 438 women have a total of 1022 spectrum readings, with 41 women having just one spectrum reading, 210 women having only two spectrum readings, and 187 women having three spectrum readings. All EIS spectra of each woman were processed in terms of the proposed principle for determination of labels for EIS spectrum readings (Principle 4.1 in Chapter 4), to emphasize the differences cervical tissues between preterm birth women and on-term birth women, and the statistics of amplitudes and phases of selected EIS spectrum readings of preterm birth women and on-term birth women over 14 frequencies were calculated and concluded in Figure 5.3 and Figure 5.4, respectively. After labelling for EIS spectra, the EIS dataset included selection results (label of '1' for selected spectrum and label of '0' for unselected spectrum) and delivery results (label of '1' for preterm birth and label of '0' for on-term birth).

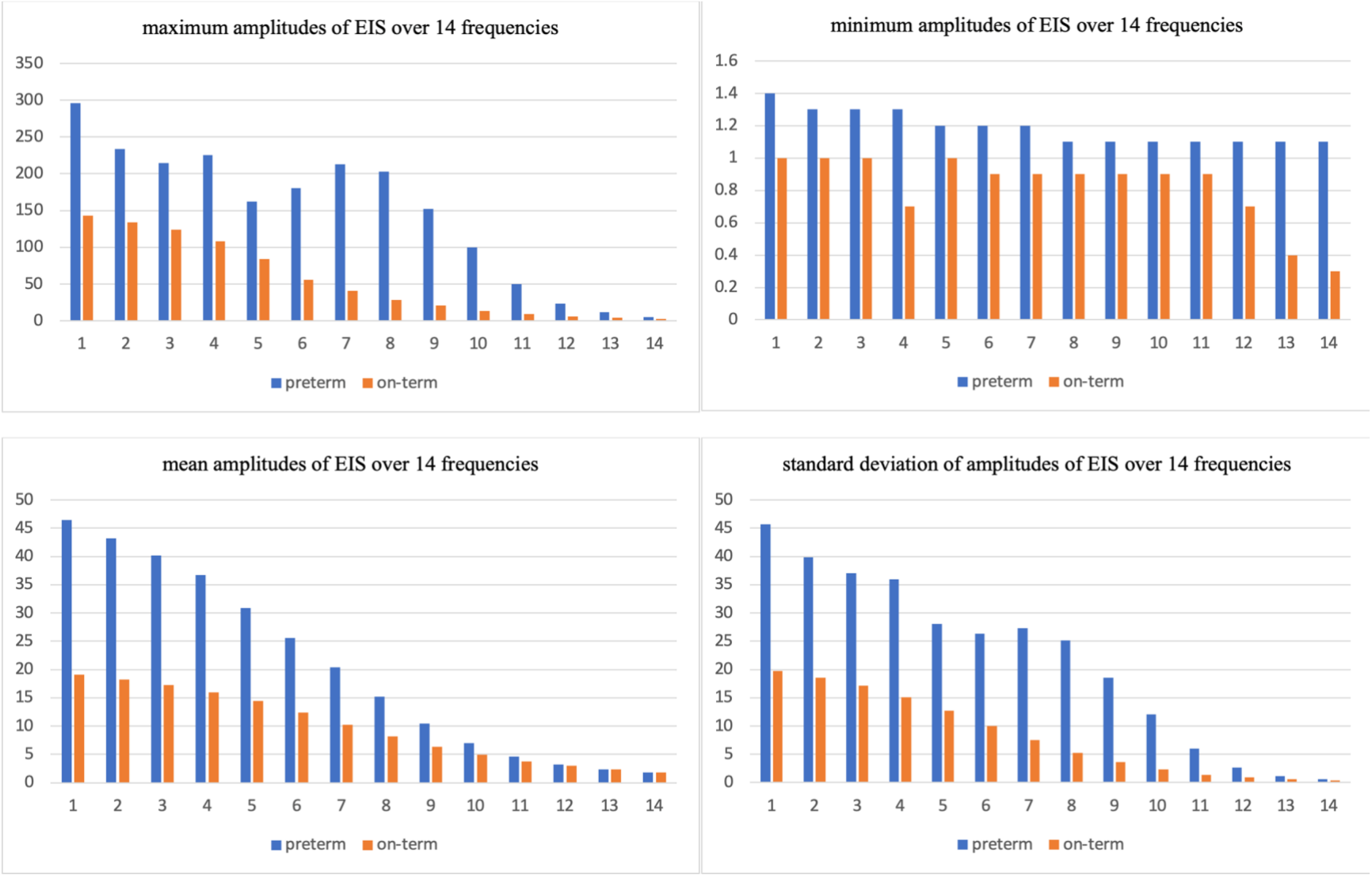

Figure 5.3: The histogram showing maximums (top-left), minimums (top-right), means (bottom-left) and standard deviations (bottom-right) of amplitudes of electrical impedance spectroscopy (EIS) over 64 preterm women and 374 on-term women.

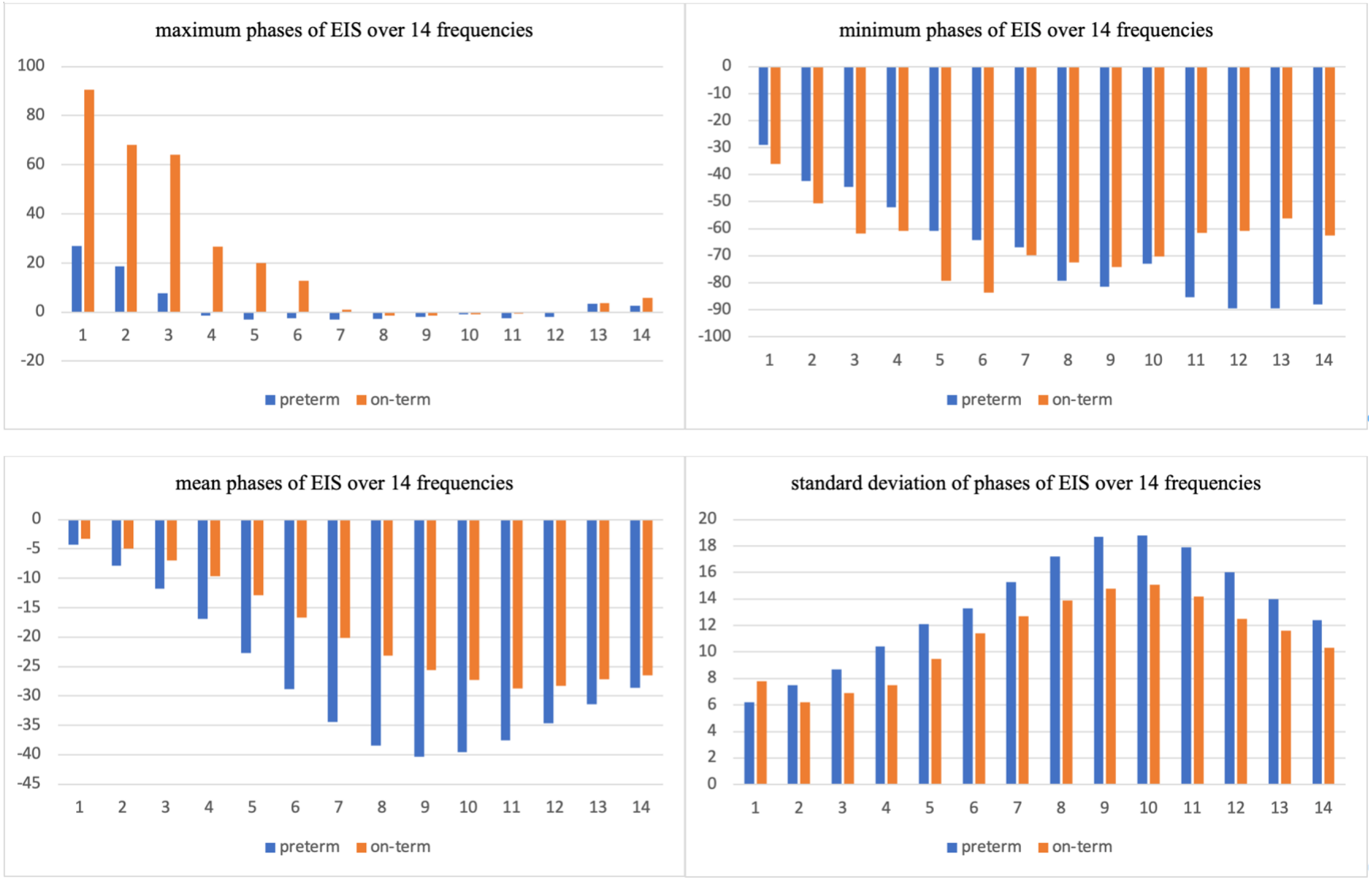

Figure 5.4: The histogram showing maximums (top-left), minimums (top-right), means (bottom-left) and standard deviations (bottom-right) of phrases of electrical impedance spectroscopy (EIS) over 64 preterm women and 374 on-term women.

5.2 Methodology

This work was to 1) train maternal characteristics-based, CL-based and fFN-based PTB risk prediction models, and 2) combine maternal characteristics-based and EIS-based PTB risk prediction models. To achieve it, firstly, three machine learning pipelines were proposed below, for training maternal characteristics-based, CL/fFN-based and EIS-based PTB risk prediction models respectively. Then, two approaches of combining models were proposed in details below, including the support vector machine as a combiner and the logistic regression as a combiner to combine maternal characteristics-based and EIS-based PTB risk prediction models, by inputting probabilities from the maternal characteristics-based model and the EIS-based PTB risk prediction model into SVM and LR classifier and outputting a final probability of PTB risk.

5.2.1 Machine Learning Pipelines

The training sets and testing sets were generated according to 100 trainIDs and 100 testIDs (split in Chapter 4 in advance). Each trainID had 290 women including 43 preterm births and 247 on-term births, and each testID had 148 women including 21 preterm births and 127 on-term births.For training maternal characteristics-based model, a maternal characteristics training set included maternal characteristics data of a trainID and their delivery results (label of '1' for PTB and label of '0' for on-term birth). For training CL/fFN-based model, a CL/fFN training set included CL/fFN data of a trainID and their delivery results (label of '1' for PTB and label of '0' for on-term birth). Also, a trainsetI and a trainsetII (used in Chapter 4) were used to train optimal-EIS spectrum filter model and EIS-based PTB risk prediction model respectively.

1) A machine learning pipeline for training maternal characteristics-based PTB risk prediction models: This machine learning pipeline for training maternal characteristics-based PTB risk prediction model is illustrated in Figure 5.5. Firstly, to balance the classes '0' and '1', each maternal characteristics training set was oversampled using the Borderline-SMOTE1

technique [112] (principle explained in Chapter 4), which produced new synthetics of the PTB EIS data and added them into the maternal characteristics training set. The balanced maternal characteristics training set was not transformed to polynomial feature structure, since the maternal characteristics training set consisted of integers ranging from 0 to 45 and more than half of values were 0, so that the transformation to polynomial feature structure would mostly lead to a sparse matrix. Therefore, the balanced maternal characteristics training set was used directly to train the SVM model. Its kernel function was set to 'rbf' and the other parameters of 'BoxConstraint' and 'KernelScale' were optimized using Bayesian Optimization with a custom function of computing 5-fold cross validation AUC.

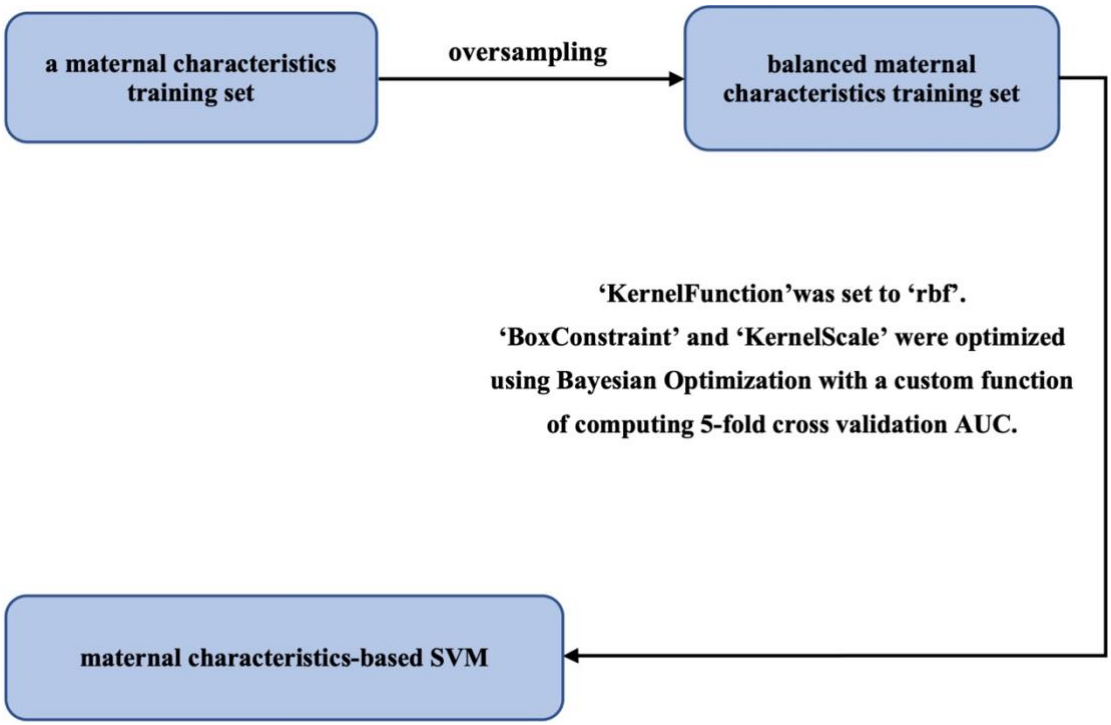

Figure 5.5: The proposed machine learning pipeline for training maternal characteristics-based PTB risk prediction model, and the training process repeated on 100 maternal characteristics training sets, so 100 maternal characteristics-based SVM models were trained and saved. AUC: area under the receiver operating characteristic curve, SVM: support vector machine.

2) A machine learning pipeline for training CL/ fFN-based PTB risk prediction models: The machine learning pipeline for training CL-based and fFN-based PTB risk prediction models is

presented in Figure 5.6, which was a simple machine learning pipeline without polynomial feature structure or any feature selection approach. Each CL/fFN training set was oversampled using the Borderline-SMOTE1 technique [112], and then the balanced CL/fFN training set was used to train a CL/fFN-based logistic regression model, with 5-fold cross validation determining the logistic regression's parameters.

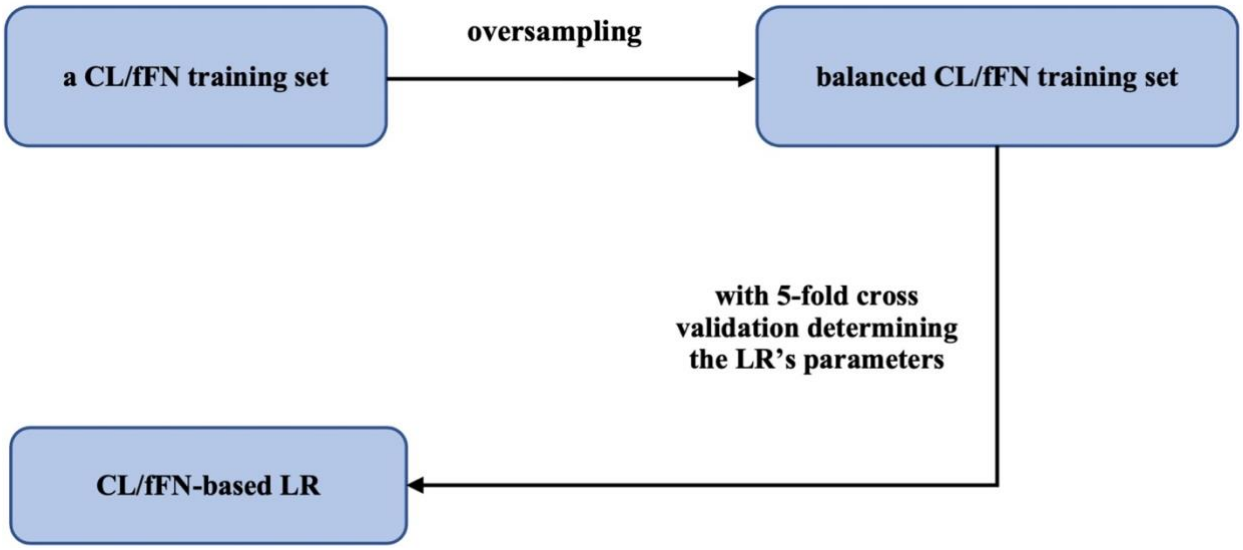

Figure 5.6: The machine learning pipeline for training CL/fFN-based PTB risk prediction models, and the training process repeated on 100 CL/fFN training sets, so 100 CL/fFN-based LR models were trained and saved. CL: cervical length, fFN: fetal fibronectin, LR: logistic regression.

3) A machine learning pipeline for training EIS-based PTB risk prediction models: The training process of optimal-EIS spectrum filter model was proposed in Chapter 4.2.1. For each trainsetI, the mean, standard deviation and centroid of amplitudes and phases of EIS spectra were calculated respectively. Six statistics features were added into trainsetI. The trainsetI was then extended to 7770 features by using a polynomial feature structure with three nonlinear degrees. To identify the optimal combination of features, the OBCC-based feature selection method with $n_{max} = 16$ features and 3-fold cross validation AUCs of SVM was applied. Lastly, a SVM model (optimal-EIS spectrum filter model) was trained on the trainsetI with the selected combination of features.

Meanwhile, since the proposed OBCC-based LR model in Chapter 4 achieved the best performance, this model was applied in this chapter. For training EIS-based PTB risk prediction model (Figure 5.7)**,** each trainsetII was oversampled using the Borderline-SMOTE1 algorithm [112] firstly, which produced 204 new synthetics of PTB data and added them into trainsetII to balance the classes '0' and '1'. The nonlinear correlation of features was then represented by the polynomial feature structure. After that, the OBCC-based feature selection approach [105] was used to choose 16 features, then, the optimal number of features was determined by the highest 3-fold cross validation AUC of LR, Finally, a LR model serving as the EIS-based PTB risk prediction model was trained on the balanced trainsetII with the optimal features.

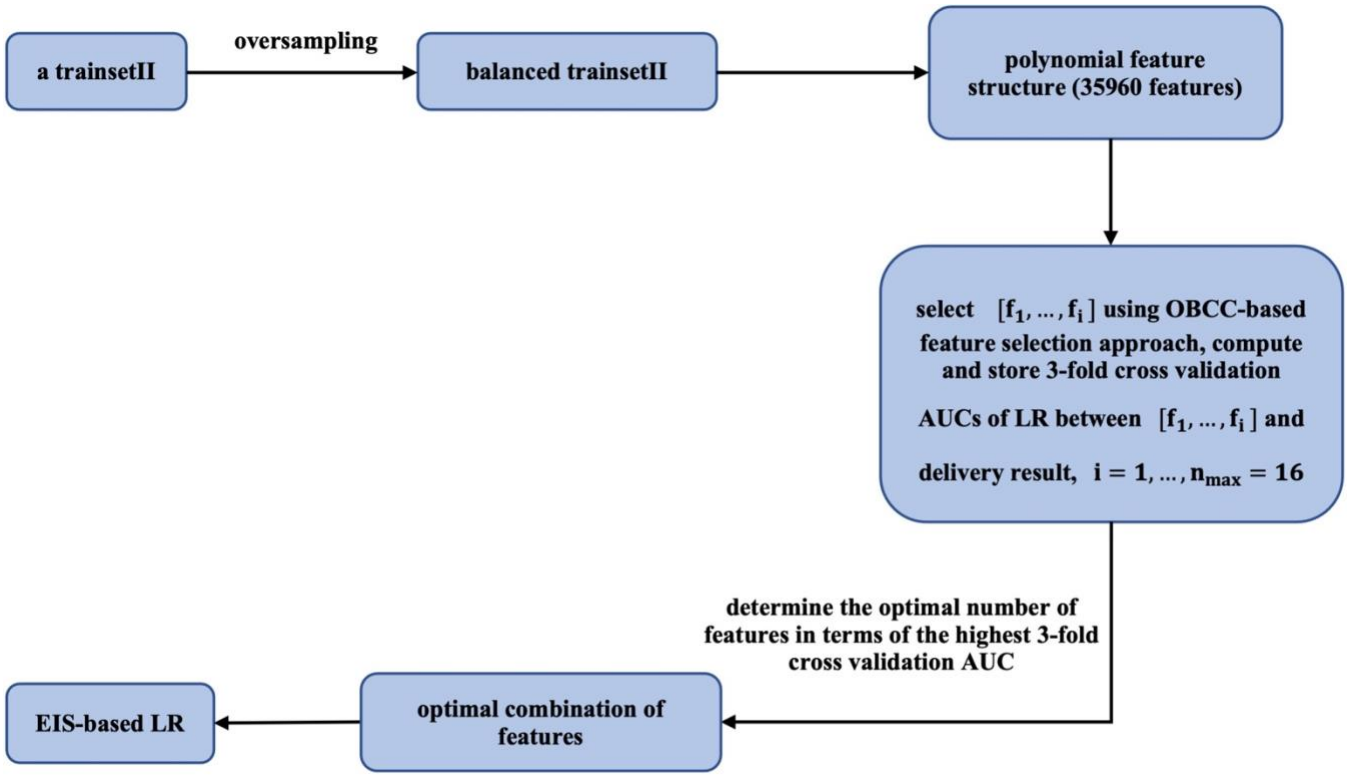

Figure 5.7: The proposed machine learning pipeline for training EIS-based PTB risk prediction models, and the training process repeated on 100 trainsetIIs, so 100 EIS-based LR models were trained and saved. EIS: electrical impedance spectroscopy, OBCC: orthogonal point-biserial correlation coefficient, LR: logistic regression.

5.2.2 Combining Models

Once the maternal characteristics-based and EIS-based PTB risk prediction models have been trained, maternal characteristics and EIS features would be input to the corresponding prediction models respectively. After that, these models output two probabilities of PTB risk, namely probability1 for maternal characteristics, probability2 for EIS features. Then, a SVM and LR were chosen as classifiers to combine probability1 and probability2, and two combined models were built as shown in Table 5.3, and the setups of the SVM and the LR were presented as follows.

Table 5.3: The list of two combined models: co-model1 and co-model2. EIS: electrical impedance spectroscopy, LR: logistic regression, SVM: support vector machine.

