Automatic Fetal Distress Assessment During Labor Based on Modal and Parametrical Analysis of the Cardiotocographic Recording

DISSERTATION

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

angenommen durch die Fakultät für Informatik

der Otto-von-Guericke-Universität Magdeburg

von Dipl.-Ing. Patricio Fuentealba Ortiz

geb. am 11.11.1988 in Osorno, Chile

Gutachterinnen/Gutachter:

Prof. Dr. Frank Ortmeier
Prof. Dr. Francisco Guerra
Prof. Dr. Bernhard Preim
Prof. Dr. Dirk Hoyer

Magdeburg, den 16.11.2020

Fuentealba Ortiz, Patricio:
Automatic Fetal Distress Assessment During Labor Based on Modal and Parametrical Analysis of the Cardiotocographic Recording
DISSERTATION
zur Erlangung des akademischen Grades Doktoringenieur (Dr.-Ing.)
Otto-von-Guericke Universität, Magdeburg, 2020.

Zusammenfassung

Das Hauptziel der fötalen Überwachung während der Wehen ist die rechtzeitige Identifizierung potenziell azidotischer Föten. Ein gutes Verständnis des fötalen Zustands hilft den Ärzten, negative fötale Ergebnisse aufgrund von Sauerstoffmangel zu verhindern und gleichzeitig unnötige Eingriffe zu vermeiden. Dieses Verfahren zur Beurteilung des fötalen Wohlbefindens wird üblicherweise mit Hilfe eines Kardiotokographen (CTG) durchgeführt, der die gemeinsame Aufzeichnung der fötalen Herzfrequenz (FHR) und der Uteruskontraktion (UC) ermöglicht. Gegenwärtig umfasst die Analyse der CTG-Daten die visuelle Interpretation verschiedener morphologischer Signalmuster auf der Grundlage vorgeschlagener medizinischer Richtlinien. Die CTG-Interpretation durch diese Methodik hat jedoch große Uneinigkeit zwischen den Beobachtern und eine geringe Spezifizität gezeigt, was zu einer schlechten Reproduzierbarkeit der Interpretation führt. Jüngste Fortschritte in der klinischen Forschung weisen darauf hin, dass jeder Fötus sein eigenes Kontrollverhalten hat, das auch von seinem Zustand abhängt. Demzufolge erfordert die korrekte Identifizierung einer Hypoxämie ein gutes Verständnis der fötalen Kompensationsmechanismen, die durch das autonome Nervensystem (ANS) moduliert werden. Die Haupthypothese dieser Arbeit ist, dass die durch das ANS modulierte FHR-Reaktion wichtige zeitvariante Merkmale beschreibt, deren Verhalten im Laufe der Zeit bei der Bewertung der fötalen Notlage helfen kann. Mit anderen Worten postulieren wir, dass die aus der ANS-Modulation resultierenden Schlag-zu-Schlag-FHR-Variationen signifikante Informationen über den Zustand des Fötus beinhalten können. In Anbetracht dieses Phänomens könnten konventionelle Methoden, die diese Merkmale nicht integrieren, nicht für eine korrekte CTG-Analyse geeignet sein, da sie nur eine Momentaufnahme des gesamten zeitvarianten Prozesses betrachten. In diesem Zusammenhang sind für die CTG-Analyse mehrere Ansätze vorgeschlagen worden, die auf zeitvariablen Signalverarbeitungsmethoden basieren. Sie konzentrieren sich jedoch hauptsächlich auf die fötale Reaktivität als Antwort auf ein UC-Ereignis, ohne die Untersuchung der dynamischen Veränderungen des FHR-Signals, die sich aus der ANS-Modulation im Laufe der Zeit ergeben, zu berücksichtigen. Der Hauptbeitrag dieser Arbeit beruht auf der Charakterisierung solcher Dynamiken durch die Verfolgung ihrer zeitvariablen Merkmale. Für diese Operation setzen wir im Wesentlichen zwei Signalverarbeitungsmethoden ein: Verbesserte Vollständige Ensemble-Empirische-Bandzerlegung mit adaptivem Rauschen (ICEEMDAN) und zeitvariable autoregressive (TV-AR) Modellierung. Die zentrale Idee ist, die Dynamik des FHR-Signals durch Zerlegung in intrinsische modale Funktionen (IMFs) zu untersuchen und dann ihre Merkmale sowohl im Zeitbereich als auch im Spektralbereich zu analysieren. Zuerst wird das FHR-Signal in eine endliche Anzahl von IMFs zerlegt. Dann wird für jede IMF das TV-AR-Spektrum berechnet, um die beteiligten Frequenzkomponenten zu verfolgen und einen möglichen Zusammenhang zwischen ihrem dynamischen Verhalten und der fötalen Notlage zu untersuchen. Der vorgeschlagene Ansatz ist auf der Grundlage realer CTG-Daten aus der frei zugänglichen CTU-UHB-Datenbank bewertet worden. Die erhaltenen Ergebnisse zeigen, dass es möglich ist, zeitvariant modulierte Dynamiken im FHR-Signal zu identifizieren und zu verfolgen, deren Verhalten signifikant (p-Wert $< .05$) vom Gesundheitszustand des Fötus abhängen kann. Ebenfalls könnten diese untersuchten Dynamiken mit den durch das ANS modulierten fötalen Kompensationsmechanismen in Verbindung gebracht werden. Zusammenfassend hat sich unsere vorgeschlagene Methodik, die auf der ICEEMDAN- und TV-AR-Spektralanalyse basiert, als vielversprechender Ansatz für die Abschätzung der fötalen Notlage während der Wehen erwiesen. Außerdem hat sie eine Leistungsverbesserung der CTG-Klassifikation im Vergleich zur konventionellen Analyse ermöglicht.