Combined models	Classifier	Member models
co-model1	SVM	maternal characteristics-based SVM + EIS-based LR
co-model2	LR	maternal characteristics-based SVM + EIS-based LR

1) Support Vector Machine Classifier: A SVM was used to combine probabilities from two PTB risk prediction models, its kernel function was set to 'rbf' and the other parameters of 'BoxConstraint' and 'KernelScale' were optimized using Bayesian Optimization with a custom function of computing 10-fold cross validation AUC.

2) Logistic Regression Classifier: A LR classifier on probabilities from two PTB risk prediction models was trained by adjusting its parameters using 10-fold cross validation. The values of parameters with the highest 10-fold CV AUC were chosen for the LR model. Then, combined models of co-model1 and co-model2 can output a final probability of PTB risk respectively. The illustration of combining a maternal characteristics-based model and an EIS-based PTB risk prediction model is shown in Figure 5.8.

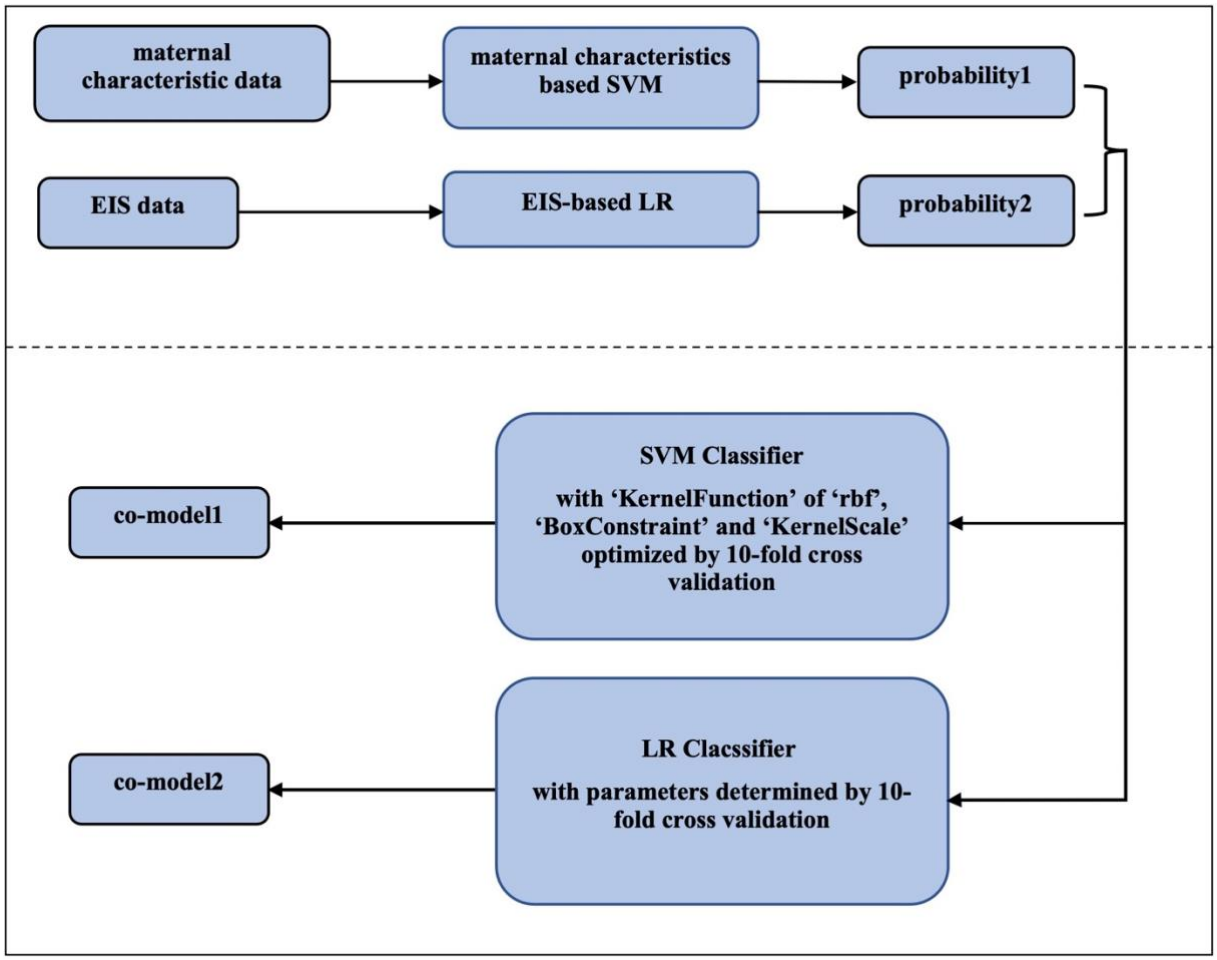

Figure 5.8: The illustration of combining a maternal characteristics-based model and an EIS-based PTB risk prediction model. EIS: electrical impedance spectroscopy, LR: logistic regression, SVM: support vector machine.

5.3 Results of Analysis

In this section, performances of training and testing maternal characteristics-based PTB risk prediction models, CL-based and fFN-based PTB risk prediction models, EIS-based PTB risk prediction models, and combined PTB risk prediction models were displayed as follows, in terms of median, mean, minimum and maximum of training AUCs on 100 trainIDs, respectively, and of testing AUCs on 100 testIDs (generated in Chapter 4.2 in advance). Also, the best training and testing performances of maternal characteristics-based PTB risk prediction

model, CL-based and fFN-based PTB risk prediction models, EIS-based PTB risk prediction model, and combined PTB risk prediction models were illustrated including training and testing AUCs, sensitivity and specificity.

5.3.1 Results of Maternal Characteristics-Based PTB Risk Prediction Models

The training and testing performances of 100 maternal characteristics-based PTB risk prediction models were illustrated in Table 5.4 below. Training AUCs of 100 maternal characteristics-based PTB risk prediction models ranged from 0.66 to 0.79, and testing AUCs were from 0.51 to 0.73. Additionally, Table 5.5 showed that a maternal characteristics-based PTB risk prediction model had the highest testing AUC of 0.73 (95% CI: 0.67 – 0.78, sensitivity: 0.38, specificity: 0.80), and performed a training AUC of 0.63 (95% CI: 0.57 – 0.73, sensitivity: 0.58, specificity: 0.84).

5.3.2 Results of CL-based and fFN-Based PTB Risk Prediction Models

The training and testing performances of 100 CL/fFN-based PTB risk prediction models were illustrated in Table 5.6 below. Training AUCs of 100 CL-based PTB risk prediction models ranged from 0.70 to 0.85, and testing AUCs were from 0.62 to 0.83. Training AUCs of 100 fFN-based PTB risk prediction models ranged from 0.53 to 0.71, and testing AUCs were from 0.50 to 0.81.

Additionally, Table 5.7 showed that a CL-based PTB risk prediction model had the highest testing AUC of 0.83 (95% CI: 0.71 – 0.95, sensitivity: 0.50, specificity: 0.89), and performed a training AUC of 0.70 (95% CI: 0.64 – 0.76, sensitivity: 0.37, specificity: 0.94); a fFN-based PTB risk prediction model gave the highest testing AUC of 0.81 (95% CI: 0.67 – 0.94,

sensitivity: 0.45, specificity: 0.93), and performed a training AUC of 0.54 (95% CI: 0.47 – 0.62, sensitivity: 0.27, specificity: 0.91).

5.3.3 Results of EIS-based PTB Risk Prediction Models

The training and testing performances of 100 EIS-based LR models were illustrated in Table 5.8 below. Training AUCs of 100 EIS-based LR models ranged from 0.76 to 0.88. Testing AUCs of 100 EIS-based LR models were between 0.23 and 0.85. Additionally, Table 5.9 showed that an EIS-based LR model with the highest testing AUC of 0.85 (95% CI: 0.77 – 0.93, sensitivity: 0.70, specificity: 0.87) was trained with six features and performed a training AUC of 0.81 (95% CI: 0.74 – 0.86, sensitivity: 0.85, specificity: 0.68).

5.3.4 Results of Combined PTB Risk Prediction Models

The training and testing performances of 200 combined PTB risk prediction models were illustrated in Table 5.10 below. The co-model1 was using a SVM model to combine two probabilities of a maternal characteristics-based PTB risk prediction model and an EIS-based PTB risk prediction model. The co-model2 used a LR model to combine two probabilities of a maternal characteristics-based PTB risk prediction model and an EIS-based PTB risk prediction model. Training AUCs of 100 co-model1s ranged from 0.80 to 0.89, and their testing AUCs were from 0.33 to 0.89. Training AUCs of 100 co-model2s ranged from 0.77 to 0.90, and testing AUCs were from 0.32 to 0.87.

Additionally, Table 5.11 showed that a co-model1 had the highest testing AUC of 0.89 (95% CI: 0.80 – 0.97, sensitivity: 0.60, specificity: 0.90), and performed a training AUC of 0.82 (95% CI: 0.76 – 0.89, sensitivity: 0.67, specificity: 0.85); a co-model2 gave the highest testing AUC of 0.87 (95% CI: 0.81 – 0.93, sensitivity: 0.64, specificity: 0.89), and performed a training AUC of 0.83 (95% CI: 0.76 – 0.90, sensitivity: 0.69, specificity: 0.85).

Table 5.4: The training and testing performances of 100 maternal characteristics-based PTB risk prediction models. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. AUC: area under the receiver operating characteristic curve, SVM: support vector machine.

100 maternal characteristics-based PTB risk prediction models			
PTB model	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 maternal characteristics-based SVM models	0.73, 0.73 (0.66, 0.79)	0.65, 0.63 (0.38, 0.81)	0.75, 0.73 (0.41, 0.83)
PTB model	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 maternal characteristics-based SVM models	0.68, 0.66 (0.51, 0.73)	0.60, 0.62 (0.32, 0.86)	0.68, 0.67 (0.33, 0.92)

Table 5.5: The best training and testing performances of a maternal characteristics-based PTB risk prediction model. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set (optimal operating point of maternal characteristics-based SVM model = 0.24), p-values were obtained using the Student's t-test.

The best maternal characteristics-based PTB risk prediction model				
PTB model	Training AUC (95% CI)	Sensitivity	Specificity	p-value
maternal characteristics-based SVM model	0.63 (0.57 – 0.73)	0.38	0.80	<0.001
PTB model	Testing AUC (95% CI)	Sensitivity	Specificity	p-value
maternal characteristics-based SVM model	0.73 (0.67 – 0.78)	0.58	0.84	<0.01

Table 5.6: The training and testing performances of 100 CL/fFN-based PTB risk prediction models. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. AUC: area under the receiver operating characteristic curve, SVM: support vector machine.

100 CL/fFN-based PTB risk prediction models			
PTB model	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 CL-based LR models	0.77, 0.77 (0.70, 0.85)	0.74, 0.73 (0.37, 0.99)	0.70, 0.68 (0.40, 0.96)
100 fFN-based LR models	0.63, 0.63 (0.53, 0.71)	0.41, 0.44 (0.27, 0.79)	0.90, 0.87 (0.56, 0.96)
PTB model	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 CL-based LR models	0.73, 0.73 (0.62, 0.83)	0.78, 0.75 (0.38, 0.95)	0.54, 0.55 (0.25, 0.90)
100 fFN-based LR models	0.63, 0.63 (0.50, 0.81)	0.39, 0.45 (0.18, 1)	0.84, 0.74 (0, 0.94)

Table 5.7: The best training and testing performances of CL/fFN-based PTB risk prediction model. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set (optimal operating point of CL-based LR model = 0.67 and optimal operating point of fFN-based LR model = 0.54), p-values were obtained using the Student's t-test.

The best CL/fFN-based PTB risk prediction model				
PTB model	Training AUC (95% CI)	Sensitivity	Specificity	p-value
CL-based LR model	0.70 (0.64 – 0.76)	0.37	0.94	<0.001
fFN-based LR model	0.54 (0.47 – 0.62)	0.27	0.91	<0.001
PTB model	Testing AUC (95% CI)	Sensitivity	Specificity	p-value
CL-based LR models	0.83 (0.71 – 0.95)	0.50	0.89	<0.01
fFN-based LR model	0.81 (0.67 – 0.94)	0.45	0.93	<0.001

Table 5.8: The training and testing performances of 100 EIS-based PTB risk prediction models. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. AUC: area under the receiver operating characteristic curve, LR: logistic regression.

100 EIS-based PTB risk prediction models			
PTB model	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 EIS-based LR models	0.82, 0.82 (0.76, 0.88)	0.80, 0.79 (0.61, 0.93)	0.75, 0.75 (0.59, 0.92)
PTB model	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 EIS-based LR models	0.72, 0.68 (0.23, 0.85)	0.67, 0.64 (0.32, 0.95)	0.72, 0.65 (0.01, 0.90)

Table 5.9: Table: The best training and testing performances of an EIS-based PTB risk prediction model. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set (optimal operating point of EIS-based LR model = 0.13), p-values were obtained using the Student's t-test.

The best EIS-based PTB risk prediction model				
PTB model	Training AUC (95% CI)	Sensitivity	Specificity	p-value
EIS-based LR model	0.81 (0.74 – 0.86)	0.85	0.68	<0.00001
PTB model	Testing AUC (95% CI)	Sensitivity	Specificity	p-value
EIS-based LR model	0.85 (0.77 – 0.93)	0.70	0.87	<0.0001

Table 5.10: The training and testing performances of 200 combined PTB risk prediction models (100 co-model1s and 100 co-model2s). Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. AUC: area under the receiver operating characteristic curve.

200 combined PTB risk prediction models			
PTB model	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 co-model1s	0.84, 0.84 (0.80, 0.89)	0.79, 0.79 (0.55, 1)	0.80, 0.79 (0.66, 0.92)
100 co-model2s	0.82, 0.81 (0.77, 0.90)	0.74, 0.75 (0.58, 0.92)	0.81, 0.80 (0.63, 0.96)
PTB model	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 co-model1s	0.78, 0.73 (0.33, 0.89)	0.65, 0.63 (0.14, 0.87)	0.77, 0.71 (0.13, 0.92)
100 co-model2s	0.76, 0.73 (0.32, 0.87)	0.66, 0.64 (0.21, 0.84)	0.75, 0.70 (0.26, 0.87)

Table 5.11: The best training and testing performances of combined PTB risk prediction models. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set (optimal operating point of co-model1 = 0.14 and optimal operating point of co-model2 = 0.18), p-values were obtained using the Student's t-test.

The best combined PTB risk prediction models				
PTB model	Training AUC (95% CI)	Sensitivity	Specificity	p-value
co-model1	0.82 (0.76 – 0.89)	0.67	0.85	<0.00001
co-model2	0.83 (0.76 – 0.90)	0.69	0.85	<0.00001
PTB models	Testing AUC (95% CI)	Sensitivity	Specificity	p-value
co-model1	0.89 (0.80 – 0.97)	0.60	0.90	<0.001
co-model2	0.87 (0.81– 0.93)	0.64	0.89	<0.001

5.4 Discussions

Table 5.12 presented the training and testing performance comparison of combined PTB prediction models, maternal characteristics-based PTB prediction models, CL/fFN based PTB prediction models, and EIS-based PTB prediction models over 100 times of training and testing. From comparing, the co-model1s showed a better performance on median and mean training AUCs, than PTB prediction models only based on maternal characteristics, CL, fFN and EIS. Also, the median and mean AUCs of combined PTB prediction models (both co-model1 and co-model2) outperformed that of PTB prediction models only based on maternal characteristics, CL, fFN and EIS.

From the Table 5.13, either training performances or testing performances of the combined PTB prediction models on EIS data for 438 recruited women were better than that of PTB prediction models only based on maternal characteristics, CL, fFN and EIS. The co-model1 and co-model2 had the training AUCs of 0.82 and 0.83, which were better than training AUCs of 0. 63, 0.70, 0.54 and 0.81 (given by a maternal characteristics-based PTB prediction model, a CL-based PTB prediction model, a fFN-based PTB prediction model and an EIS-based PTB prediction model, respectively). Also, the co-model1 and co-model2 had the testing AUCs of 0.89 and 0.87, which were better than testing AUCs of 0. 73, 0.83, 0.81 and 0.85 (given by a maternal characteristics-based PTB prediction model, a CL-based PTB prediction model, a fFN-based PTB prediction model and an EIS-based PTB prediction model, respectively). Besides, from the performance comparison between the proposed combined models of this work and existing models [20,86] in Table 5.14, the combined PTB risk prediction models also had greater predictions than the existing models.

Table 5.12: The training and testing performances of 600 PTB risk prediction models (200 combined PTB prediction models, 100 maternal characteristics-based PTB prediction models, 100 CL/fFN-based PTB prediction models and 100 EIS-based PTB prediction models). Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. AUC: area under the receiver operating characteristic curve, LR: logistic regression, SVM: support vector machine.

600 PTB risk prediction models			
PTB model	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 co-model1s	0.84, 0.84 (0.80, 0.89)	0.79, 0.79 (0.55, 1)	0.80, 0.79 (0.66, 0.92)
100 co-model2s	0.82, 0.81 (0.77, 0.90)	0.74, 0.75 (0.58, 0.92)	0.81, 0.80 (0.63, 0.96)
100 maternal characteristics-based SVM models	0.73, 0.73 (0.66, 0.79)	0.65, 0.63 (0.38, 0.81)	0.75, 0.73 (0.41, 0.83)
100 CL-based LR models	0.77, 0.77 (0.70, 0.85)	0.74, 0.73 (0.37, 0.99)	0.70, 0.68 (0.40, 0.96)
100 fFN-based LR models	0.63, 0.63 (0.53, 0.71)	0.41, 0.44 (0.27, 0.79)	0.90, 0.87 (0.56, 0.96)
100 EIS-based LR models	0.82, 0.82 (0.76, 0.88)	0.80, 0.79 (0.61, 0.93)	0.75, 0.75 (0.59, 0.92)
PTB model	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
100 co-model1s	0.78, 0.73 (0.33, 0.89)	0.65, 0.63 (0.14, 0.87)	0.77, 0.71 (0.13, 0.92)
100 co-model2s	0.76, 0.73 (0.32, 0.87)	0.66, 0.64 (0.21, 0.84)	0.75, 0.70 (0.26, 0.87)
100 maternal characteristics-based SVM models	0.68, 0.66 (0.51, 0.73)	0.60, 0.62 (0.32, 0.86)	0.68, 0.67 (0.33, 0.92)
100 CL-based LR models	0.73, 0.73 (0.62, 0.83)	0.78, 0.75 (0.38, 0.95)	0.54, 0.55 (0.25, 0.90)
100 fFN-based LR models	0.63, 0.63 (0.50, 0.81)	0.39, 0.45 (0.18, 1)	0.84, 0.74 (0, 0.94)
100 EIS-based LR models	0.72, 0.68 (0.23, 0.85)	0.67, 0.64 (0.32, 0.95)	0.72, 0.65 (0.01, 0.90)

Table 5.13: The best training and testing performances of six PTB prediction model. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set (optimal operating point of co-model1 = 0.14, optimal operating point of co-model2 = 0.18, optimal operating point of maternal characteristics-based SVM model = 0.24, optimal operating point of CL-based LR model = 0.67, optimal operating point of fFN-based LR model = 0.54, and optimal operating point of EIS-based LR model = 0.13), p-values were obtained using the Student's t-test.

The best PTB risk prediction models				
PTB model	Training AUC (95% CI)	Sensitivity	Specificity	p-value
co-model1	0.82 (0.76 – 0.89)	0.67	0.85	<0.00001
co-model2	0.83 (0.76 – 0.90)	0.69	0.85	<0.00001
maternal characteristics-based SVM model	0.63 (0.57 – 0.73)	0.38	0.80	<0.001
CL-based LR model	0.70 (0.64 – 0.76)	0.37	0.94	<0.001
fFN-based LR model	0.54 (0.47 – 0.62)	0.27	0.91	<0.001
EIS-based LR model	0.81 (0.74 – 0.86)	0.85	0.68	<0.00001
PTB model	Testing AUC (95% CI)	Sensitivity	Specificity	p-value
co-model1	0.89 (0.80 – 0.97)	0.60	0.90	<0.001
co-model2	0.87 (0.81– 0.93)	0.64	0.89	<0.001
maternal characteristics-based SVM model	0.73 (0.67 – 0.78)	0.58	0.84	<0.01
CL-based LR model	0.83 (0.71 – 0.95)	0.50	0.89	<0.01
fFN-based LR model	0.81 (0.67 – 0.94)	0.45	0.93	<0.001
EIS-based LR model	0.85 (0.77 – 0.93)	0.70	0.87	<0.0001

Table 5.14: Performance comparison between the proposed combined models in this work and existing models. EIS: electrical impedance spectroscopy, TVS-CL: transvaginal sonography cervical length, fFN: fetal fibronectin, AUC: area under the receiver operating characteristic curve.

Reference	Feature	Model	Testing AUC (95% CI)
This Work	maternal characteristics and EIS	co-model1	0.89 (0.80 – 0.97)
		co-model2	0.87 (0.81 – 0.93)
Anumba, D et al. [86]	EIS	logistic regression	0.76 (0.71 – 0.81)
Anumba, D et al. [86]	TVS-CL	logistic regression	0.72 (0.66 – 0.76)
Anumba, D et al. [86]	fFN	logistic regression	0.62 (0.56 – 0.72)
Anumba, D et al. [86]	EIS, TVS-CL and fFN	logistic regression	0.79 (0.74 – 0.83)
Esplin, M Sean et al. [20]	TVS-CL	threshold	0.67 (0.64 - 0.70)
Esplin, M Sean et al. [20]	fFN	threshold	0.59 (0.56 - 0.62)
Esplin, M Sean et al. [20]	TVS-CL and fFN	logistic regression	0.67 (0.64 - 0.70)

5.5 Conclusion

This chapter has applied a models-combining approach to achieve the combination of a cervical EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction model. Cervical EIS data were measured from 438 women at 19 to 23 weeks' gestation and their maternal characteristics data included demographics and obstetric data. Also, CL and fFN data of the 438 women were also used as a comparison for PTB risk prediction. Based on three machining learning pipelines of training the maternal characteristics-based PTB risk prediction model, the CL/fFN-based PTB risk prediction model, and the EIS-based PTB risk prediction model, 100 maternal characteristics-based PTB risk prediction models, 100 CL/fFN-based PTB

risk prediction models, and 100 EIS-based PTB risk prediction models were obtained on 100 training sets. Then, in terms of the models-combining approach, 100 SVM-based combined models and 100 LR-based combined models were also obtained by combining two probabilities of a maternal characteristics-based PTB risk prediction model and an EIS-based PTB risk prediction model.

The best SVM-based combined model (co-model1) had the highest testing AUC of 0.89 (95% CI: 0.80 – 0.97, sensitivity: 0.60, specificity: 0.90), and the best LR-based combined model (co-model2) had the highest testing AUC of 0.87 (95% CI: 0.81 – 0.93, sensitivity: 0.64, specificity: 0.89). These two combined models had an improvement on PTB prediction comparing to PTB prediction models only based on maternal characteristics, CL, fFN and EIS, and outperformed than the state-of-the-art models [20, 86].

Chapter 6

6. A Feasibility Study on the MIS-based PTB Risk Prediction

In the present work, a feasibility study on the application of the magnetic impedance spectroscopy technique to PTB risk prediction is conducted. The MIS data were collected from 84 women using two different MIS devices. In this group of women, 44 were AHR women at 18 to 28 weeks of pregnancy and 40 were SYMP women at 18 to 35 weeks of pregnancy. The objectives were to train and apply a MIS data-based machine learning model for PTB prediction and to study the effects of different MIS data pre-processing approaches on the performance of the PTB prediction. The results have demonstrated the feasibility of the MIS-based PTB prediction and showed that an appropriate approach was needed to pre-process MIS data so that a more effective machine learning model can be built and used for PTB risk prediction.

Before application of machine leaning to MIS data to build a machine leaning model for PTB risk prediction, the directly collected MIS data (named as real-time MIS data) were pre-processed to calibrate the MIS measurements with respect to air and ferrite references. This calibration was conducted by subtracting air and ferrite references from measured MIS spectrum. In this study, a new calibration was introduced, which divided the MIS spectrum by air and ferrite references and uses the calibrated MIS spectrum for PTB risk prediction. The two calibrations were called “calibration via subtraction” and “calibration via division”, respectively. Then, a further method was proposed, which built an ARX model using raw MIS data and only used model features representing cervical magnetic impedance rather than MIS device circuit properties for PTB prediction. This further calibration method was called “calibration via modelling”. After the MIS data calibration, a SVM based machine learning classification model was trained on 66% of the 84 women. Then the trained classification

model was applied to the remaining 34% of the 84 women to evaluate its performance on PTB risk prediction. Six kinds of SVM classification models were trained and evaluated where MIS data calibrated using the three different calibration methods were used, respectively, to perform the feasibility study on MIS-based PTB risk prediction.

6.1 Information of 84 Recruited Women

The ECCLIPPx II project recruited a total of 84 women (Table 6.1) visiting between 18 to 35 weeks of their gestation, with an average visit of 25 weeks of gestation, 59 of them were measured by Probe 2a and 25 of them were measured by Probe 2c. Of these, 44 of 84 women were at an AHR of premature birth (52.4%) and 40 of women were SYMP of preterm birth, besides, 63 of women delivered on term (75%) and 21 of women delivered prematurely (25%). 9.53% of PTB women gave births within 28 completed weeks of gestation, 23.81% gave birth between 28 and 32 weeks of gestation, 33.33% delivered between 32 and 34 weeks and 33.33% delivered between 34 and 37 weeks.

Among the 44 women (Table 6.2) who were at an AHR of premature birth and all of whom visited twice to take measurements, where the first visit was during 18 to 22 weeks of gestation (average: 20 weeks) and the second visit was between 23 and 28 weeks (average: 26 weeks). Of these, 77.3% AHR women (34 out of 44) gave births on term and 22.3% AHR women (10 out of 44) gave births prematurely. 20% of PTB women were very preterm birth, 50% of PTB women were moderate preterm birth and 30% of them were late preterm birth. Besides, they gave birth prematurely at between 11 and 18 weeks after their first visit

Among the 40 women (Table 6.3) who were symptomatic of premature birth and all of whom had one visit only to take measurements, where their visits were during 18 to 35 weeks of gestation (average: 30 weeks). Of these, 72.5% SYMP women (29 out of 40) gave births on term and 27.5% SYMP women (11 out of 40) gave births prematurely. 18.2% of PTB women were extreme preterm birth (<28 completed weeks of gestation), 27.3% of PTB women were very preterm birth (between 28 and 32 weeks of gestation), 18.2% of PTB women were moderate preterm birth (32 to 43 weeks of gestation) and 36.3% of them were late preterm birth

(34 to 37 weeks of gestation). Additionally, all of them gave birth prematurely at between 1 and 12 weeks after their visit (average: 4 weeks).

Table 6.1: The statistics of 84 recruited women.

Content	Description
Recruits:	84 women
Gestation of visit:	18 to 35 weeks of gestation (average: 25 weeks)
Measured by Probe 2a:	59 of 84 women
Measured by Probe 2c:	25 of 84 women
Asymptomatic High Risk (AHR):	44 of 84 women (52.4%)
Symptomatic (SYMP):	40 of 84 women (47.6%)
On-term birth:	63 women (75%)
Preterm birth:	21 women (25%)
Extreme PTB (<28 weeks):	2 of 21 PTB women (9.53%)
Very PTB (28 to 32 weeks):	5 of 21 PTB women (23.81%)
Moderate PTB (32 to 34 weeks):	7 of 21 PTB women (33.33%)
Late PTB (34 to 37 weeks):	7 of 21 PTB women (33.33%)

Table 6.2: The statistics of 44 Asymptomatic High Risk women.

Content	Description
Asymptomatic High Risk:	44 women
Visit times:	2 times of visit
Gestation of first visit:	18 to 22 weeks of gestation (average: 20 weeks)
Gestation of second visit:	23 to 28 weeks of gestation (average: 26 weeks)
On-term birth:	34 women (77.3%)
Preterm birth:	10 women (22.3%)
Extreme PTB (<28 weeks):	0 of 10 PTB women (0%)
Very PTB (28 to 32 weeks):	2 of 10 PTB women (20%)
Moderate PTB (32 to 34 weeks):	5 of 10 PTB women (50%)
Late PTB (34 to 37 weeks):	3 of 10 PTB women (30%)
PTB after first visit:	11 to 18 weeks (average: 14 weeks)
PTB after second visit:	4 to 11 weeks (average: 8 weeks)

Table 6.3: The statistics of 40 Symptomatic women.

Content	Description
Symptomatic:	40 women
Visit times:	1 time of visit
Gestation of visit:	18 to 35 weeks of gestation (average: 30 weeks)
On-term birth:	29 women (72.5%)
Preterm birth:	11 women (27.5%)
Extreme PTB (<28 weeks):	2 of 11 PTB women (18.2%)
Very PTB (28 to 32 weeks):	3 of 11 PTB women (27.3%)
Moderate PTB (32 to 34 weeks):	2 of 11 PTB women (18.2%)
Late PTB (34 to 37 weeks):	4 of 11 PTB women (36.3%)
PTB after visit:	1 to 12 weeks (average: 4 weeks)

6.2 Measurement and Record of MIS

As introduced in Chapter 3, the MIS data were measured by MIS2 device using Probe 2a or Probe 2c, over 16 discrete frequencies from 21 kHz to 1013 kHz (15 discrete frequencies were used in this work since the highest frequency (1013 kHz) cycled through once). During the measurement process, the AC served as the input to the system, and the induced voltage, which was affected by the cervix tissue, served as the output. In general, the amplitude of AC decreased with an increase in frequency to keep the induced voltage within a range that was safe for pregnant women. Each frequency lasted for 12 milliseconds, as a result, one frame consisting of 16 frequencies took a total of 0.192 seconds to complete. 480,000 data were captured at each frequency with the sampling frequency of 40 MHz, so that a total of 7,680,000 sample points were captured over 16 frequencies at time domain. Once the footswitch was pressed, MIS2 device kept taking measurements of what the probe closed to. Basically, one measurement consisted of seven events for Ferrite Target, Air, Cervix 1, Vaginal Reference 1, Cervix 2, Vaginal Reference 2 and Cervix 3 respectively, and was recorded as two types of data, one of that was the raw MIS data at time domain, and the other one of that was the Real-Time MIS data at frequency domain.

Raw MIS Data: The raw MIS data consisted of seven events of Ferrite Target, Air, Cervix 1, Vaginal Reference 1, Cervix 2, Vaginal Reference 2 and Cervix 3, which were saved independently. Each event consisted of alternating currents as input and induced voltages as output, where 7,680,000 sample points were recorded over 0.192 seconds. In this work, MIS data of cervix 1, cervix 2 and cervix 3 were used for modelling and PTB risk prediction. Figure 6.1 illustrated two women's raw MIS data measured by Probe 2a and measured by Probe 2c respectively. From Figure 6.1, the transient states between two different frequencies can be obviously observed. Because of the differences between two outputs, the modelling should be implemented respectively, for raw data measured by Probe 2a and Probe 2c.

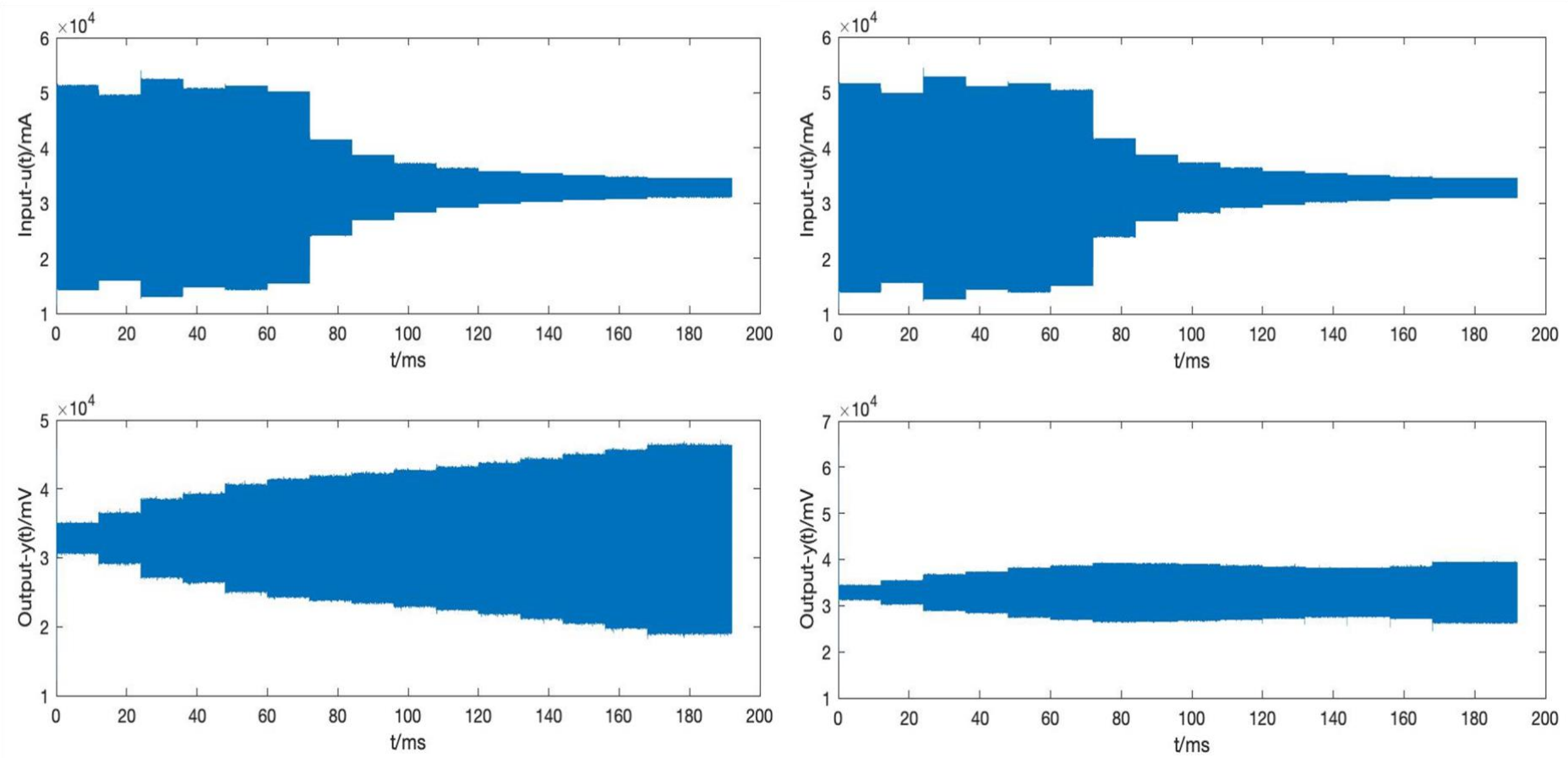

Figure 6.1: The plots of a woman's raw data measured by Probe 2a (left) and a woman's raw data measured by Probe 2c (right). Input was alternating currents and output was induced voltages.

Real-Time MIS Data: Real-Time MIS data were continuously converted from raw MIS data at time domain by FPGA once the footswitch was pressed, and recorded in frames during the whole measurement process. The whole measurement for a woman needed 2000 to 5000 frames. Since one frame consisted 16 frequencies, there were 16 complex numbers to represent real part and imaginary part. Therefore, Real-Time MIS data were saved respectively at 16 frequencies over frames. Figure 6.2 illustrated a woman's imaginary part of Real-Time MIS data at 1013 kHz, where the whole measurement at 1013 kHz was presented and seven events were marked.

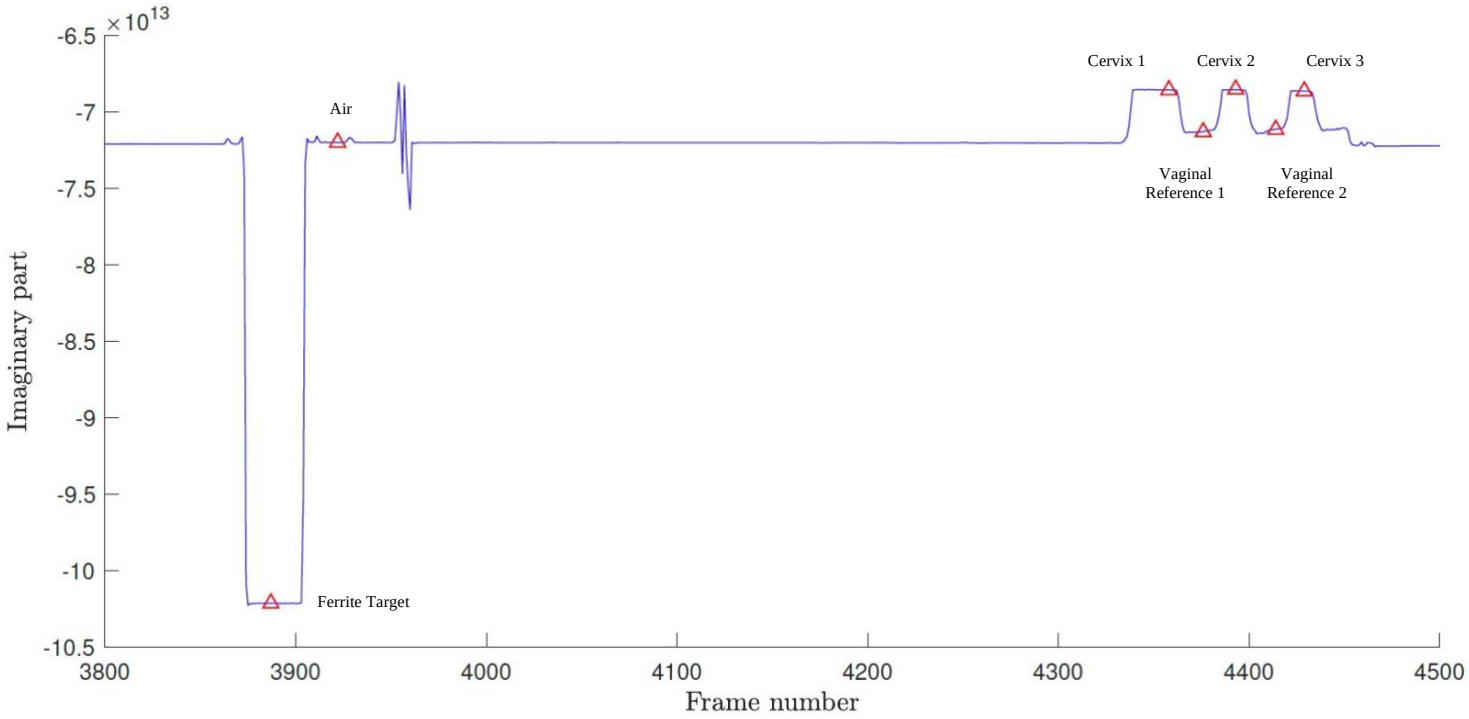

Figure 6.2: A woman's imaginary part of Real-Time MIS data at 1013 kHz. The marks were seven events of Ferrite Target, Air, Cervix 1, Vaginal Reference 1, Cervix 2, Vaginal Reference 2 and Cervix 3, from left to right.

Cervical MIS Data Structure: The 44 AHR women visited twice to take measurements and the 40 SYMP women had only one visit to take a measurement, and for each measurement, cervix 1, cervix 2 and cervix 3 were recorded, therefore, an AHR woman had six records of cervix and a SYMP woman had three records of cervix. Table 6.4 and Table 6.5 present an AHR woman's data structure and a SYMP woman's data structure respectively.

Table 6.4: The illustration of an AHR woman's data structure. AHR: asymptomatic high risk, RT-data: Real-Time MIS data.

An AHR woman	Visit 1	Cervix 1	Raw-data form RT-data form
		Cervix 2	Raw-data form RT-data form
		Cervix 3	Raw-data form RT-data form
	Visit 2	Cervix 1	Raw-data form RT-data form
		Cervix 2	Raw-data form RT-data form
		Cervix 3	Raw-data form RT-data form

Table 6.5: The illustration of a SYMP woman's data structure. SYMP: symptomatic, RT-data: Real-Time MIS data.

A SYMP woman	SYMP Visit	Cervix 1	Raw-data form RT-data form
		Cervix 2	Raw-data form RT-data form
		Cervix 3	Raw-data form RT-data form

6.3 Methodology

6.3.1 Calibration on Real-Time MIS Data

The ECCLIPPx II group advised that real-time MIS data need to carry out calibration before using. The real part of the transimpedance from the drive current to the grad voltage should be proportional to the conductivity of the target [97]. The existing method for generating a number proportional to target conductivity from measurements required obtaining measurements of both the background and the ferrite target. As this was considered to be a measure of all zero conductivity errors, the grad voltage and related current were measured with the probe retained in air and then removed from all subsequent readings. Also, the phase of the transimpedance must be corrected relative to a zero-phase reference using a ferrite target that produced a known phase shift (-90^{o}). Therefore, the ECCLIPPx II group recommended a method of calibration via subtraction, and in this work, a novel method was proposed as calibration via division, both of that were described as follows.

A. Calibration via subtraction

Firstly, the target measurement of cervix was calibrated with air reference by subtracting the air reference,

$$Target\_to\_Air = \frac{gradV}{currentsenseV}\Big|_{Target} - \frac{gradV}{currentsenseV}\Big|_{Air\ Ref} \tag{6.1}$$

where $\frac{gradV}{currentsenseV}\Big|_{Target}$ is the target measurement of cervix (complex number), and $\frac{gradV}{currentsenseV}\Big|_{Air\ Ref}$ is the air reference (complex number). Then, the ferrite reference was calibrated by subtracting air reference (Equation 6.2), and the phase was corrected to the zero-phase by (Equation 6.3),

$$Ferrite\_to\_Air = \frac{gradV}{currentsenseV}\Big|_{Ferrite\ Ref} - \frac{gradV}{currentsenseV}\Big|_{Air\ Ref} \tag{6.2}$$

$$\Delta\theta = -\angle(Ferrite\_to\_Air) - 90° \quad (6.3)$$

where $\frac{gradV}{currentsenseV}\Big|_{Ferrite\ Ref}$ is the ferrite reference (complex number).

Finally, the measurement of cervix calibrated with air reference was corrected by phase adjustment as follows.

$$Phase_{adjust} = cos(\Delta\theta) + j \times \sin(\Delta\theta) \quad (6.4)$$

$$calibraed\_target = Target\_to\_Air \times Phase_{adjust} \quad (6.5)$$

Before application of machine leaning to MIS data to build a machine leaning model for PTB risk prediction, the real-time MIS data over 15 frequencies including cervix 1, cervix 2 and cervix 3 of each woman were calibrated firstly using this calibration method.

B. Calibration via division

Firstly, the target measurement of cervix was calibrated with air reference divided by the air reference,

$$Target\_to\_Air = \frac{gradV}{currentsenseV}\Big|_{Target} \div \frac{gradV}{currentsenseV}\Big|_{Air\ Ref} \quad (6.6)$$

where $\frac{gradV}{currentsenseV}\Big|_{Target}$ is the target measurement of cervix (complex number), and $\frac{gradV}{currentsenseV}\Big|_{Air\ Ref}$ is the air reference (complex number). Then, the ferrite reference was calibrated with air reference, divided by air reference (Equation 6.7), and the phase was then corrected to the zero-phase by (Equation 6.8),

$$Ferrite\_to\_Air = \frac{gradV}{currentsenseV}\Big|_{Ferrite\ Ref} \div \frac{gradV}{currentsenseV}\Big|_{Air\ Ref} \quad (6.7)$$

$$\Delta\theta = -\angle(Ferrite\_to\_Air) - 90° \quad (6.8)$$

where $\frac{gradV}{currentsenseV}\Big|_{Ferrite\ Ref}$ is the ferrite reference (complex number).

Finally, the measurement of cervix calibrated with air reference was corrected by phase adjustment as follows.

$$Phase_{adjust} = cos(\Delta\theta) + j \times \sin(\Delta\theta) \quad (6.9)$$

$$calibraed\_target = Target\_to\_Air \times Phase_{adjust} \quad (6.10)$$

Similarly, the real-time MIS data over 15 frequencies including cervix 1, cervix 2 and cervix 3 of each woman were calibrated using calibration via division.

6.3.2 Calibration on Raw MIS Data

Before application of calibration on raw MIS data, the raw MIS data were used to implement system modelling of an autoregressive with exogenous input model to simulate measurements. The terms of the ARX model were determined by forward regression with orthogonal least-squares (FROLS) algorithm [121], thereafter, the parameters of terms were determined by least-squares. Once the system modelling of the ARX model has been completed, stability analysis with pole-zero identification was carried out to remove the unstable component of the ARX model, meanwhile, the calibration via modelling was also implemented using pole-zero analysis by removing components which do not belong to cervix. The follows described the process of system modelling firstly, including ARX model, FROLS algorithm, determination of parameters with least-squares and stability analysis with pole-zero identification, and then described calibration via modelling with pole-zero analysis.