Schlüsselwörter— Kardiotokograph, Fötale Herzfrequenz, Bewertung des fötalen Wohlbefindens, Biomedizinische Signalverarbeitung, Spektralanalyse, Empirische Bandzerlegung, Zeitvariable autoregressive Modellierung, Klassifizierung von kardiotokographischen Aufzeichnungen, Automatisches Lernen

Abstract

The main aim of fetal surveillance during labor is to timely identify potential acidotic fetuses. A good understanding of the fetal condition helps clinicians to prevent adverse fetal outcomes due to oxygen insufficiency, and at the same time to avoid unnecessary interventions. The fetal welfare assessment procedure is commonly performed by using a cardiotocograph (CTG), which provides the joint recording of fetal heart rate (FHR) and uterine contraction (UC) data. Currently, CTG data analysis involves the visual interpretation of several morphological signal patterns based on proposed medical guidelines. However, CTG interpretation by this methodology has shown a wide intra- and inter-observer disagreement and low specificity, leading to a poor interpretation reproducibility. Recent advances in clinical research indicate that each fetus has its own control behavior, which also depends on its condition. Therefore, the correct identification of hypoxemia requires a good understanding of the fetal compensatory mechanisms modulated by the autonomic nervous system (ANS). The main hypothesis of this thesis work is that the FHR response modulated by the ANS describes important time-variant characteristics, whose behavior can help to assess fetal distress. In other words, we postulate that the beat-to-beat FHR variations resulting from the ANS modulation can involve significant information about the fetal condition. Considering this phenomenon, conventional methods that do not integrate these characteristics could not be appropriated for a correct CTG analysis, because they consider only a snapshot of the complete time-variant process. In this context, several approaches based on time-varying signal processing methods have been proposed for CTG analysis. However, they are mainly focused on fetal reactivity as a response to a UC event, without considering the study of the dynamical changes involved in the FHR signal resulting from the ANS modulation over time. The main contribution of this work consists of characterizing such dynamics by tracking their time-variant characteristics. For this operation, we employ, essentially, two signal processing methods: the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) and time-varying autoregressive (TV-AR) modeling. The core idea is to investigate the dynamics involved in the FHR signal by decomposing it into intrinsic mode functions (IMFs), and then analyze their characteristics in both the time-domain and the spectral-domain. First, the FHR signal is decomposed into a finite number of IMFs. Then, for each IMF, the TV-AR spectrum is computed in order to track the involved frequency components and investigate a potential connection between their dynamical behavior and fetal distress. The proposed approach has been evaluated on real CTG data extracted from the open-access CTU-UHB database. Obtained results show that it is possible to identify and track time-variant modulated dynamics in the FHR signal, whose behavior can depend significantly (p-value $< .05$) on the fetal health condition. Likewise, the studied dynamics might be linked to the fetal compensatory mechanisms modulated by the ANS. Conclusively, the proposed methodology based on ICEEMDAN and TV-AR modeling has shown to be a promising approach for fetal distress estimation during labor. Besides, it has allowed improving the CTG classification performance compared with the conventional analysis.

***Keywords*—** Cardiotocograph, Fetal heart rate, Fetal welfare assessment, Biomedical signal processing, Spectral analysis, Empirical mode decomposition, Time-varying autoregressive modeling, Cardiotocographic recording classification, Machine learning