A. System Modelling of Autoregressive with Exogenous Input Model

1) Autoregressive with Exogenous Input Model: The equation of ARX model is presented [121] as,

$$A(z^{-1})y(k) = B(z^{-1})u(k - nk) + \varepsilon(k) \quad (6.11)$$

where $y(k)$ and $u(k\ -\ nk)$ refer to the output and input signals of ARX model, $\varepsilon(k)$ is the noise sequence, and nk is the time delay between output and input signals. Furthermore, $A(z^{-1})$ and $B(z^{-1})$ are given by,

$$A\ (z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_{na} z^{-na} \tag{6.12}$$

$$B\ (z^{-1}) = b_1 + b_2 z^{-1} + b_3 z^{-2} + \cdots + b_{nb+1} z^{-nb} \tag{6.13}$$

where z^{-1} is the backward shift operator, na is the time delay of output signals, and nb is the time delay of input signals. Let substitute $A\ (z^{-1})$ and $B\ (z^{-1})$ into Equation 6.11, obtain Equation 6.14 below. By setting $z = e^{j\omega T}$, where T is the sampling time, the time-domain model (6.11) can be converted into frequency-domain model.

$$y(\ k) + a_1 y(\ k-1) + \cdots + a_{na} y(k-na) = b_1 u(\ k-nk) + \cdots + b_{nb+1} u(k-nb-nk) + \varepsilon(k) \tag{6.14}$$

2) Forward Regression with Orthogonal Least-Squares Algorithm: The FROLS algorithm is identical to the standard OLS algorithm [121], with the exception that, at each stage, a full search of all remaining or unselected model terms is conducted to identify the best candidate term, regardless of the order in which the model terms are added [122]. More details about FROLS algorithm were shown as follows.

Let support the maximum output time delay na was m, the maximum input time delay nb was n, and the time delay between output and input nk is 0. The noise sequence $\varepsilon(k)$ was not taken into consideration, then Equation 6.14 was transformed to,

$$y(\ k) = a_1 y(\ k-1) + \cdots + a_m y(k-m) + b_1 u(\ k) + \cdots + b_n u(k-n) \tag{6.15}$$

In this case, $\{y(\ k-1), \dots, y(k-m), u(\ k), \dots, u(k-n)\}$ were candidate terms to be determined by FROLS algorithm. Suppose $y = y(\ k)$, and $[y(\ k-1), \dots, y(k-m), u(\ k), \dots, u(k-n)] = [p_1, p_2, \dots, p_M] = P$, where $M = m + n$. The Algorithm 6.1 was described the FROLS algorithm [121].

Algorithm 6.1 Forward Regression with Orthogonal Least-Squares

Input: the candidate terms of $P = [p_1, p_2, \dots, p_M]$, output y and a set-up threshold δ

Output: the matrix of determined terms W

Step 1: 1. **for** all p_i in P do

2. $q_i = \frac{p_i}{\|p_i\|}$, where $\|\cdot\|$ is the Euclidean length

3. $g_i = \frac{\langle y, q_i \rangle}{\langle q_i, q_i \rangle}$, where g_i is the auxiliary parameter, and $\langle \cdot , \cdot \rangle$ is the inner product of two vectors

4. $ERR[i] = \frac{g_i^2 \langle q_i , q_i \rangle}{\langle y , y \rangle}$, where ERR is the error reduction ratio

5. $l_1 = \arg\max \{ERR[i]\}$

6. **Define** $W[1] = q_{l_1}$, $D[1] = p_{l_1}$, $err[1] = ERR[l_1]$ and remove p_{l_1} from P

7. **end**

Step s ($s \geq 2$): 1. **for** all p_i in P do

2. $q_i = p_i - \sum_{j=1}^{s-1} \frac{\langle p_i, w_j \rangle}{\langle w_i, w_i \rangle} w_i$, which is Gram-Schmidt orthogonalisation

3. $g_i = \frac{\langle y, q_i \rangle}{\langle q_i, q_i \rangle}$

4. $ERR[i] = \frac{g_i^2 \langle q_i , q_i \rangle}{\langle y , y \rangle}$

5. $l_s = \arg\max \{ERR[i]\}$

6. **Define** $W[s] = q_{l_s}$, $D[s] = p_{l_s}$, $err[s] = ERR[l_s]$ and remove p_{l_s} from P

The search will **stop** at the N step when the ESR is less than the set-up threshold δ,

$$ESR = 1 - \sum_{s=1}^{N} err[s] < \delta$$

where ESR is the error-to-signal ratio.

Return the matrix of determined terms W

3) Determination of Parameters with Least-Squares: After the terms were determined in Equation 6.15, the parameters $\Theta = [a_1, \dots, a_m, b_1, \dots, b_n]^T$ would be determined by least-squares [121] which minimises the cost function, and the cost function has been improved to focus on fitting the points of transient states,

$$J(\Theta) = \frac{1}{|C|}\sum_{k\in C}[y(k) - \hat{y}(k)]^2 \tag{6.16}$$

where $y(k)$ is the measured output, $\hat{y}(k) = P(k)\Theta$ is the simulated output, and $P(k) = [y(k-1), \dots, y(k-m), u(k), \dots, u(k-n)]$ and $|C|$ is the cardinality of the set C, which is $C = \{k | y(k-1), \dots, y(k-m), u(k), \dots, u(k-n) \in transient\ states\}$.

4) Stability Analysis with Pole-Zero Identification: After the parameters $\Theta = [a_1, \dots, a_m, b_1, \dots, b_n]^T$ were determined, substitute Θ into Equation 6.12 and 6.13,

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_m z^{-m}$$

$$B(z^{-1}) = b_1 + b_2 z^{-1} + \cdots + b_n z^{-n}$$

then substitute them into Equation 6.11,

$$(1 + a_1 z^{-1} + \cdots + a_m z^{-m}) y(k) = (b_1 + b_2 z^{-1} + \cdots + b_n z^{-n}) u(k)$$

$$y(k) = \frac{(b_1 + b_2 z^{-1} + \cdots + b_n z^{-n})}{(1 + a_1 z^{-1} + \cdots + a_m z^{-m})} u(k)$$

$$H(z) = \frac{b_1 + b_2 z^{-1} + \cdots + b_n z^{-n}}{1 + a_1 z^{-1} + \cdots + a_m z^{-m}}$$

then the zero-pole-gain form of the transfer function was,

$$H(z) = k\frac{(z - zero_1)(z - zero_2) \dots (z - zero_L)}{(z - pole_1)(z - pole_2) \dots (z - pole_L)}$$

where $L = arg\max\{m, n\}$, $zero_i$ and $pole_i$ are the zeros and poles of system. Once all $|pole_i| < 1$, for $i = 1, 2, \dots, L$, where $|\cdot|$ is the length, the system is stable, otherwise the

system is unstable, the unstable component of $|pole_i| > 1$ shall be removed from the system. In this study, the instability issue was dealt with by calibration via modelling.

B. Calibration via Modelling

Theoretically, the built ARX model using raw MIS data consisted of the components of woman's cervix, MIS device circuit properties and other components like air and vagina which should be removed from the cervical magnetic impedance. In this study, I proposed a hypothesis that the built ARX model comprised the components of cervix, MIS device circuit and others, which were able to be presented as independent sub-models and were connected in series as follows. Each of them were presented in a zero-pole-gain form of the transfer function.

$$H(z) = H(z)_{cervix} \times H(z)_{circuit} \times H(z)_{others} \tag{6.17}$$

The pole-zero plot of ARX model $H(z)$ was analysed, and based on this hypothesis and prior knowledge of MIS device circuit and cervix, where the model of MIS device circuit was more like a second-order system and the model of cervix was more like a first-order system, the sub-model of cervix may be identified by pole-zero analysis. Therefore, a model of cervix alone $H(z)_{cervix}$ was obtained by removing other components like $H(z)_{circuit}$ and $H(z)_{others}$, to achieve the calibration of MIS data. Finally, cervical magnetic impedances produced by the ARX model of cervix over 15 frequencies including cervix 1, cervix 2 and cervix 3 of each woman were structured as the calibrated MIS dataset. The cervical magnetic impedance spectrum was computed by,

$$Z\left(e^{j\omega_i T}\right) = \frac{Y(j\omega_i T)}{U(j\omega_i T)} \tag{6.18}$$

where $i = 1, 2, \dots, 15$, and $Y(j\omega_i T)$ and $U(j\omega_i T)$ are input signals and output signals of cervical MIS in steady states at frequency ω_i.

6.3.3 A Machine Learning Pipeline for training MIS-based PTB Risk Prediction Models

After the calibration, Table 6.4 and Table 6.5 were updated to calibrated data structures of an AHR women and a SYMP woman presented in Table 6.6 and Table 6.7. After that, 40 SYMP women were combined with 44 AHR women at their first visit and with 44 AHR women at their second visit, respectively, so that two sets of 84 women were obtained. Each woman's MIS data were calibrated by subtraction, division, and modelling respectively. Therefore, the two sets of 84 women's MIS data were constructed as shown in Table 6.8, and there were six MIS datasets created in total.

Table 6.6: The illustration of an AHR woman's calibrated data structure. AHR: asymptomatic high risk.

An AHR woman	Visit 1	Cervix 1	calibrated via subtraction calibrated via division calibrated via modelling
		Cervix 2	calibrated via subtraction calibrated via division calibrated via modelling
		Cervix 3	calibrated via subtraction calibrated via division calibrated via modelling
	Visit 2	Cervix 1	calibrated via subtraction calibrated via division calibrated via modelling
		Cervix 2	calibrated via subtraction calibrated via division calibrated via modelling
		Cervix 3	calibrated via subtraction calibrated via division calibrated via modelling

Table 6.7: The illustration of an SYMP woman's calibrated data structure. SYMP: symptomatic.

A SYMP woman	SYMP Visit	Cervix 1	calibrated via subtraction calibrated via division calibrated via modelling
		Cervix 2	calibrated via subtraction calibrated via division calibrated via modelling
		Cervix 3	calibrated via subtraction calibrated via division calibrated via modelling

Table 6.8: The constructure of two sets of MIS datasets. AHR: asymptomatic high risk.

MIS dataset	Combination	Calibration Method
MIS dataset1a	40 SYMP women and 44 AHR women at 1st visit	calibrated via subtraction
MIS dataset1b		calibrated via division
MIS dataset1c		calibrated via modelling
MIS dataset2a	40 SYMP women and 44 AHR women at 2nd visit	calibrated via subtraction
MIS dataset2b		calibrated via division
MIS dataset2c		calibrated via modelling

For each MIS dataset, 84 women were randomly split into 100 subsets of 66% women as training sets for training PTB risk prediction models and 100 subsets of 34% women as testing sets for testing models. Each subset of 66% women consisted of 66% PTB women and 66% on-term birth women, similarly, each subset of 34% women also consisted of 34% PTB women and 34% on-term birth women. The following provided further explanation.

84 recruited women at PTB risk were,

- of 21 preterm births (25%),
- of 63 on-term births (75%),
- split into a subset of 56 women including 14 preterm births and 42 on-term births,
- split into a subset of 28 women including 7 preterm births and 21 on-term births.

Since each woman had three records of cervix 1, cervix 2 and cervix 3, in general, the three records would be averaged to obtain a final record. In this study, the three records were regarded as three independent samples during training a machine learning model, which meant that 56 women were extended to 168 women (42 preterm birth women and 126 on-term birth women) (shown in Figure 6.3).

no.	women	real part			imaginary part		
1	woman 1	$p_{1,1}$	…	$p_{1,15}$	$p_{1,16}$	…	$p_{1,30}$
2	woman 2	$p_{2,1}$	…	$p_{2,15}$	$p_{2,16}$	…	$p_{2,30}$
…	…	…		…	…		…
56	woman 56	$p_{56,1}$	…	$p_{56,15}$	$p_{56,16}$	…	$p_{56,30}$

no.	women	real part			imaginary part		
1	woman 1	$f_{1,1}$	…	$f_{1,15}$	$f_{1,16}$	…	$f_{1,30}$
2	woman 1	$f_{2,1}$	…	$f_{2,15}$	$f_{2,16}$	…	$f_{2,30}$
3	woman 1	$f_{3,1}$	…	$f_{3,15}$	$f_{3,16}$	…	$f_{3,30}$
…	…	…		…	…		…
168	woman 56	$f_{168,1}$	…	$f_{168,15}$	$f_{168,16}$	…	$f_{168,30}$

Figure 6.3: The illustration of extending 56 samples to 168 samples on a training set. A woman' MIS data had 15 complex numbers over 15 different frequencies. The upper table consisted of 56 samples of 56 women when used the averaged record, and for example, $[p_{1,1} \ldots p_{1,30}]$ was the averaged record from the first woman. The bottom table consisted of 168 samples of 56 women when used three records of cervix 1, cervix 2 and cervix 3. For example, $[f_{1,1} \ldots [f_{1,30}]$, $[f_{2,1} \ldots [f_{2,30}]$ and $[f_{3,1} \ldots [f_{3,30}]$ were the three records from the first woman.

Figure 6.4 provided a visual representation of the machine learning pipeline for training a MIS-based PTB risk prediction model. First, each training set was oversampled using the Borderline-SMOTE1 technique [112] (introduced in Chapter 4) which produced random new synthetics of the PTB data and added them into the training set to balance the classes ‘0’ and ‘1’. The balanced training set was used directly to train a machine learning model, where a SVM was used, its kernel function was set to ‘rbf’, and the remaining parameters of 'BoxConstraint' and 'KernelScale' were optimised using Bayesian Optimization with a custom function for computing a 3-fold cross validation AUC of SVM.

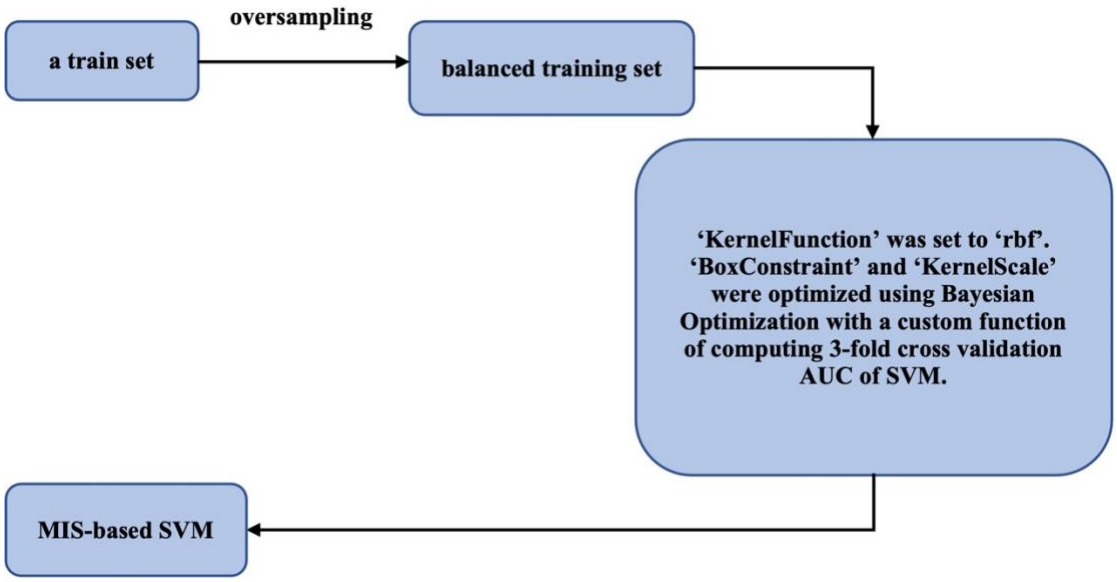

Figure 6.4: A machine learning pipeline for training MIS-based PTB risk prediction models. MIS: magnetic impedance spectroscopy, AUC: AUC: area under the receiver operating characteristic curve, SVM: support vector machine.

6.4 Results of Calibration

6.4.1 Results of Calibration on Real-Time MIS Data

The achievements of calibration on real-time MIS data were from two approaches, one of that was calibration via subtraction recommended by the ECCLIPPx II group, computed by Equation 6.1 to Equation 6.5, and the other was calibration via division proposed in this study, computed by Equation 6.6 to Equation 6.10. The comparison of original real time MIS data (no calibration), MIS data calibrated via subtraction and MIS data calibrated via division was illustrated in Figure 6.5.

Comparing the set of amplitudes of the real time MIS data, the amplitudes of MIS data calibrated via subtraction and MIS data calibrated via division almost lose the differences over 15 frequencies, so that the values are very similar over 15 frequencies, while comparing the set of phases of the real time MIS data, the phases of MIS data calibrated via subtraction and MIS data calibrated via division remain the differences over 15 frequencies. The phases of MIS data calibrated via division change slightly from the phases of original MIS data, and the phases of MIS data calibrated via subtraction have greater values than the phases of original MIS data, but its changing trend over lower frequencies is corresponding to that of the phases of original MIS data.

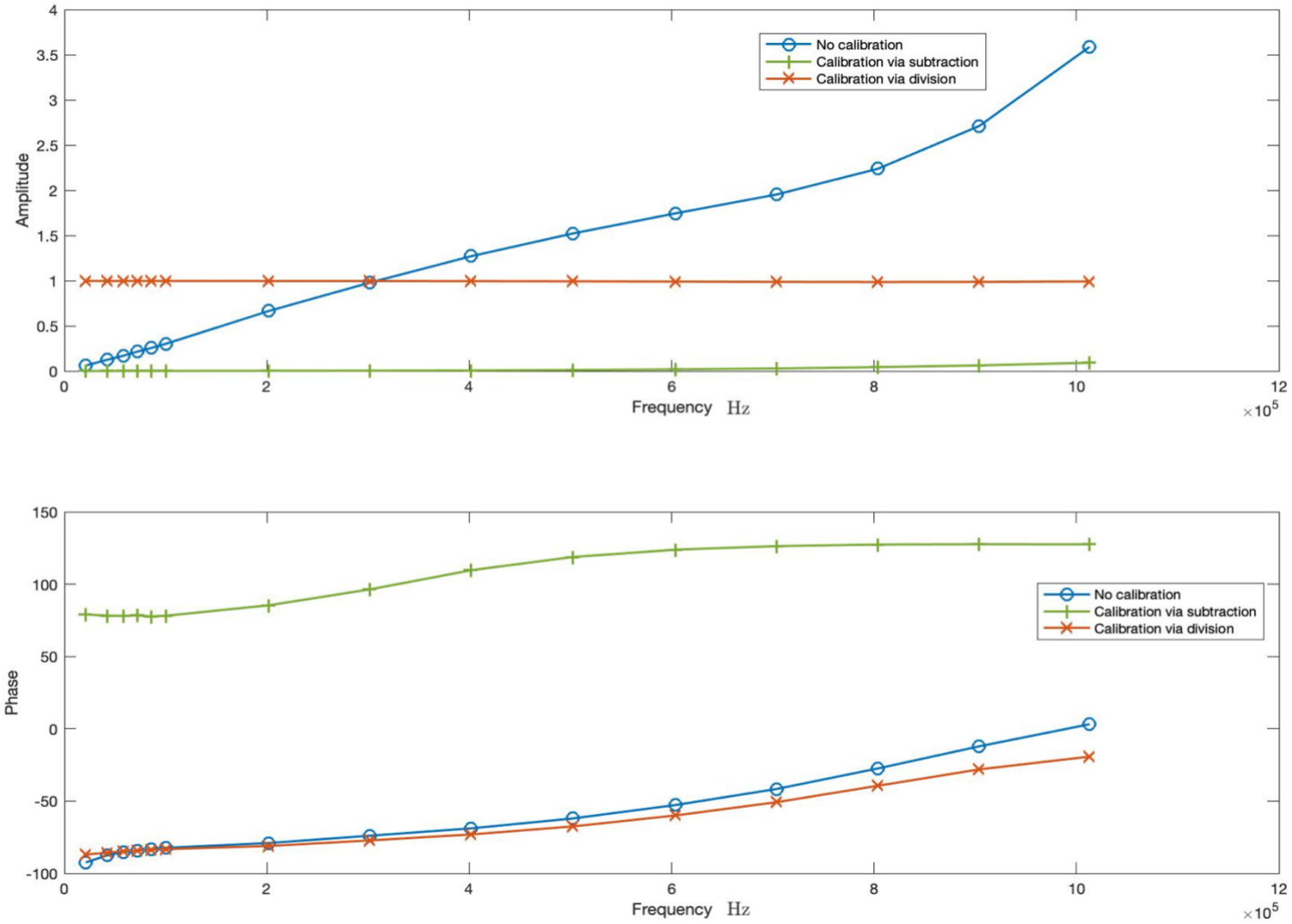

Figure 6.5: The comparison of original real time MIS data (no calibration), MIS data calibrated via subtraction and MIS data calibrated via division, from a woman randomly selected from 84 women. Results presented amplitudes (ohm) and phases (degree) of MIS spectra over 15 different frequencies ranging from 21 kHz to 1013 kHz.

6.4.2 Results of Calibration on Raw MIS Data

The results of system modelling using raw MIS data collected by Probe 2a and Probe 2c were presented respectively, including input and output signals, the ARX models built by system identification, the adjusted ARX models, and their simulating performances. Thereafter,

calibration via modelling was implemented on the adjusted ARX models and results were illustrated. All results were obtained from the record of Cervix 1 from two women, one of whom was randomly chosen from women measured by Probe 2a and another of whom was randomly chosen from women measured by Probe 2c. The results obtained from other records were similar to the corresponding results proposed as follows.

A. System Modelling for ARX model

The raw MIS data were collected under the sampling frequency of 400 MHz, whose high sampling frequency can lead to an issue of linear dependent neighbour terms. These neighbour terms may be rejected while carrying out FROLS algorithm to select terms for system, so that high-order models were more possible to be determined, however, these models did not adapt to the cervical MIS device circuit. Hence, the raw MIS data were down sampled to five MHz firstly, and maximum time delays of input and output signals were both set to five. Then, the down-sampled and zero-mean raw MIS data were used to build an ARX model by FROLS algorithm and least-squares method.

With FROLS algorithm, the ARX model structure was determined as follows, for raw MIS data of 84 women measured by either Probe 2a or Probe 2c,

$$y(k) = \theta_1 y(k-1) + \theta_2 y(k-2) + \theta_3 y(k-5) + \theta_4 u(k) + \theta_5 u(k-1) + \theta_6 u(k-5) \tag{6.19}$$

where $\Theta = [\theta_1, \dots, \theta_3, \dots, \theta_6]^T$ are the parameters of the ARX model. Each set of parameters were determined by the raw MIS data of corresponding women using least-squares method which minimised the cost function of Equation 6.16, where $\Theta = [\theta_1, \dots, \theta_3, \dots, \theta_6]^T$ and $P(k) = [y(k-1), y(k-2), y(k-5), u(k), u(k-1), u(k-5)]$. After determining parameters, the fitting performance of each ARX model can be evaluated in time domain and in frequency domain. In time domain, the normalized root mean square error (NRMSE) between measured outputs and simulated outputs was employed to validate the fitting performance and computed by Equation 6.20, which was one of validation methods.

$$NRMSE = \sqrt{\frac{sum((y(k)-\hat{y}(k))^2)}{sum((y(k)-mean(y))^2)}} \tag{6.20}$$

In frequency domain, the cervical impedance spectra (frequency response) of the ARX model of each woman over 15 frequencies were firstly computed by Equation 6.18, then which were compared by the cervical impedance spectra of corresponding women obtained by FPGA. Also, the NRMSEs of amplitude and phase between them (Equation 6.21 and Equation 6.21) were used to validate the fitting performance of the ARX model in frequency domain. More details about the fitting performance and steady analysis were shown as follows.

$$NRMSE = \sqrt{\frac{sum((Amp(k) - \widehat{Amp}(k))^2)}{sum((Amp(k) - mean(Amp))^2)}}, k = 1, \dots, 15 \quad (6.21)$$

$$NRMSE = \sqrt{\frac{sum((Pha(k) - \widehat{Pha}(k))^2)}{sum((Pha(k) - mean(Pha))^2)}}, k = 1, \dots, 15 \quad (6.22)$$

B. ARX model built using Raw MIS Data collected by Probe 2a

The fitting performance of an ARX model in time domain, the model structure of that was shown as Equation 6.19 and parameters of that were identified by least-squares method, was illustrated in Figure 6.6, which compared measured outputs to simulated outputs. The ARX model was built using a woman's raw MIS data collected by Probe 2a and its NRMSE computed by Equation 6.20 was 0.04. For the same woman, the fitting performance of the ARX model in frequency domain was illustrated in Figure 6.7, which compared amplitudes and phases obtained by ARX model and FPGA over 15 frequencies, respectively. The NRMSEs of amplitude and phase computed by Equation 6.21 and 6.22 were 0.0067 and 0.0824 respectively.

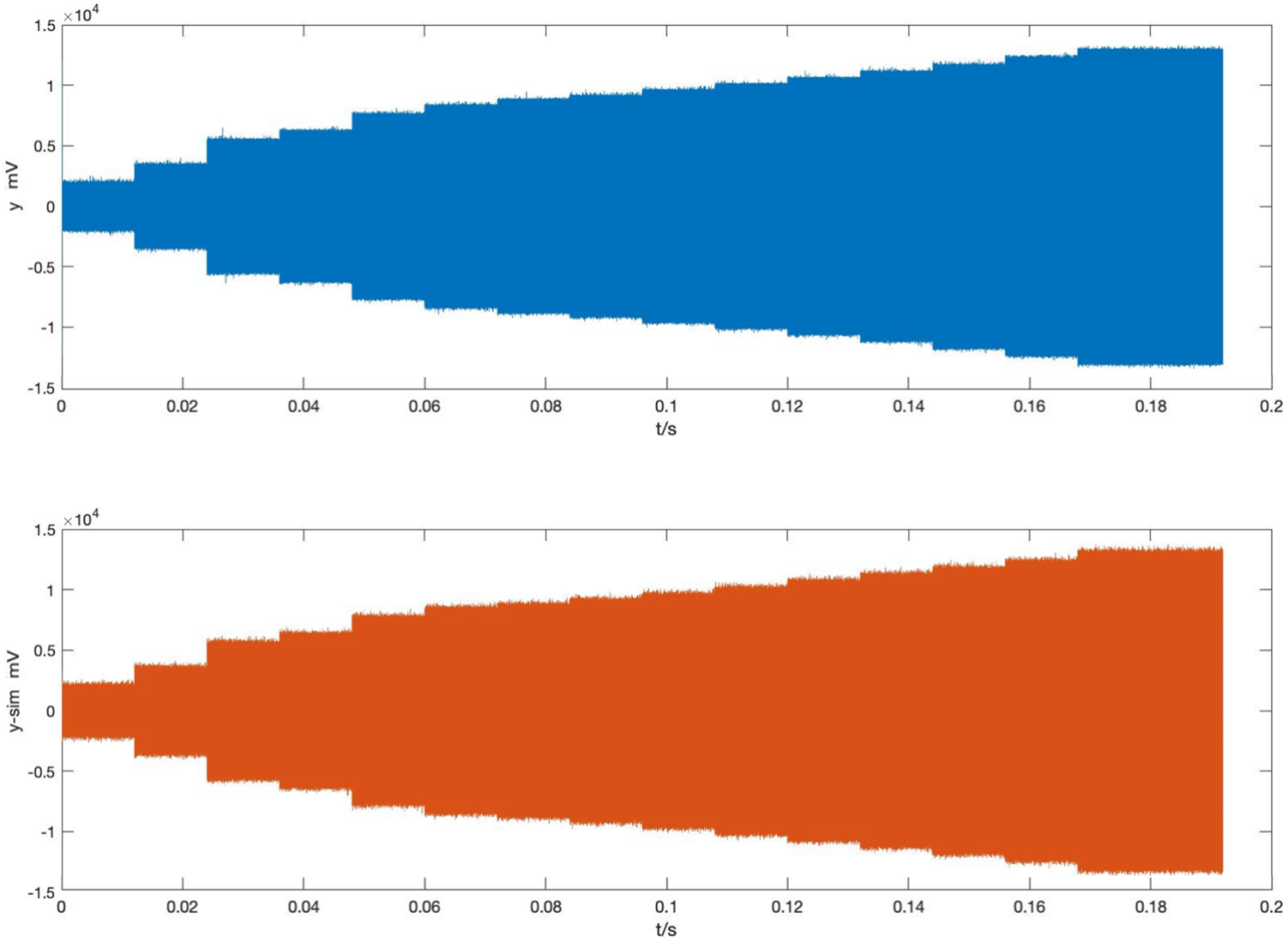

Figure 6.6: The time-domain fitting performance of the ARX model built using Raw MIS Data collected by Probe 2a. Results presented between zero-mean measured voltages (blue) and zero-mean simulated voltages (orange) over 0.192 seconds.

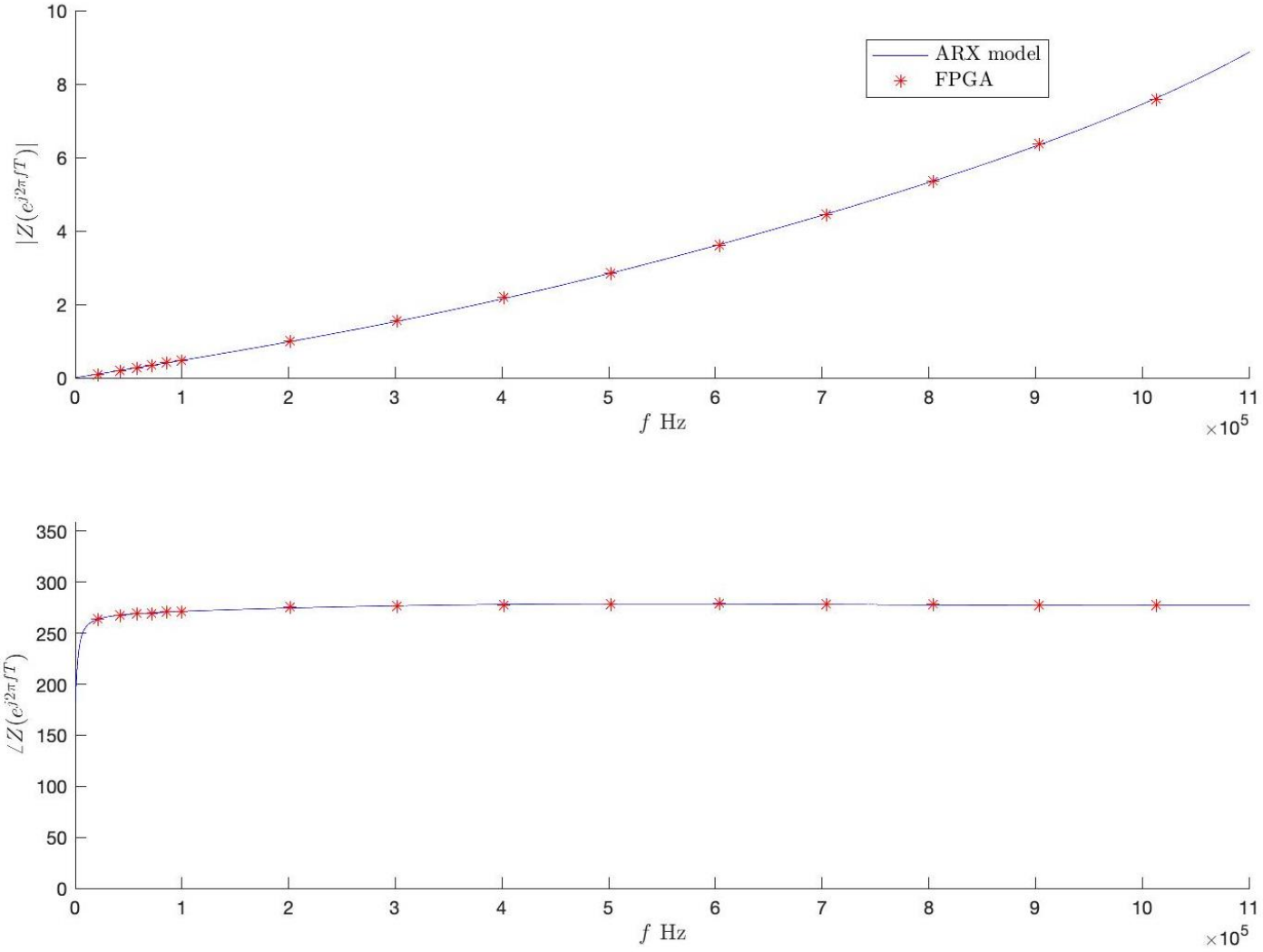

Figure 6.7: The frequency-domain fitting performance of the ARX model built using Raw MIS Data collected by Probe 2a. Results presented amplitudes (upper plot (ohm)) and phases (bottom plot (degree)) obtained by the ARX model and FPGA respectively, over 15 frequencies ranging from 21 kHz to 1013 kHz.

C. ARX model built using Raw MIS Data collected by Probe 2c

The fitting performances of ARX model in time domain, the model structure of that was used as shown by Equation 6.19 and parameters of that were identified by least-squares method, was illustrated in Figure 6.8, which compared measured outputs to simulated outputs. The ARX model was built using a woman's raw MIS data collected by Probe 2c. The ARX model built

using raw MIS Data collected by Probe 2c was unstable in time domain, however, for the same woman, the ARX model was fitted well in frequency domain and its performance was illustrated in Figure 6.9, which compared amplitudes and phases obtained by ARX model and FPGA over 15 frequencies, respectively, and its NRMSEs of amplitude and phase were 0.1416 and 0.1247 respectively.

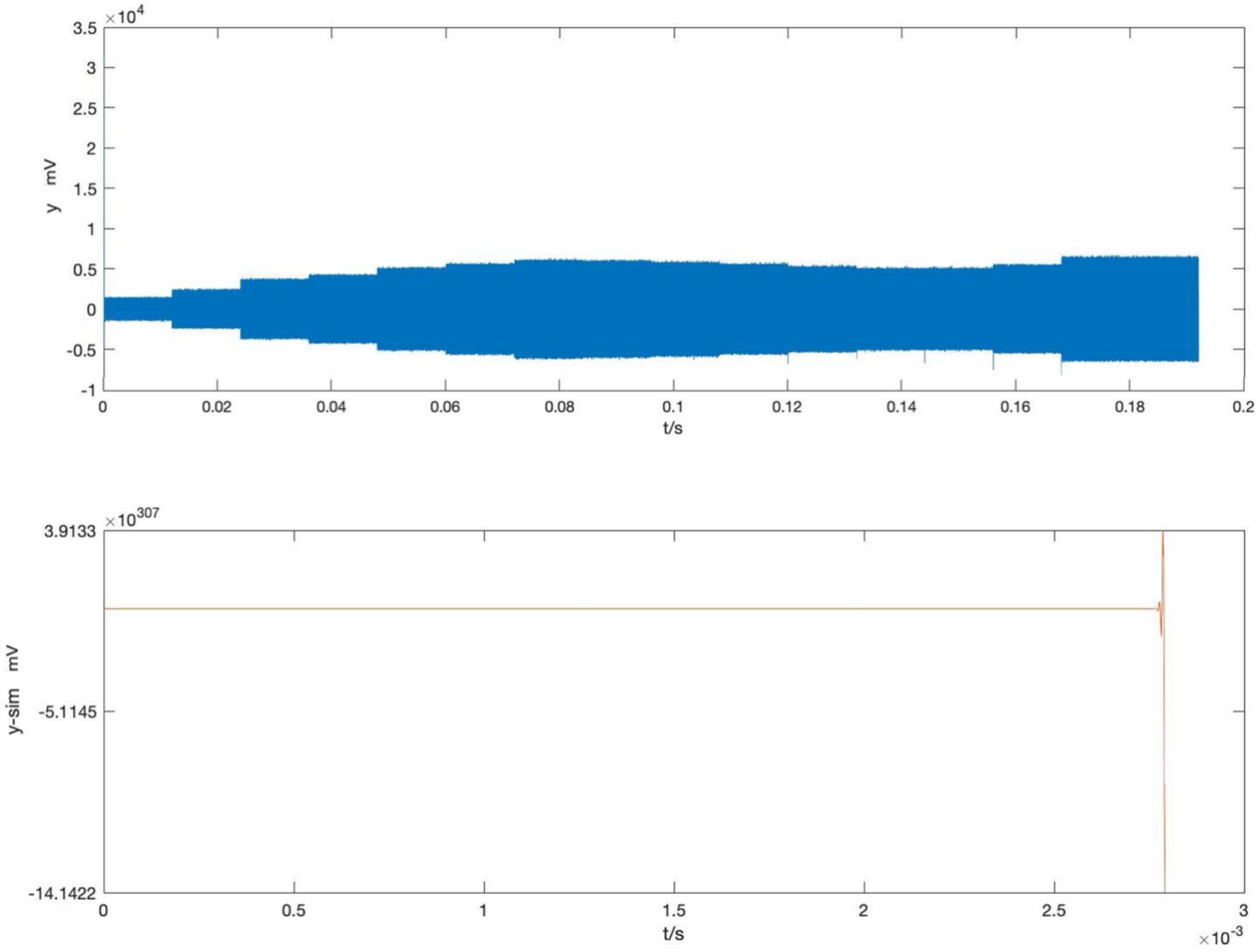

Figure 6.8: The time-domain fitting performance of the ARX model built using raw MIS Data collected by Probe 2c. Results presented between zero-mean measured voltages (blue) and simulated voltages (orange) over 0.192 seconds. The ARX model was unstable.

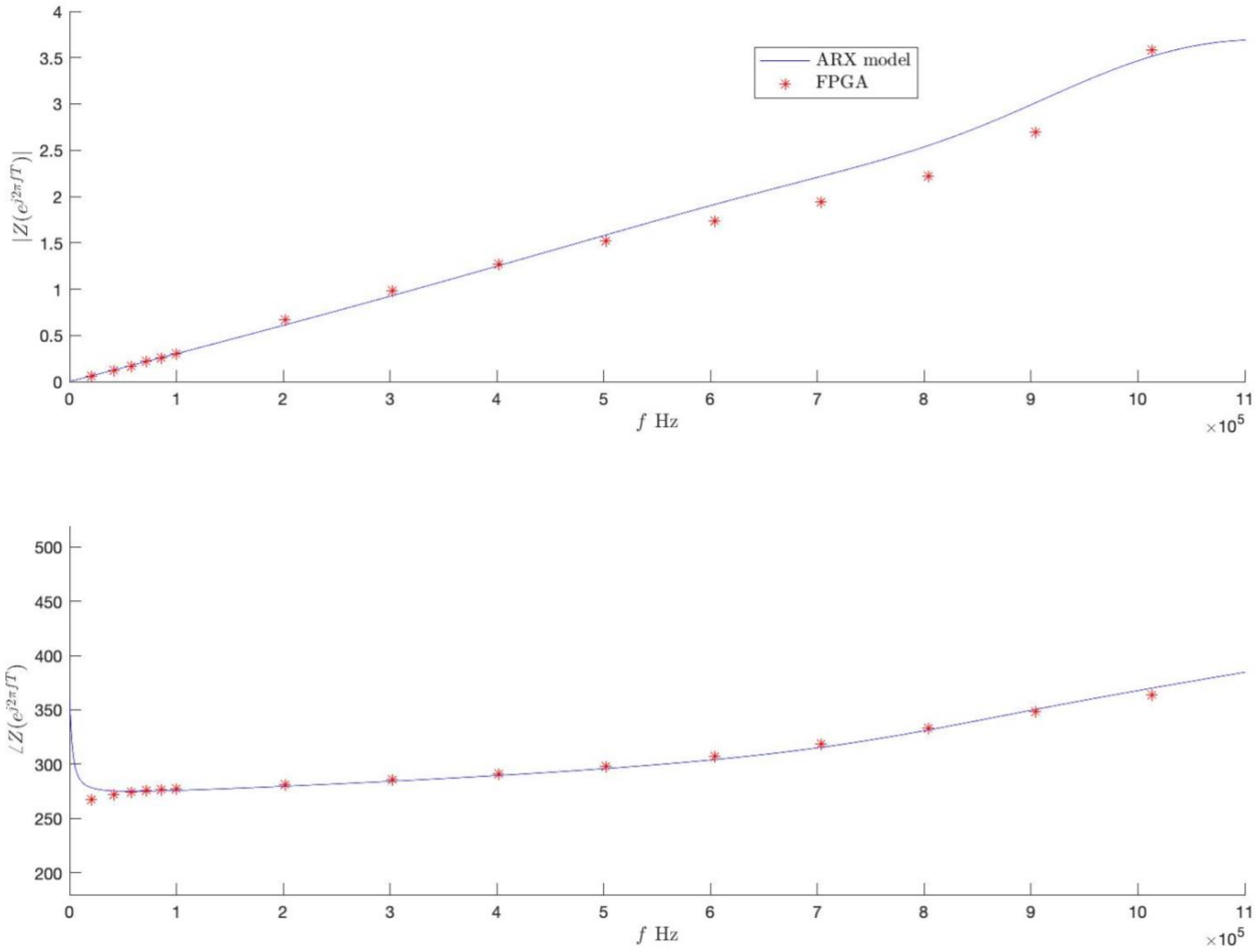

Figure 6.9: The frequency-domain fitting performance of the ARX model built using Raw MIS Data collected by Probe 2c. Results presented amplitudes (upper plot (ohm)) and phases (bottom plot (degree)) obtained by the ARX model and FPGA respectively, over 15 frequencies ranging from 21 kHz to 1013 kHz.

The pole-zero plot was present in Figure 6.10, where there were five poles and five zeros of the ARX model, of which, a complex-conjugate pole pair was out of the unit circle. The instability of the ARX model may be caused by the design of cervical MIS device circuit of Probe 2c. In order to gather cervical MIS data measured by Probe 2a and Probe 2c together training a PTB risk prediction model, the key was to extract the same properties of cervical MIS data measured by Probe 2a and Probe 2c. In this case, the proposed calibration method

via modelling was used to correct the cervical MIS data, meanwhile the calibration via modelling can also deal with the issue of unstable system.

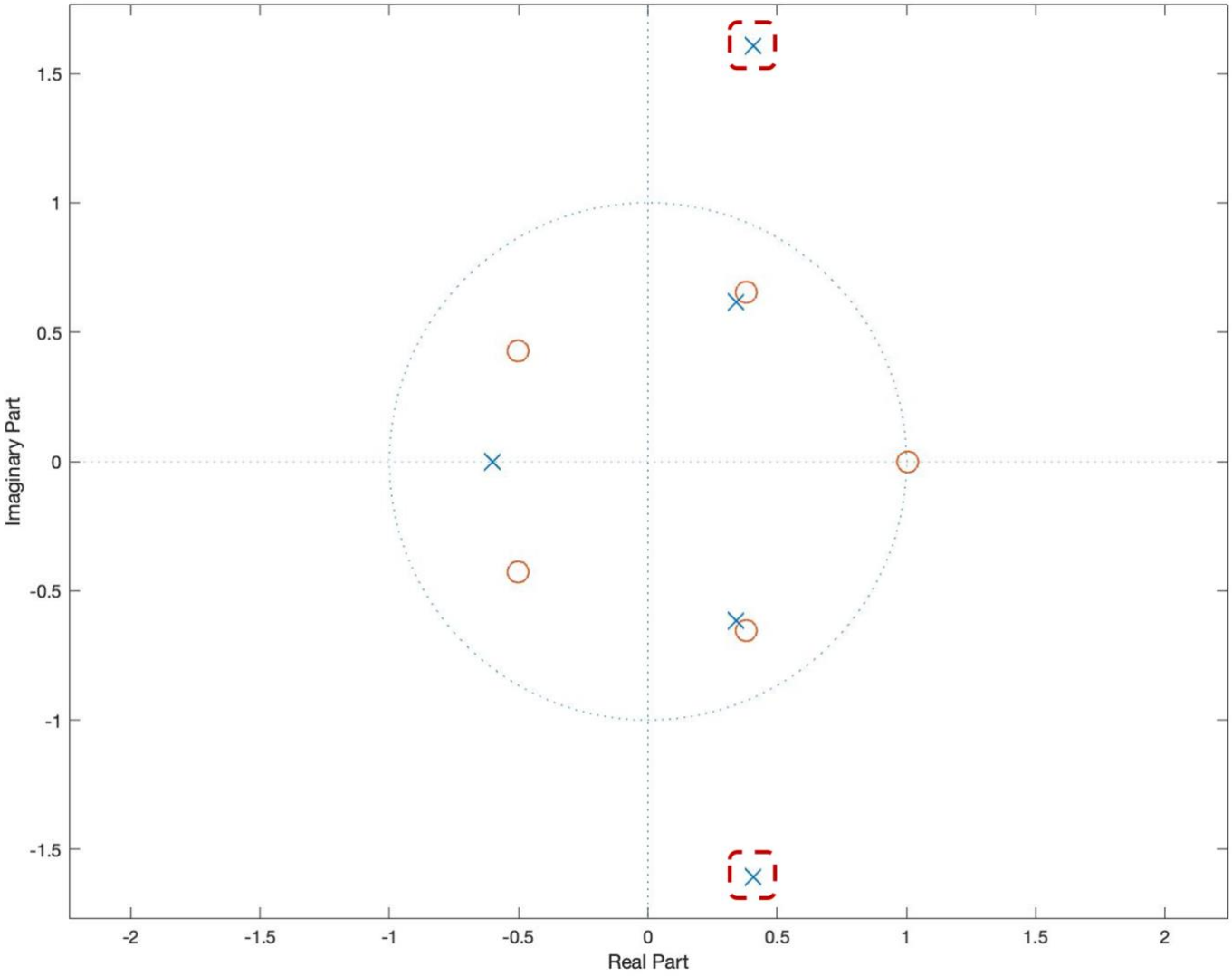

Figure 6.10: The pole-zero plot of the ARX model built using the Raw MIS Data collected by Probe 2c. Results presented that there was a complex-conjugate pole pair (circled by red) outside the unit circle. (orange circle: zero, blue cross: pole)

D. Calibration via modelling

The calibration via modelling was to remove components of the MIS device circuit and other measurements which do not belong to the cervix. With comparison of pole-zero plots of the ARX model built using raw MIS data measured by Probe 2a and the ARX model built using

raw MIS data measured by Probe 2c, the component of the cervix was identified in Figure 6.11, where the upper plot was the pole-zero plot of the ARX model built using raw MIS data measured by Probe 2a and the bottom plot was the pole-zero plot of the ARX model built using raw MIS data measured by Probe 2c. The common pole and zero presented in two pole-zero plots represented the common component of two ARX models, which were assumed to be the cervical component to be remained. The other poles and zeros were supposed to be the MIS device circuit and other components unrelated to cervix, which were second-order systems to be removed. In this case, the adjusted pole-zero plots were presented in Figure 6.12, where the common pole and zero of two ARX models were remained. The adjusted structure of the ARX model was represented by a zero-pole-gain form of transfer function as $H(z) = k\frac{z+zero_1}{z+pole_1}$.

Figure 6.13 showed the comparison of non-calibrated MIS spectra and calibrated MIS spectra, where the non-calibrated MIS spectra were generated by the original ARX model built using raw MIS data measured by Probe 2a, and the calibrated MIS spectra were generated by the corresponding adjusted ARX model. Though the comparison, the calibrated MIS spectra changed slightly from the non-calibrated MIS spectra.

Similarly, Figure 6.14 illustrated the comparison of non-calibrated MIS spectra and calibrated MIS spectra, where the non-calibrated MIS spectra were generated by the original ARX model built using raw MIS data measured by Probe 2c, and the calibrated MIS spectra were generated by the corresponding adjusted ARX model. Unlike the ARX model built using raw MIS data measured by Probe 2a, the adjusted ARX model of Probe 2c generated calibrated MIS spectra which were different from the non-calibrated MIS spectra. The calibrated MIS spectra become more similar to the calibrated MIS spectra generated by adjusted ARX model of Probe 2a. This phenomenon would prove that calibration via modelling achieved the aim to gather cervical MIS data measured by Probe 2a and Probe 2c together training a PTB risk prediction model.

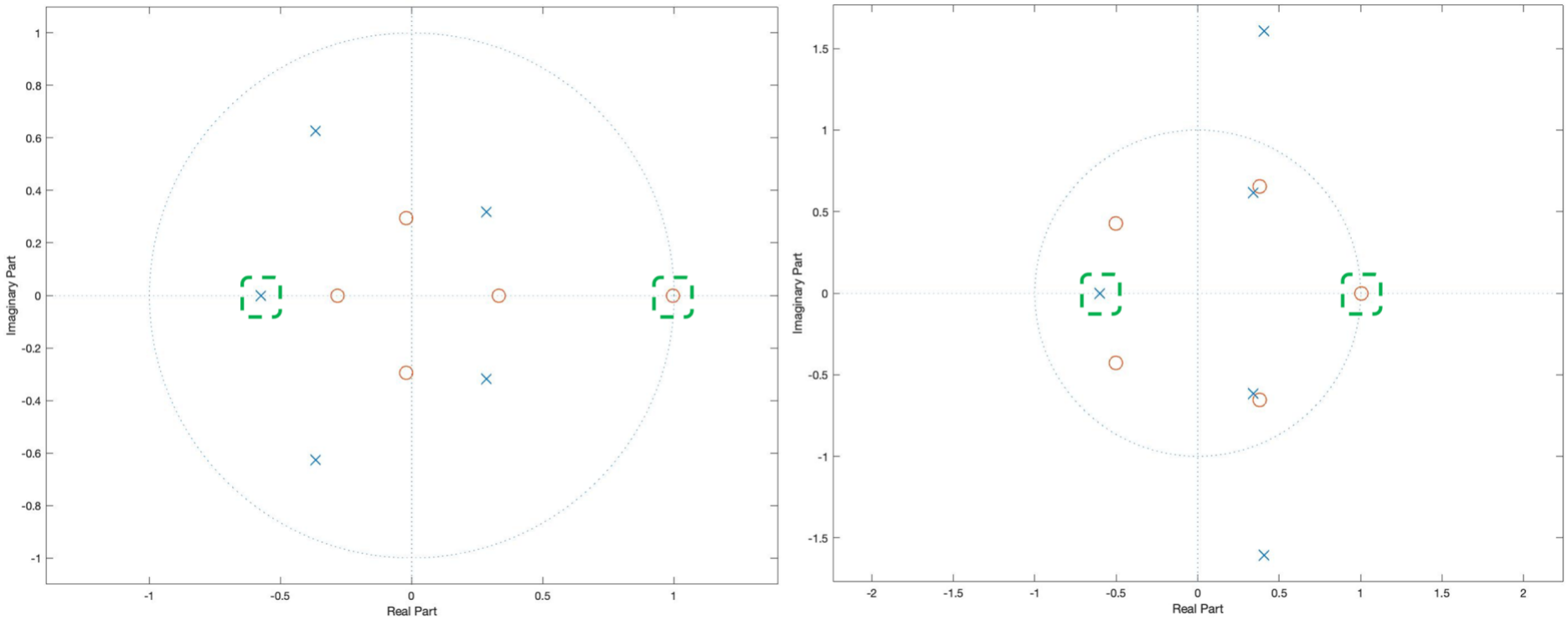

Figure 6.11: The comparison of pole-zero plots of the ARX model built using raw MIS data measured by Probe 2a (Left plot) and the ARX model built using raw MIS data measured by Probe 2c (Right plot). Results showed that the poles and zeros circled by green from two ARX models were similar, which can be supposed as cervical component to be remained.

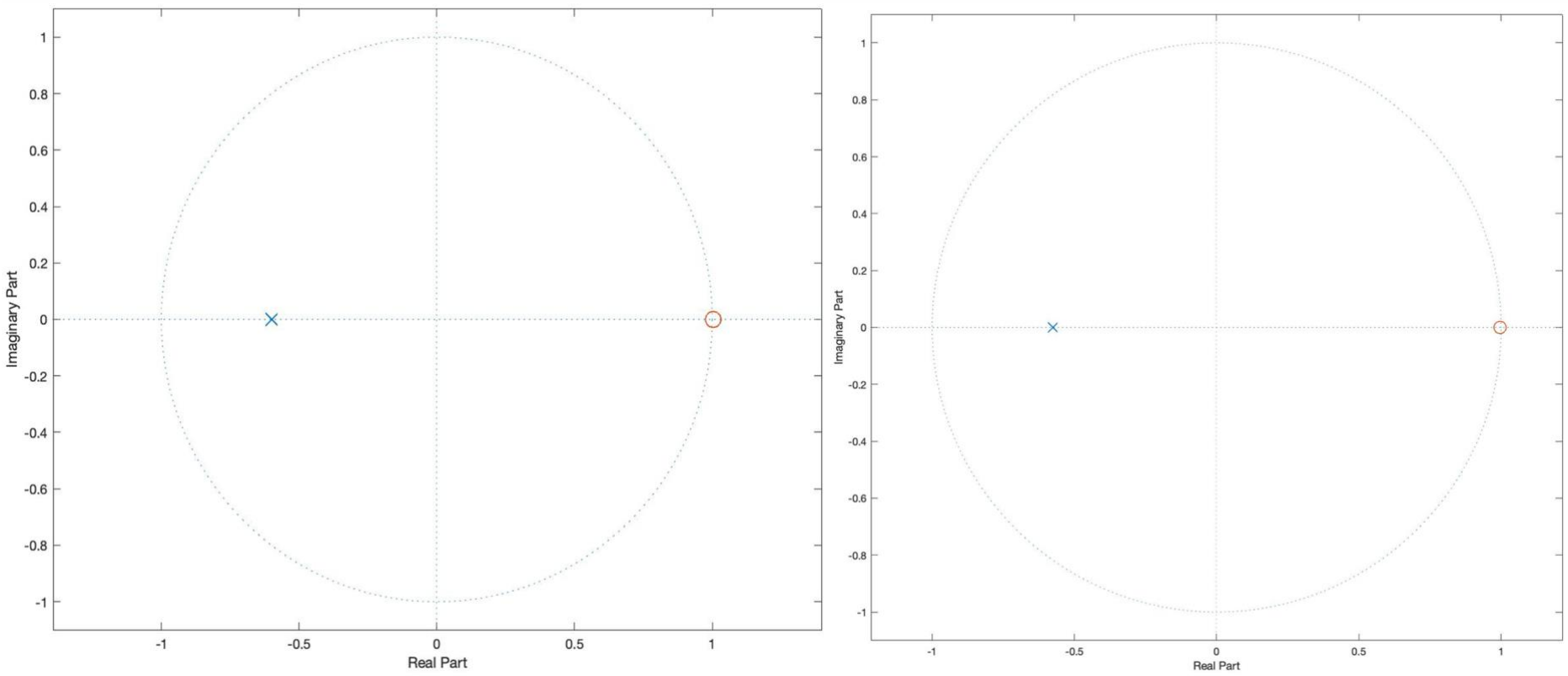

Figure 6.12: The adjusted pole-zero plots of the ARX model built using raw MIS data measured by Probe 2a (Left plot) and the ARX model built using raw MIS data measured by Probe 2c (Right plot). The two ARX models remained the common component of cervix.

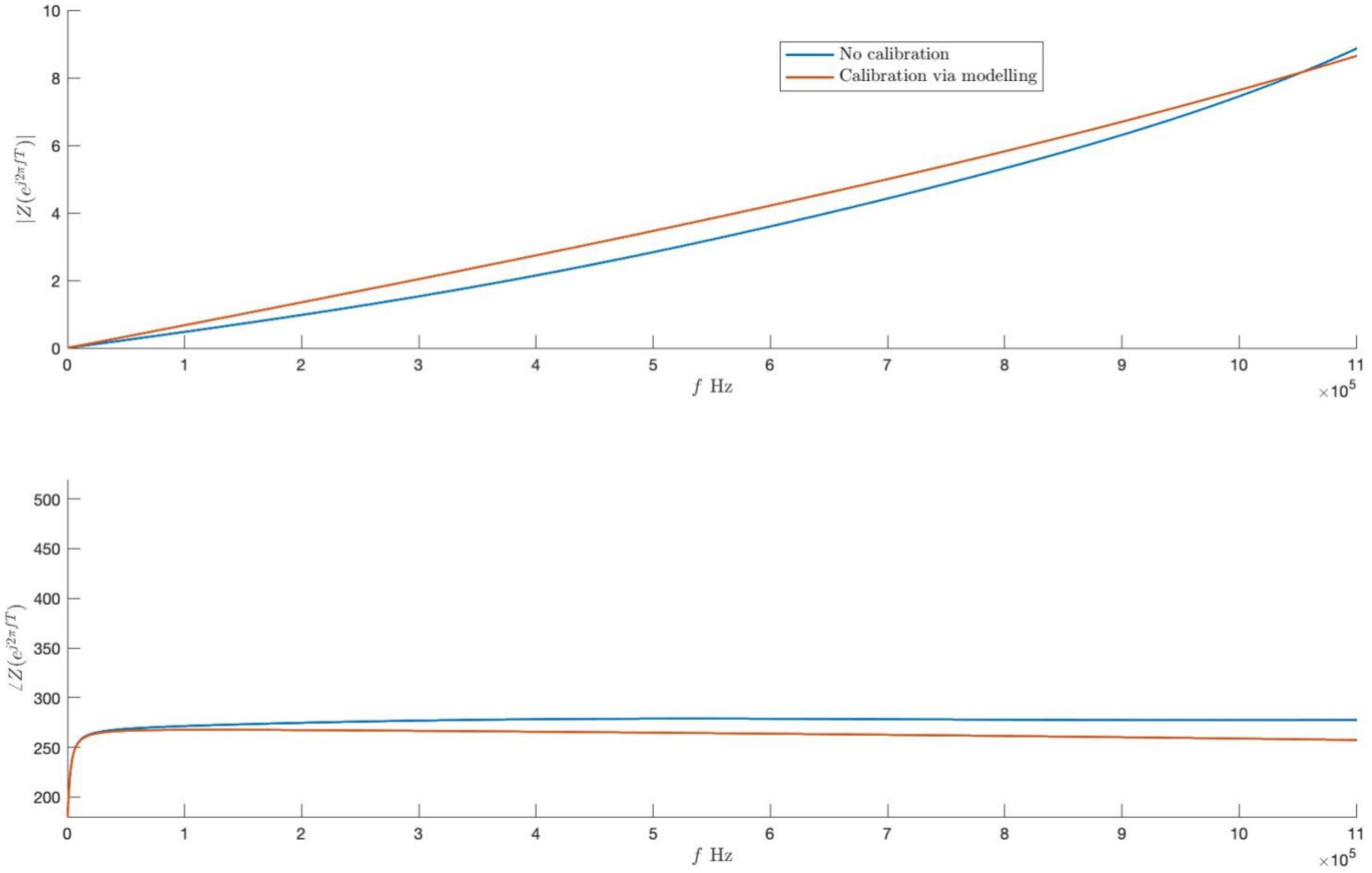

Figure 6.13: The comparison of non-calibrated MIS data and MIS data calibrated via modelling, and both were from the same woman measured by Probe 2a. The non-calibrated MIS data were generated by the original ARX model, and the MIS data calibrated via modelling was generated by the corresponding adjusted ARX model. Results presented amplitudes (Upper plot (ohm)) and phases (Bottom plot (degree)) of MIS data over 21 kHz to 1013 kHz.

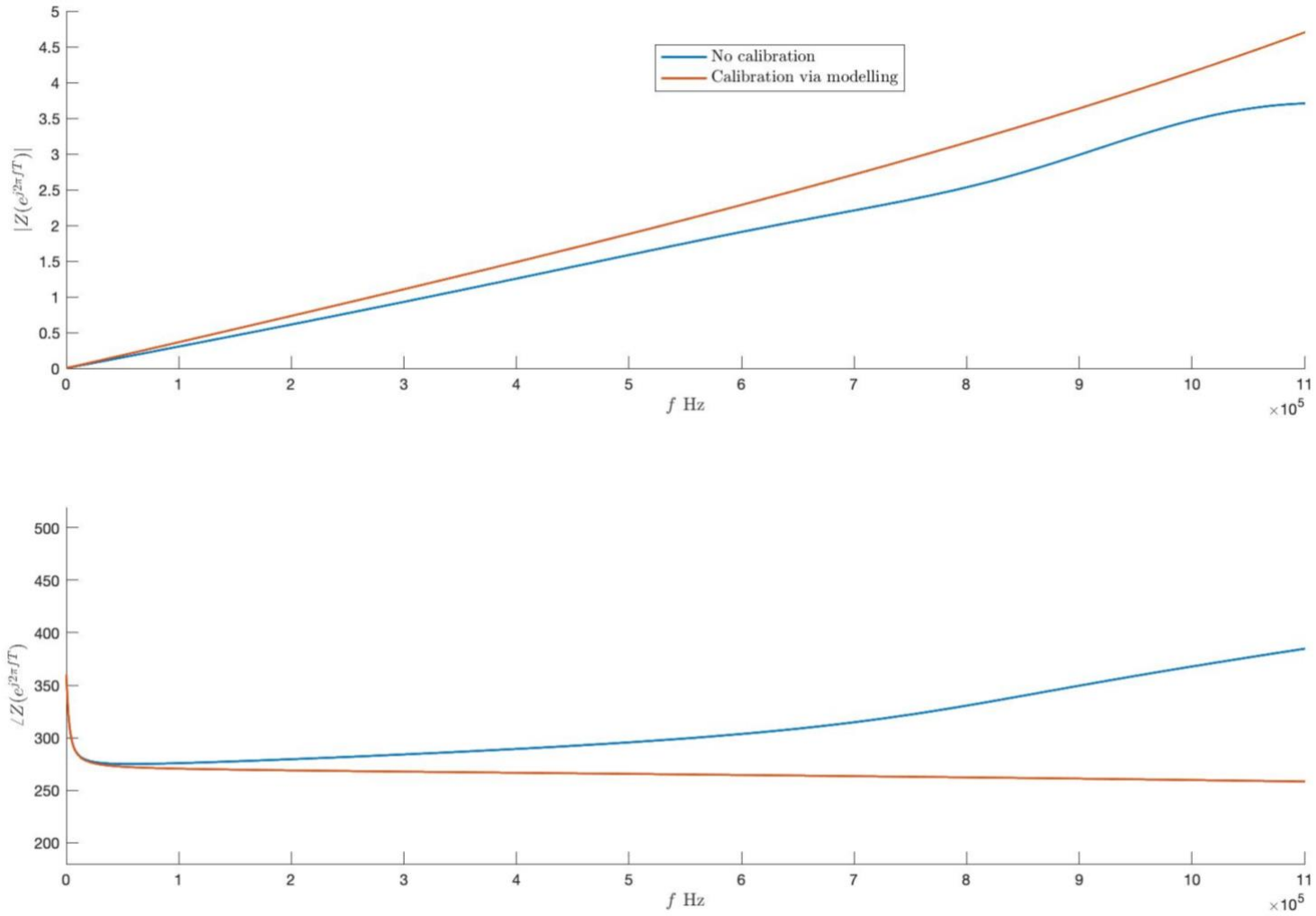

Figure 6.14: The comparison of non-calibrated MIS data and MIS data calibrated via modelling, and both were from the same woman measured by Probe 2c. The non-calibrated MIS data were generated by the original ARX model, and the MIS data calibrated via modelling was generated by the corresponding adjusted ARX model. Results presented amplitudes (Upper plot (ohm)) and phases (Bottom plot (degree)) of MIS data over 21 kHz to 1013 kHz.

6.5 Results of MIS-based PTB Risk Prediction

After MIS data (both real-time MIS data and raw MIS data) calibrated, six MIS datasets were constructed as described in Chapter 6.3.3, which were MIS dataset1a, dataset1b, dataset1c, dataset2a, dataset2b, and dataset2c. Also, 100 training sets and 100 testing sets were created from each MIS dataset. Therefore, 600 training sets and 600 testing sets were obtained, and 600 MIS-based PTB risk prediction models were trained according to the proposed machine learning pipeline.

Unlike the training process where each woman's records of cervix 1, cervix 2 and cervix 3 were regarded as three independent samples, in testing process, the three records of cervix 1, cervix 2 and cervix 3 were input to a MIS-based PTB risk prediction model respectively, then the model outputted three probabilities of PTB risk, which were averaged as the woman's final probability of PTB risk (illustrated in Figure 6.15).

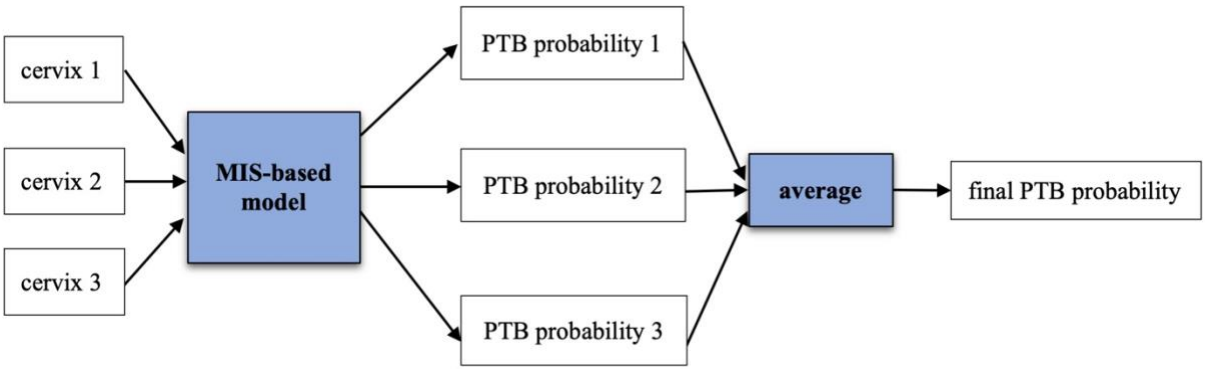

Figure 6.15: The illustration of testing process for a MIS-based PTB risk prediction model. MIS: magnetic impedance spectroscopy, PTB: preterm birth.

The training and testing performances of 600 MIS-based PTB risk prediction models were illustrated in Table 6.9 below. For datasets of combining 40 SYMP women and 44 AHR women at their first visit, training AUCs of 100 MIS-based PTB risk prediction models trained

by MIS dataset1a ranged from 0.72 to 0.87, and testing AUCs were from 0.30 to 0.68; training AUCs of 100 MIS-based PTB risk prediction models trained by MIS dataset1b ranged from 0.74 to 0.87, and testing AUCs were from 0.43 to 0.75; and training AUCs of 100 MIS-based PTB risk prediction models trained by MIS dataset1c ranged from 0.81 to 0.95, and testing AUCs were from 0.42 to 0.78.

For datasets of combining 40 SYMP women and 44 AHR women at their second visit, training AUCs of 100 MIS-based PTB risk prediction models trained by MIS dataset2a ranged from 0.71 to 0.88, and testing AUCs were from 0.29 to 0.69; training AUCs of 100 MIS-based PTB risk prediction models trained by MIS dataset2b ranged from 0.81 to 0.93, and testing AUCs were from 0.44 to 0.76; and training AUCs of 100 MIS-based PTB risk prediction models trained by MIS dataset2c ranged from 0.82 to 0.97, and testing AUCs were from 0.43 to 0.77.

From results, datasets of combining 40 SYMP women and 44 AHR women at their second visit could improve a little on training MIS-based PTB risk prediction models and the models' predicting ability. Besides, calibration via division and via modelling were better than calibration via subtraction for MIS-based PTB risk prediction.

Table 6.9: The training and testing performances of 600 MIS-based PTB prediction models trained by MIS dataset1a, MIS dataset1b, MIS dataset1c, MIS dataset2a, MIS dataset2b and MIS dataset2c respectively. Sensitivity and specificity were obtained using the optimal operating point by maximizing Youden's index on training set. MIS: magnetic impedance spectroscopy, AUC: area under the receiver operating characteristic curve.

600 MIS-based PTB risk prediction models				
MIS dataset that SVM model trained by	Calibration Method	Training AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
MIS dataset1a	calibrated via subtraction	0.77, 0.77 (0.72, 0.87)	0.76, 0.75 (0.59, 0.91)	0.61, 0.63 (0.42, 0.98)
MIS dataset1b	calibrated via division	0.85, 0.85 (0.74, 0.92)	0.84, 0.87 (0.55, 0.95)	0.69, 0.70 (0.51, 0.84)
MIS dataset1c	calibrated via modelling	0.86, 0.86 (0.81, 0.95)	0.87, 0.88 (0.61, 0.95)	0.70, 0.69 (0.62, 0.87)
MIS dataset2a	calibrated via subtraction	0.78, 0.78 (0.71, 0.88)	0.76, 0.76 (0.60, 0.94)	0.62, 0.62 (0.45, 0.97)
MIS dataset2b	calibrated via division	0.88, 0.87 (0.81, 0.93)	0.84, 0.85 (0.58, 0.92)	0.71, 0.70 (0.53, 0.93)
MIS dataset2c	calibrated via modelling	0.90, 0.91 (0.82, 0.97)	0.89, 0.90 (0.55, 0.98)	0.71, 0.71 (0.56, 0.97)
MIS dataset that SVM model trained by	Calibration Method	Testing AUC median, mean (min, max)	Sensitivity median, mean (min, max)	Specificity median, mean (min, max)
MIS dataset1a	calibrated via subtraction	0.47, 0.43 (0.30, 0.68)	0.41, 0.42 (0.01, 0.84)	0.62, 0.60 (0.22, 0.86)
MIS dataset1b	calibrated via division	0.59, 0.58 (0.43, 0.75)	0.51, 0.49 (0.12, 0.85)	0.66, 0.63 (0.30, 0.93)
MIS dataset1c	calibrated via modelling	0.60, 0.58 (0.42, 0.78)	0.59, 0.55 (0.27, 0.98)	0.65, 0.62 (0.25, 0.90)
MIS dataset2a	calibrated via subtraction	0.50, 0.46 (0.29, 0.69)	0.40, 0.42 (0.04, 0.92)	0.63, 0.61 (0.21, 0.84)
MIS dataset2b	calibrated via division	0.61, 0.59 (0.44, 0.76)	0.54, 0.51 (0.24, 0.90)	0.64, 0.64 (0.27, 0.92)
MIS dataset2c	calibrated via modelling	0.61, 0.58 (0.43, 0.77)	0.61, 0.60 (0.30, 0.99)	0.67, 0.64 (0.31, 0.98)

6.6 Conclusion

This chapter has applied calibration approaches to implement the MIS data pre-processing, where the first calibration was via substruction recommended by the ECCLIPPx II, and the additional two calibration approaches were proposed in this work, which were calibration via division and calibration via modelling. Furthermore, this work applied the system identification technique to create an ARX model using raw MIS data, and the ARX model was adjusted to remain the cervical component only. After that, real time MIS data and raw MIS data were calibrated for PTB risk prediction. Based on a machining learning pipeline of training the MIS-based PTB risk prediction model, 600 MIS-based PTB risk prediction models using SVM classifiers were trained on 600 training sets. From results, calibration via division and calibration via modelling outperformed calibration via substruction for PTB risk prediction, met the challenge of MIS data pre-processing and achieved the aim of gathering cervical MIS data measured by Probe 2a and Probe 2c together to train a PTB risk prediction model. The MIS-based PTB risk prediction models had good performances on training but presented poor performances on predicting, and needed a more comprehensive further study.

Chapter 7

7. Conclusions and Future Works

The PhD study work was based on the ECCLIPPx^{TM} study and the ECCLIPPx II study of Electrical Impedance Spectroscopy and Magnetic Impedance Spectroscopy technologies for PTB risk prediction. The aim of this study was to analyse Electrical Impedance Spectroscopy signals and Magnetic Impedance Spectroscopy signals and develop machine learning-based PTB risk prediction models. The significances, limitations and future works of this study were concluded as follows.

Significances

(1) I analysed EIS data collected from 438 recruited women at 19 to 23 weeks of gestation. A principle of identifying the optimum EIS spectrum among several recorded EIS spectra of one woman was proposed to improve the quality of EIS data and the accuracy of PTB risk prediction. The optimal-EIS spectrum filter model has been trained to select optimal EIS spectrum, and the EIS-based PTB risk prediction model has been trained to predict PTB risk. The best EIS-based PTB risk prediction model (an OBCC-based LR model) performed the highest testing AUC of 0.85 (95% CI: 0.77 – 0.93, sensitivity: 0.70, specificity: 0.87), and a training AUC of 0.81 (95% CI: 0.74 – 0.86, sensitivity: 0.85, specificity: 0.68).

The proposed EIS-based PTB risk prediction model overcame the problem of EIS data heterogeneity by applying the optimal-EIS spectrum filter model, compared with the general EIS-based PTB risk prediction model without using the optimal-EIS spectrum filter model. Also, the proposed EIS-based PTB risk prediction model guaranteed the generalizability in some way, since the training data were from different age ranges and different risk levels, and prolonged the prediction period from 21 days to 12 weeks on average.

(2) I further considered a models-combining approach to achieve the combination of a cervical EIS-based PTB risk prediction model and a maternal characteristics-based PTB risk prediction model, improving the accuracy of PTB risk prediction. The maternal characteristics of the same 438 recruited women were analysed and used to train the maternal characteristics-based PTB risk prediction model. The EIS-based PTB risk prediction was the OBCC-based LR model, which has performed the best performances for PTB risk prediction. Based on the models-combining approach, the SVM-based combined model and the LR-based combined model were trained. The best combined model (a SVM-based combined model) had the highest testing AUC of 0.89 (95% CI: 0.80 – 0.97, sensitivity: 0.60, specificity: 0.90), and a training AUC of 0.82 (95% CI: 0.76 – 0.89, sensitivity: 0.67, specificity: 0.85).

The combined model had an improvement on PTB prediction compared with PTB prediction models only based on maternal characteristics, CL, fFN and EIS, and outperformed the state-of-the-art models. Also, the combined model achieved to apply multiple PTB risk factors to predict PTB risk and improve the accuracy.

(3) Finally, I analysed the MIS data collected from 84 recruited women at 18 to 35 weeks of gestation. Before application of machine leaning for PTB risk prediction, two calibration methods were proposed in this study to process both directly collected MIS data and raw MIS data, which were calibration via division and calibration via modelling. After MIS data calibrated, six MIS-based PTB risk prediction models were trained on different types of MIS training sets to evaluate the feasibility of using MIS data for PTB risk prediction.

From results, two proposed calibration methods were better than calibration via substruction recommended by the ECCLIPPx II study. And calibration via modelling can achieve remove information unrelated to cervix, to only remain cervical information. Therefore, this calibration method can be regarded as a data pre-processing technique before building PTB risk prediction model. The MIS-based PTB risk prediction models had good training performances, but had poor testing performances. In this case, a further study is needed to carry on.

Limitations and Future Works

(1) For EIS-based PTB risk prediction models and combined models, most of their median testing AUCs were greater than their mean testing AUCs. Also, testing AUCs of the best EIS-based PTB risk prediction models and the best combined models were greater than training

AUCs. The models need to be more generalized. In the future work, firstly I will analyse the EIS data to ensure if there are outliers identified. If there are outliers, I will remove them and then repeat the implements in Chapter 4 and 5 to check if there is any improvement. Secondly, I will evaluate the training set and testing set to guarantee the sample can represent the population. Thirdly, the non-linear correlation between EIS data and PTB risk needs to be employed and evaluate to improve the feature selection approaches applied in this study.

(2) When applying the feature selection approaches applied in this study, the number of features to select needs to be setup firstly according to prior knowledge, since these feature selection approaches have no regularization term or penalty term to determine the number of features. In future work, advanced feature selection methods or embedded feature selection methods can be employed.

(3) The sample size of MIS data was not large enough to create a generalized machine learning model. More MIS data need to be collected in the future. In some cases, cervical MIS signals were weak, that means the differences between cervical values and air reference's values or between cervical values and vaginal values were not obvious. Only limited information can be remained after calibration. In future work, the MIS device needs to be improved or evaluate the quality of MIS data after measuring.

Bibliography

[1] Anonymous. WHO; recommended definitions, terminology and format for statistical tables related to the perinatal period and use of a new certificate for cause of perinatal deaths. Modifications recommended by FIGO as amended October 14, 1976. Acta Obstet Gynecol Scand. 1977; 56(3):247–53.

[2] Liu, Li, PhD et al., "Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000," The Lancet (British edition), vol. 379, no. 9832, pp. 2151–2161, 2012, doi: 10.1016/S0140-6736(12)60560-1.

[3] S. Beck et al., "The worldwide incidence of preterm birth: A systematic review of maternal mortality and morbidity," Bulletin of the World Health Organization, vol. 88, no. 1, pp. 31–38, 2010, doi: 10.2471/BLT.08.062554.

[4] Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, Narwal R, Adler A, Vera Garcia C, Rohde S, Say L, Lawn JE: 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 2012, 379:2162-2172.

[5] M. Saifon Chawanpaiboon et al., "Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis," The Lancet global health, vol. 7, no. 1, pp. e37–e46, 2019.

[6] C. P. Howson, M. V. Kinney, L. McDougall, J. E. Lawn, and T. S. P. B. A. G. Born Too Soon Preterm Birth Action Group, "Born too soon: preterm birth matters," Reproductive health, vol. 10, no. S1, pp. S1–S1, 2013, doi: 10.1186/1742-4755-10-S1-S1.

[7] Mangham, L. et al., 2009. The Cost of Preterm Birth Throughout Childhood in England and Wales. Pediatrics, 123(2), pp.312–27.

[8] W. Huang, S. Ural, and Y. Zhu, “Preterm labor tests: current status and future directions,” Critical reviews in clinical laboratory sciences, pp. 1–19, 2022, doi: 10.1080/10408363.2022.2027864.

[9] E. Tsoi, S. Akmal, S. Rane, C. Otigbah, and K. H. Nicolaides, “Ultrasound assessment of cervical length in threatened preterm labor,” Ultrasound in obstetrics & gynecology, vol. 21, no. 6, pp. 552–555, 2003, doi: 10.1002/uog.131.

[10] I. B. Fuchs, W. Henrich, K. Osthues, and J. W. Dudenhausen, “Sonographic cervical length in singleton pregnancies with intact membranes presenting with threatened preterm labor,” Ultrasound in obstetrics & gynecology, vol. 24, no. 5, pp. 554–557, 2004, doi: 10.1002/uog.1714.

[11] H. M. Georgiou, M. K. W. Di Quinzio, M. Permezel, and S. P. Brennecke, “Predicting Preterm Labour: Current Status and Future Prospects,” Disease markers, vol. 2015, pp. 435014–9, 2015, doi: 10.1155/2015/435014.

[12] Shah J. et al., 2016. Identification of biomarkers for prediction of preterm delivery. Journal of Medical Society, 30:3-14.

[13] N. Suff, L. Story, and A. Shennan, “The prediction of preterm delivery: What is new?,” Seminars in fetal & neonatal medicine, vol. 24, no. 1, pp. 27–32, 2019, doi: 10.1016/j.siny.2018.09.006.

[14] Honest H, Bachmann LM, Coomarasamy A, Gupta JK, Kleijnen J, and Khan KS, “Accuracy of cervical transvaginal sonography in predicting preterm birth: a systematic review,” The ultrasound review of obstetrics and gynecology, vol. 3, no. 3, p. 305, 2003.

[15] A. C. Lim, M. A. [=Maud A. Hegeman, M. A. Huis In 't Veld, B. C. Opmeer, H. W. Bruinse, and B. W. J. Mol, “Cervical length measurement for the prediction of preterm birth in multiple pregnancies: a systematic review and bivariate meta-analysis,” Ultrasound in obstetrics & gynecology, vol. 38, no. 1, pp. 10–17, 2011, doi: 10.1002/uog.9013.

[16] Boots, Amy B., DO, Sanchez-Ramos, Luis, MD, Bowers, Dawn M., MD, Kaunitz, Andrew M., MD, Zamora, Javier, PhD, and Schlattmann, Peter, MD, PhD, “The short-term prediction of preterm birth: a systematic review and diagnostic

metaanalysis," American journal of obstetrics and gynecology, vol. 210, no. 1, pp. 54.e1–54.e10, 2014, doi: 10.1016/j.ajog.2013.09.004.

[17] Vandermolen et al., 2016. Quantitative fetal fibronectin and cervical length to predict preterm birth in asymptomatic women with previous cervical surgery. American Journal of Obstetrics and Gynecology, 215(4), pp.480.e1–480.e10.

[18] N. Melamed, L. Hiersch, I. Meizner, R. Bardin, A. Wiznitzer, and Y. Yogev, "Is measurement of cervical length an accurate predictive tool in women with a history of preterm delivery who present with threatened preterm labor?" Ultrasound in obstetrics & gynecology, vol. 44, no. 6, pp. 661–668, 2014, doi: 10.1002/uog.13395.

[19] Conde-Agudelo, Agustin, MD, MPH, PhD and Romero, Roberto, MD, DMedSci, "Predictive accuracy of changes in transvaginal sonographic cervical length over time for preterm birth: a systematic review and metaanalysis," American journal of obstetrics and gynecology, vol. 213, no. 6, pp. 789–801, 2015, doi: 10.1016/j.ajog.2015.06.015.

[20] M. S. Esplin et al., "Predictive Accuracy of Serial Transvaginal Cervical Lengths and Quantitative Vaginal Fetal Fibronectin Levels for Spontaneous Preterm Birth Among Nulliparous Women," JAMA : the journal of the American Medical Association, vol. 317, no. 10, pp. 1047–1056, 2017, doi: 10.1001/jama.2017.1373.

[21] L. Hiersch, N. Melamed, A. Aviram, R. Bardin, Y. Yogev, and E. Ashwal, "Role of Cervical Length Measurement for Preterm Delivery Prediction in Women With Threatened Preterm Labor and Cervical Dilatation," Journal of ultrasound in medicine, vol. 35, no. 12, pp. 2631–2640, 2016, doi: 10.7863/ultra.15.12007.

[22] Hiersch, Liran, MD, Yogev, Yariv, MD, Domniz, Noam, MD, Meizner, Israel, MD, Bardin, Ron, MD, and Melamed, Nir, MD, "The role of cervical length in women with threatened preterm labor: is it a valid predictor at any gestational age?" American journal of obstetrics and gynecology, vol. 211, no. 5, pp. 532.e1–532.e9, 2014, doi: 10.1016/j.ajog.2014.06.002.

[23] H. Leitich, C. Egarter, A. Kaider, M. Hohlagschwandtner, P. Berghammer, and P. Husslein, "Cervicovaginal fetal fibronectin as a marker for preterm delivery: A meta-

analysis," American journal of obstetrics and gynecology, vol. 180, no. 5, pp. 1169–1176, 1999, doi: 10.1016/S0002-9378(99)70612-5.

[24] S. N. Deshpande et al., "Rapid fetal fibronectin testing to predict preterm birth in women with symptoms of premature labour: a systematic review and cost analysis," Health technology assessment (Winchester, England), vol. 17, no. 40, p. 1–+, 2013, doi: 10.3310/hta17400.

[25] V. Berghella, E. Hayes, J. Visintine, and J. K. Baxter, "Fetal fibronectin testing for reducing the risk of preterm birth," Cochrane database of systematic reviews, vol. 2008, no. 4, pp. CD006843–CD006843, 2008, doi: 10.1002/14651858.CD006843.pub2.

[26] Zhou, Mei-Xi et al., 2015. Evaluation of the ability of cervical length and fetal fibronectin measurement to predict preterm delivery in asymptomatic women with risk factors. The Journal of Maternal-Fetal & Neonatal Medicine, 28(2), pp.153–157.

[27] Abbott et al., 2012. Evaluation of a quantitative fetal fibronectin test for spontaneous preterm birth in symptomatic women. American Journal of Obstetrics and Gynecology, 208(2), pp.122.e1–e6.

[28] Jwala, S. et al., 2016. Evaluation of additive effect of quantitative fetal fibronectin to cervical length for prediction of spontaneous preterm birth among asymptomatic low-risk women. Acta Obstetricia et Gynecologica Scandinavica, 95(8), pp.948–955.

[29] N. L. Hezelgrave and A. H. Shennan, "Quantitative fetal fibronectin to predict spontaneous preterm birth: a review," Women's Health, vol. 12, no. 1, pp. 121–128, 2016, doi: 10.2217/whe.15.74.

[30] G. Rizzo, A. Capponi, D. Arduini, C. Lorido, and C. Romanini, "The value of fetal fibronectin in cervical and vaginal secretions and of ultrasonographic examination of the uterine cervix in predicting premature delivery for patients with preterm labor and intact membranes," American journal of obstetrics and gynecology, vol. 175, no. 5, pp. 1146–1151, 1996, doi: 10.1016/S0002-9378(96)70020-0.

[31] D. Dudley, "Serial transvaginal cervical length measurements and quantitative vaginal fetal fibronectin concentrations did not predict spontaneous preterm birth in low-risk

nulliparous women," Evidence-based medicine (English ed.), vol. 22, no. 5, pp. 188–188, 2017, doi: 10.1136/ebmed-2017-110761.

[32] F. Fuchs, M. Houllier, S. Leparco, A. Guyot, M.-V. Senat, and H. Fernandez, "Performance of cervical phIGFBP-1 test alone or combined with short cervical length to predict spontaneous preterm birth in symptomatic women," Scientific reports, vol. 7, no. 1, pp. 10856–6, 2017, doi: 10.1038/s41598-017-11447-y.

[33] M. Goyal, A. Kriplani, G. Kachhawa, and S. Badiger, "Prediction of preterm labor by a rapid bedside test detecting phosphorylated insulin-like growth factor-binding protein 1 in cervical secretions," International journal of gynecology and obstetrics, vol. 134, no. 2, pp. 165–168, 2016, doi: 10.1016/j.ijgo.2016.01.019.

[34] Conde-Agudelo, Agustin, MD, MPH, PhD and Romero, Roberto, MD, DMedSci, "Cervical phosphorylated insulin-like growth factor binding protein-1 test for the prediction of preterm birth: a systematic review and metaanalysis," American journal of obstetrics and gynecology, vol. 214, no. 1, pp. 57–73, 2016, doi: 10.1016/j.ajog.2015.06.060.

[35] S. Cooper, I. Lange, S. Wood, S. Tang, L. Miller, and S. Ross, "Diagnostic accuracy of rapid phIGFBP-I assay for predicting preterm labor in symptomatic patients," Journal of perinatology, vol. 32, no. 6, pp. 460–465, 2012, doi: 10.1038/jp.2011.133.

[36] F. Riboni, A. Vitulo, M. Dell'avanzo, M. Plebani, G. Battagliarin, and D. Paternoster, "Biochemical markers predicting pre-term delivery in symptomatic patients: phosphorylated insulin-like growth factor binding protein-1 and fetal fibronectin," Archives of gynecology and obstetrics, vol. 284, no. 6, pp. 1325–1329, 2011, doi: 10.1007/s00404-011-1839-4.

[37] L. Danti, F. Prefumo, A. Lojacono, S. Corini, A. Testori, and T. Frusca, "The combination of short cervical length and phIGFBP-1 in the prediction of preterm delivery in symptomatic women," The journal of maternal-fetal & neonatal medicine, vol. 24, no. 10, pp. 1262–1266, 2011, doi: 10.3109/14767058.2010.547962.

[38] M. Brik, A. I. M. Hernández, C. C. Pedraz, and A. Perales, "Phosphorylated insulin-like growth factor binding protein-1 and cervical measurement in women with

threatening preterm birth," Acta obstetricia et gynecologica Scandinavica, vol. 89, no. 2, pp. 268–274, 2010, doi: 10.3109/00016340903443668.

[39] J. C. Melchor, A. Khalil, D. Wing, E. Schleussner, and D. Surbek, "Prediction of preterm delivery in symptomatic women using PAMG-1, fetal fibronectin and phIGFBP-1 tests: systematic review and meta-analysis," Ultrasound in obstetrics & gynecology, vol. 52, no. 4, pp. 442–451, 2018, doi: 10.1002/uog.19119.

[40] Wing DA, Haeri S, Silber AC, Roth CK, Weiner CP, Echebiri NC, Franco A, Pappas LM, Yeast JD, Brebnor AA, Quirk JG, Murphy AM, Laurent LC, Field NT, Norton ME. Placental alpha microglobulin-1 compared with fetal fibronectin to predict preterm delivery in symptomatic women. Obstet Gynecol 2017; 130: 1183–1191.

[41] Nikolova T, Uotila J, Nikolova N, Borisova VY, Bolotskikh Nikolova T, Uotila J, Nikolova N, Borisova VY, Bolotskikh VM. Do PAMG-1 or phIGFBP-1 biomarkers improve the prediction of imminent spontaneous preterm delivery in PTL symptomatic women with non-obvious cervical length (CL)? Am J Obstet Gynecol 2017; 216 (1 suppl): S11–S12.

[42] J. C. Melchor et al., "Predictive performance of PAMG-1 vs fFN test for risk of spontaneous preterm birth in symptomatic women attending an emergency obstetric unit: retrospective cohort study," Ultrasound in obstetrics & gynecology, vol. 51, no. 5, pp. 644–649, 2018, doi: 10.1002/uog.18892.

[43] T. Nikolova, O. Bayev, N. Nikolova, and G. C. Di Renzo, "Evaluation of a novel placental alpha microglobulin-1 (PAMG-1) test to predict spontaneous preterm delivery," Journal of perinatal medicine, vol. 42, no. 4, pp. 473–477, 2014, doi: 10.1515/jpm-2013-0234.

[44] M. Ravi, M. Beljorie, and K. El Masry, "Evaluation of the quantitative fetal fibronectin test and PAMG-1 test for the prediction of spontaneous preterm birth in patients with signs and symptoms suggestive of preterm labor," The journal of maternal-fetal & neonatal medicine, vol. 32, no. 23, pp. 3909–3914, 2019, doi: 10.1080/14767058.2018.1476485.

[45] S. Yoneda et al., “Prediction of exact delivery time in patients with preterm labor and intact membranes at admission by amniotic fluid interleukin-8 level and preterm labor index,” The journal of obstetrics and gynaecology research, vol. 37, no. 7, pp. 861–866, 2011, doi: 10.1111/j.1447-0756.2010.01453.x.

[46] S. Yoneda, M. Sakai, Y. Sasaki, A. Shiozaki, T. Hidaka, and S. Saito, “Interleukin-8 and glucose in amniotic fluid, fetal fibronectin in vaginal secretions and preterm labor index based on clinical variables are optimal predictive markers for preterm delivery in patients with intact membranes,” The journal of obstetrics and gynaecology research, vol. 33, no. 1, pp. 38–44, 2007, doi: 10.1111/j.1447-0756.2007.00474.x.

[47] P. C. Ng et al., “Early Prediction of Sepsis-Induced Disseminated Intravascular Coagulation with Interleukin-10, Interleukin-6, and RANTES in Preterm Infants,” Clinical chemistry (Baltimore, Md.), vol. 52, no. 6, pp. 1181–1189, 2006, doi: 10.1373/clinchem.2005.062075.

[48] N. Thomakos, G. Daskalakis, A. Papapanagiotou, N. Papantoniou, S. Mesogitis, and A. Antsaklis, “Amniotic fluid interleukin-6 and tumor necrosis factor-α at mid-trimester genetic amniocentesis: Relationship to intra-amniotic microbial invasion and preterm delivery,” European journal of obstetrics & gynecology and reproductive biology, vol. 148, no. 2, pp. 147–151, 2009, doi: 10.1016/j.ejogrb.2009.10.027.

[49] R.-M. Holst, H. Hagberg, U.-B. Wennerholm, K. Skogstrand, P. Thorsen, and B. Jacobsson, “Prediction of spontaneous preterm delivery in women with preterm labor: Analysis of multiple proteins in amniotic and cervical fluids,” Obstetrics and gynecology (New York. 1953), vol. 114, no. 2, pp. 268–277, 2009, doi: 10.1097/AOG.0b013e3181ae6a08.

[50] G.-H. Son, Y.-A. You, E.-J. Kwon, K.-Y. Lee, and Y. J. Kim, “Comparative Analysis of Midtrimester Amniotic Fluid Cytokine Levels to Predict Spontaneous Very Pre-term Birth in Patients with Cervical Insufficiency,” American journal of reproductive immunology (1989), vol. 75, no. 2, pp. 155–161, 2016, doi: 10.1111/aji.12451.

[51] L. Huang et al., “Serum multiple cytokines for the prediction of spontaneous preterm birth in asymptomatic women: A nested case-control study,” Cytokine (Philadelphia, Pa.), vol. 117, pp. 91–97, 2019, doi: 10.1016/j.cyto.2019.02.007.

[52] M. Pandey, M. Chauhan, and S. Awasthi, “Interplay of cytokines in preterm birth,” Indian journal of medical research (New Delhi, India : 1994), vol. 146, no. 3, pp. 316–327, 2017, doi: 10.4103/ijmr.IJMR_1624_14.

[53] Q. Zhu, J. Sun, and Y. Chen, “Preterm birth and single nucleotide polymorphisms in cytokine genes,” Translational pediatrics, vol. 3, no. 2, pp. 120–134, 2014, doi: 10.3978/j.issn.2224-4336.2014.03.02.

[54] P. Amini, S. Maroufizadeh, R. O. Samani, O. Hamidi, and M. Sepidarkish, “Prevalence and determinants of preterm birth in Tehran, Iran: A comparison between logistic regression and decision tree methods,” Osong public health and research perspectives, vol. 8, no. 3, pp. 195–200, 2017, doi: 10.24171/j.phrp.2017.8.3.06

[55] Alleman, Brandon W., BS, BA et al., “A proposed method to predict preterm birth using clinical data, standard maternal serum screening, and cholesterol,” American journal of obstetrics and gynecology, vol. 208, no. 6, pp. 472.e1–472.e11, 2013, doi: 10.1016/j.ajog.2013.03.005.

[56] A. Weber et al., “Application of machine-learning to predict early spontaneous preterm birth among nulliparous non-Hispanic black and white women,” Annals of epidemiology, vol. 28, no. 11, pp. 783–789.e1, 2018, doi: 10.1016/j.annepidem.2018.08.008.

[57] A. García-Blanco, V. Diago, V. Serrano De La Cruz, D. Hervás, C. Cháfer-Pericás, and M. Vento, “Can stress biomarkers predict preterm birth in women with threatened preterm labor?,” Psychoneuroendocrinology, vol. 83, pp. 19–24, 2017, doi: 10.1016/j.psyneuen.2017.05.021.

[58] T. Tran, W. Luo, D. Phung, J. Morris, K. Rickard, and S. Venkatesh, “Preterm Birth Prediction: Deriving Stable and Interpretable Rules from High Dimensional Data,” 2016.

[59] M. A. El-Ardat et al., “Ultrasound measurement of cervical length as predictor of threatened preterm birth: A predictive model,” Acta informatica medica, vol. 22, no. 5, pp. 306–308, 2014, doi: 10.5455/aim.2014.22.306-308.

[60] Y.-Z. Zhu, G.-Q. Peng, G.-X. Tian, X.-L. Qu, and S.-Y. Xiao, “New model for predicting preterm delivery during the second trimester of pregnancy,” Scientific reports, vol. 7, no. 1, pp. 11294–9, 2017, doi: 10.1038/s41598-017-11286-x.

[61] J.-R. He et al., “Predictions of preterm birth from early pregnancy characteristics: Born in Guangzhou cohort study,” Journal of clinical medicine, vol. 7, no. 8, p. 185, 2018, doi: 10.3390/jcm7080185.

[62] R. Raja, I. Mukherjee, and B. K. Sarkar, “A Machine Learning-Based Prediction Model for Preterm Birth in Rural India,” Journal of healthcare engineering, vol. 2021, pp. 6665573–11, 2021, doi: 10.1155/2021/6665573.

[63] Z. Sharifi-Heris, J. Laitala, A. Airola, A. M. Rahmani, and M. Bender, “Machine Learning Approach for Preterm Birth Prediction Using Health Records: Systematic Review,” JMIR Medical Informatics, vol. 10, no. 4, pp. e33875–e33875, 2022, doi: 10.2196/33875.

[64] T. Włodarczyk et al., “Machine learning methods for preterm birth prediction: A review,” Electronics (Basel), vol. 10, no. 5, pp. 1–24, 2021, doi: 10.3390/electronics10050586.

[65] J.-I. Kim and J. Y. Lee, “Systematic Review of Prediction Models for Preterm Birth Using CHARMS,” Biological research for nursing, vol. 23, no. 4, pp. 708–722, 2021, doi: 10.1177/10998004211025641.

[66] J. Garcia-Casado, Y. Ye-Lin, G. Prats-Boluda, J. Mas-Cabo, J. Alberola-Rubio, and A. Perales, “Electrohysterography in the diagnosis of preterm birth: a review,” Physiological measurement, vol. 39, no. 2, pp. 02TR01–02TR01, 2018, doi: 10.1088/1361-6579/aaad56.

[67] P. Fergus, P. Cheung, A. Hussain, D. Al-Jumeily, C. Dobbins, and S. Iram, “Prediction of Preterm Deliveries from EHG Signals Using Machine Learning,” PloS one, vol. 8, no. 10, p. e77154, 2013, doi: 10.1371/journal.pone.0077154.

[68] P. Fergus, I. Idowu, A. Hussain, and C. Dobbins, “Advanced artificial neural network classification for detecting preterm births using EHG records,” Neurocomputing (Amsterdam), vol. 188, pp. 42–49, 2016, doi: 10.1016/j.neucom.2015.01.107.

[69] N. Sadi-Ahmed, B. Kacha, H. Taleb, and M. Kedir-Talha, “Relevant Features Selection for Automatic Prediction of Preterm Deliveries from Pregnancy ElectroHysterograhic (EHG) records,” Journal of medical systems, vol. 41, no. 12, pp. 204–13, 2017, doi: 10.1007/s10916-017-0847-8.

[70] P. Ren, S. Yao, J. Li, P. A. Valdes-Sosa, and K. M. Kendrick, “Improved prediction of preterm delivery using empirical mode decomposition analysis of uterine electromyography signals,” PloS one, vol. 10, no. 7, pp. e0132116–e0132116, 2015, doi: 10.1371/journal.pone.0132116.

[71] I. O. Idowu et al., "Artificial Intelligence for Detecting Preterm Uterine Activity in Gynecology and Obstetric Care," 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, 2015, pp. 215-220, doi: 10.1109/CIT/IUCC/DASC/PICOM.2015.31.

[72] A. Smrdel and F. Jager, “Separating sets of term and pre-term uterine EMG records,” Physiological measurement, vol. 36, no. 2, pp. 341–355, 2015, doi: 10.1088/0967-3334/36/2/341.

[73] Naeem SM, Seddik AF, Eldosoky MA. New technique based on uterine electromyography nonlinearity for preterm delivery detection. Journal of Engineering and Technology Research. 2014 Nov 3;6(7):107-14.

[74] A. Hussain, P. Fergus, H. Al-Askar, D. Al-Jumeily, and F. Jager, “Dynamic neural network architecture inspired by the immune algorithm to predict preterm deliveries in pregnant women,” Neurocomputing (Amsterdam), vol. 151, no. 3, pp. 963–974, 2015, doi: 10.1016/j.neucom.2014.03.087.

[75] D. Alamedine, A. Diab, C. Muszynski, B. Karlsson, M. Khalil, and C. Marque, “Selection algorithm for parameters to characterize uterine EHG signals for the detection of preterm labor,” Signal, Image and Video Processing, vol. 8, no. 6, pp. 1169–1178, 2014, doi: 10.1007/s11760-014-0655-2.

[76] A. Lemancewicz et al., “Early diagnosis of threatened premature labor by electrohysterographic recordings – The use of digital signal

processing," Biocybernetics and biomedical engineering, vol. 36, no. 1, pp. 302–307, 2015, doi: 10.1016/j.bbe.2015.11.005.

[77] M. Borowska, E. Brzozowska, P. Kuć, E. Oczeretko, R. Mosdorf, and P. Laudański, "Identification of preterm birth based on RQA analysis of electrohysterograms," Computer methods and programs in biomedicine, vol. 153, pp. 227–236, 2018, doi: 10.1016/j.cmpb.2017.10.018.

[78] Sim S, Ryou H, Kim H, Han J, Park K. Evaluation of electrohysterogram feature extraction to classify the preterm and term delivery groups. In Goh J, editor, The 15th International Conference on Biomedical Engineering, ICBME 2013. Springer Verlag. 2014. p. 675-678. (IFMBE Proceedings). https://doi.org/10.1007/978-3-319-02913-9_172

[79] S. A. P, K. Subramaniam, and N. V. Iqbal, "A review of significant researches on prediction of preterm birth using uterine electromyogram signal," Future generation computer systems, vol. 98, pp. 135–143, 2019, doi: 10.1016/j.future.2018.10.033.

[80] Jackson, G.M.J., Ludmir, J. & Bader, T., 1992. The Accuracy of Digital Examination and Ultrasound in the Evaluation of Cervical Length. Obstetrics & Gynecology, 79(2), pp.214–218.

[81] E. Bhattacharjee and A. Maitra, "Spontaneous preterm birth: the underpinnings in the maternal and fetal genomes," Npj genomic medicine, vol. 6, no. 1, pp. 43–43, 2021, doi: 10.1038/s41525-021-00209-5.

[82] Goldenberg, Robert L, Prof, Culhane, Jennifer F, PhD, Iams, Jay D, Prof, and Romero, Roberto, Prof, "Epidemiology and causes of preterm birth," The Lancet (British edition), vol. 371, no. 9606, pp. 75–84, 2008, doi: 10.1016/S0140-6736(08)60074-4.

[83] J.-M. Moutquin, "Classification and heterogeneity of preterm birth," BJOG : an international journal of obstetrics and gynaecology, vol. 110, no. s20, pp. 30–33, 2003, doi: 10.1046/j.1471-0528.2003.00021.x.

[84] R. P. Jokhi, V. V. Ghule, B. H. Brown, and D. O. C. Anumba, "Reproducibility and repeatability of measuring the electrical impedance of the pregnant human cervix-the effect of probe size and applied pressure," 2009.

[85] M. Grossi and B. Riccò, “Electrical impedance spectroscopy (EIS) for biological analysis and food characterization: A review,” Journal of sensors and sensor systems, vol. 6, no. 2, pp. 303–325, 2017, doi: 10.5194/jsss-6-303-2017.

[86] D. O. C. Anumba, V. Stern, J. T. Healey, S. Dixon, and B. H. Brown, “Value of cervical electrical impedance spectroscopy to predict spontaneous preterm delivery in asymptomatic women: the ECCLIPPx prospective cohort study,” Ultrasound in obstetrics & gynecology, vol. 58, no. 2, pp. 293–302, 2021, doi: 10.1002/uog.22180.

[87] P. Li, P. E. Highfield, Z.-Q. Lang, and D. Kell, “Cervical cancer prognosis and diagnosis using electrical impedance spectroscopy,” Journal of electrical bioimpedance, vol. 12, no. 1, pp. 153–162, 2021, doi: 10.2478/joeb-2021-0018.

[88] S. V. Gandhi, D. Walker, P. Milnes, S. Mukherjee, B. H. Brown, and D. O. C. Anumba, “Electrical impedance spectroscopy of the cervix in non-pregnant and pregnant women,” European journal of obstetrics & gynecology and reproductive biology, vol. 129, no. 2, pp. 145–149, 2006, doi: 10.1016/j.ejogrb.2005.12.029.

[89] V. Stern et al., “PLD.26 The ECCLIPPx study: Electrical Impedance Prediction of Preterm Birth,” Archives of disease in childhood. Fetal and neonatal edition, vol. 99, no. Suppl 1, pp. A113–A116, 2014, doi: 10.1136/archdischild-2014-306576.327.

[90] S. M. Moqadam, P. K. Grewal, Z. Haeri, P. A. Ingledew, K. Kohli, and F. Golnaraghi, “Cancer detection based on electrical impedance spectroscopy: A clinical study,” Journal of electrical bioimpedance, vol. 9, no. 1, pp. 17–23, 2018, doi: 10.2478/joeb-2018-0004.

[91] G. P. Stafford et al., “Spontaneous preterm birth is associated with differential expression of vaginal metabolites by lactobacilli-dominated microflora,” Frontiers in physiology, vol. 8, pp. 615–615, 2017, doi: 10.3389/fphys.2017.00615.

[92] P. P. Tarjan and R. McFee, “Electrodeless Measurements of the Effective Resistivity of the Human Torso and Head by Magnetic Induction,” IEEE transactions on biomedical engineering, vol. BME-15, no. 4, pp. 266–278, 1968, doi: 10.1109/TBME.1968.4502577.

[93] A. Richer and A. Adler, "Eddy Current Based Flexible Sensor for Contactless Measurement of Breathing," 2005 IEEE Instrumentationand Measurement Technology Conference Proceedings, 2005, pp. 257-260, doi: 10.1109/IMTC.2005.1604112.

[94] A. Barai, S. Watson, H. Griffiths, and R. Patz, “Magnetic induction spectroscopy: non-contact measurement of the electrical conductivity spectra of biological samples,” Measurement science & technology, vol. 23, no. 8, pp. 85501–1–11, 2012, doi: 10.1088/0957-0233/23/8/085501.

[95] N. G. GENÇER and M. N. TEK, “Imaging tissue conductivity via contactless measurements: A feasibility study,” Elektrik : Turkish journal of electrical engineering & computer sciences, vol. 6, no. 3, pp. 183–200, 1998.

[96] H. Scharfetter, P. Riu, M. Populo, and J. Rosell, “Sensitivity maps for low-contrast perturbations within conducting background in magnetic induction tomography,” Physiological measurement, vol. 23, no. 1, pp. 195–202, 2002, doi: 10.1088/0967-3334/23/1/320.

[97] J.-Y. Wang, T. Healey, A. Barker, B. Brown, C. Monk, and D. Anumba, “Magnetic induction spectroscopy (MIS)-probe design for cervical tissue measurements,” Physiological measurement, vol. 38, no. 5, pp. 729–744, 2017, doi: 10.1088/1361-6579/aa6b4e.

[98] H. Scharfetter, R. Casanas, and J. Rosell, “Biological tissue characterization by magnetic induction spectroscopy (MIS): requirements and limitations,” IEEE transactions on biomedical engineering, vol. 50, no. 7, pp. 870–880, 2003, doi: 10.1109/TBME.2003.813533.

[99] H. Wei and A. Wilkinson, "Design of a Sensor Coil and Measurement Electronics for Magnetic Induction Tomography", IEEE Transactions on Instrumentation and Measurement, vol. 60, no. 12, pp. 3853-3859, 2011. Available: 10.1109/tim.2011.2147590.

[100] A. Koivu and M. Sairanen, “Predicting risk of stillbirth and preterm pregnancies with machine learning,” Health information science and systems, vol. 8, no. 1, pp. 14–14, 2020, doi: 10.1007/s13755-020-00105-9.

[101] C. Gao, S. Osmundson, D. R. Velez Edwards, G. P. Jackson, B. A. Malin, and Y. Chen, "Deep learning predicts extreme preterm birth from electronic health records," Journal of biomedical informatics, vol. 100, pp. 103334–103334, 2019, doi: 10.1016/j.jbi.2019.103334.

[102] Shechtman, O, "The Coefficient of Variation as an Index of Measurement Reliability", In: Doi, S., Williams, G. (eds) Methods of Clinical Epidemiology. Springer Series on Epidemiology and Public Health. Springer, Berlin, Heidelberg. 2013, doi.org/10.1007/978-3-642-37131-8_4

[103] D J Gougeon, "Logistic regression: from introductory to advanced concepts and applications," Choice, vol. 47, no. 1. American Library Association dba CHOICE, Middletown, p. 153, 2009.

[104] F. E. Harrell, Regression modeling strategies : with applications to linear models, logistic and ordinal regression and survival analysis, Second edition. Cham: Springer, 2015.

[105] J. R. Ayala Solares, H.-L. Wei, and S. A. Billings, "A novel logistic-NARX model as a classifier for dynamic binary classification," Neural computing & applications, vol. 31, no. 1, pp. 11–25, 2017, doi: 10.1007/s00521-017-2976-x.

[106] R. Tate, "Correlation Between a Discrete and a Continuous Variable. Point-Biserial Correlation," The Annals of mathematical statistics, vol. 25, no. 3, pp. 603–607, 1954, doi: 10.1214/aoms/1177728730.

[107] Hua-Liang Wei and S. . Billings, "Feature Subset Selection and Ranking for Data Dimensionality Reduction," IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 1, pp. 162–166, 2007, doi: 10.1109/TPAMI.2007.250607.

[108] L. Liu, Q. Wang, E. Adeli, L. Zhang, H. Zhang, and D. Shen, "Feature Selection Based on Iterative Canonical Correlation Analysis for Automatic Diagnosis of Parkinson's Disease," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9901, pp. 1–8, 2016, doi: 10.1007/978-3-319-46723-8_1.

[109] K. Yoshida, J. Yoshimoto, and K. Doya, “Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data,” BMC bioinformatics, vol. 18, no. 1, pp. 108–108, 2017, doi: 10.1186/s12859-017-1543-x.

[110] Long Zhang, Kang Li, Er-Wei Bai, and G. W. Irwin, “Two-Stage Orthogonal Least Squares Methods for Neural Network Construction,” IEEE transaction on neural networks and learning systems, vol. 26, no. 8, pp. 1608–1621, 2015, doi: 10.1109/TNNLS.2014.2346399.

[111] K. Li, J.-X. Peng, and E.-W. Bai, “A two-stage algorithm for identification of nonlinear dynamic systems,” Automatica (Oxford), vol. 42, no. 7, pp. 1189–1197, 2006, doi: 10.1016/j.automatica.2006.03.004.

[112] H. Han, W.-Y. Wang, and B.-H. Mao, “Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning,” in Advances in Intelligent Computing, 2005, vol. 3644, no. I, pp. 878–887. doi: 10.1007/11538059_91.

[113] H. A. Watson, J. Carter, P. T. Seed, R. M. Tribe, and A. H. Shennan, “The QUiPP App: a safe alternative to a treat-all strategy for threatened preterm labor,” Ultrasound in obstetrics & gynecology, vol. 50, no. 3, pp. 342–346, 2017, doi: 10.1002/uog.17499.

[114] H. A. Watson et al., “EQUIPTT: The Evaluation of the QUiPP app for Triage and Transfer protocol for a cluster randomised trial to evaluate the impact of the QUiPP app on inappropriate management for threatened preterm labour,” BMC pregnancy and childbirth, vol. 19, no. 1, pp. 68–68, 2019, doi: 10.1186/s12884-019-2210-1.

[115] V. Stern and D. Anumba, “Potential incorporation of novel cervical impedance spectroscopy assessment into existing clinical algorithms for predicting preterm birth,” European journal of obstetrics & gynecology and reproductive biology, vol. 256, pp. 523–524, 2021, doi: 10.1016/j.ejogrb.2020.11.019.

[116] A. Koivu and M. Sairanen, “Predicting risk of stillbirth and preterm pregnancies with machine learning,” Health information science and systems, vol. 8, no. 1, pp. 14–14, 2020, doi: 10.1007/s13755-020-00105-9.

[117] L. J. E. Meertens et al., “Prediction models for the risk of spontaneous preterm birth based on maternal characteristics: a systematic review and independent external

validation," Acta obstetricia et gynecologica Scandinavica, vol. 97, no. 8, pp. 907–920, 2018, doi: 10.1111/aogs.13358.

[118] Hands-on ensemble learning with R : a beginner's guide to combining the power of machine learning algorithms using ensemble techniques, 1st edition. Birmingham; Mumbai: Packt Publishing, 2018.

[119] S. Dong et al., "Using Undersampling with Ensemble Learning to Identify Factors Contributing to Preterm Birth," 2020.

[120] C. Leon et al., "Evaluation of Maturation in Preterm Infants Through an Ensemble Machine Learning Algorithm Using Physiological Signals," IEEE journal of biomedical and health informatics, vol. 26, no. 1, pp. 400–410, 2022, doi: 10.1109/JBHI.2021.3093096.

[121] S. A. Billings, Nonlinear system identification [electronic resource] NARMAX methods in the time, frequency, and spatio-temporal domains. Chichester, England: Wiley, 2013.

[122] H. L. Wei, S. A. Billings, and J. Liu, "Term and variable selection for non-linear system identification," International journal of control, vol. 77, no. 1, pp. 86–110, 2004, doi: 10.1080/00207170310001639640.