Resumen

El objetivo principal de la vigilancia fetal durante el trabajo de parto es identificar oportunamente fetos potencialmente acidóticos. Una buena comprensión de la condición del feto ayuda a los médicos a prevenir resultados fetales adversos debido a insuficiencia de oxígeno, y al mismo tiempo a evitar intervenciones innecesarias. El procedimiento de evaluación del bienestar fetal se realiza comúnmente mediante el uso de un cardiotocógrafo (CTG), que proporciona el registro conjunto de los datos de la frecuencia cardíaca fetal (FHR) y la contracción uterina (UC). Actualmente, el análisis de datos CTG implica la interpretación visual de diferentes patrones morfológicos presentes en la señal basados en guías médicas propuestas. Sin embargo, la interpretación del CTG por esta metodología ha mostrado un amplio desacuerdo intra- e inter-observador y baja especificidad, lo que lleva a una mala reproducibilidad de la interpretación. Avances recientes en la investigación clínica indican que cada feto tiene su propio comportamiento de control, el cual también depende de su condición. Por lo tanto, la identificación correcta de hipoxemia requiere una buena comprensión de los mecanismos compensatorios del feto modulados por el sistema nervioso autónomo (ANS). La hipótesis principal de este trabajo de tesis es que la respuesta de la FHR modulada por el ANS describe características importantes variantes en el tiempo, cuyo comportamiento puede ayudar a evaluar el sufrimiento fetal. En otras palabras, postulamos que las variaciones de la FHR latido a latido que resultan de la modulación de ANS pueden involucrar información significativa acerca la condición fetal. Teniendo en cuenta este fenómeno, los métodos convencionales que no integran estas características podrían no ser apropiados para un correcto análisis del CTG, ya que consideran solo una instantánea del proceso completo variante en el tiempo. En este contexto, se han propuesto varios trabajos basados en métodos de procesamiento de señales variantes en el tiempo para el análisis del CTG. Sin embargo, éstos se centran principalmente en la reactividad fetal como respuesta a un evento de UC, sin considerar el estudio de los cambios dinámicos involucrados en la señal de la FHR como resultado de la modulación del ANS a lo largo del tiempo. La principal contribución de este trabajo consiste en caracterizar tales dinámicas a través del seguimiento de sus características variantes en el tiempo. Para esta operación, empleamos, esencialmente, dos métodos de procesamiento de señales: Descomposición Empírica en Modos por Conjuntos Completa con Ruido Adaptativo Mejorada (ICEEMDAN) y modelamiento autorregresivo variante en el tiempo (TV-AR). La idea central es investigar las dinámicas involucradas en la señal FHR descomponiéndola en funciones modales intrínsecas (IMFs), y luego analizar sus características tanto en el dominio del tiempo como en el dominio espectral. Primero, la señal FHR se descompone en un número finito de IMFs. Luego, para cada IMF, se calcula el espectro TV-AR para rastrear los componentes frecuenciales involucrados e investigar una posible conexión entre su comportamiento dinámico y el sufrimiento fetal. El enfoque propuesto se ha evaluado en datos CTG reales extraídos de la base de datos CTU-UHB de acceso abierto. Los resultados obtenidos muestran que es posible identificar y rastrear dinámicas moduladas variantes en el tiempo en la señal FHR, cuyo comportamiento puede depender significativamente (valor $p < .05$) de la condición de salud del feto. Del mismo modo, las dinámicas estudiadas podrían estar vinculadas a los mecanismos compensatorios del feto modulados por el ANS. En conclusión, la metodología propuesta basada en ICEEMDAN y modelamiento TV-AR ha mostrado ser un enfoque prometedor para la estimación del sufrimiento fetal durante el trabajo de parto. Además, esta ha permitido mejorar el rendimiento de la clasificación del CTG en comparación con el análisis convencional.

Palabras clave— Cardiotocografía, Frecuencia cardiaca fetal, Evaluación del bienestar fetal, Procesamiento de señales biomédicas, Análisis espectral, Descomposición modal empírica, Modelamiento autorregresivo variante en el tiempo, Clasificación de registros cardiotocográficos, Aprendizaje automático

Learning which features to select is a skill
developed by data scientists that usually only comes from months and
years of experience and can be more of an art than a science.

Alberto Artasanchez and Prateek Joshi, 2020, p. 67.

Acknowledgements

First and foremost, I would like to praise and thank God for all His blessings. I could never have done this thesis work without You, my Lord.

I want to thank my thesis supervisors. Firstly, thank Prof. Dr. Frank Ortmeier for providing me the opportunity to work at the "Chair of Software Engineering," Otto-von-Guericke University, Magdeburg, and for his guidance and assistance throughout this work. Likewise, I want to thank my adviser Dr. Alfredo Illanes for his help, time, assistance, patience, and motivation during all the process involved in this thesis work. It is not often in life to meet a person with such scientific capability and human qualities.

My special thanks to my wife for her love and support, and thanks to my family and friends for everything.

This thesis work has been supported by the Republic of Chile through the National Commission for Scientific and Technological Research CONICYT, Chilean National Scholarship Program for Graduate Studies.

I dedicate this thesis work to my beloved wife María, my parents Claudio and Francisca, my brethren in Christ, and all my friends.

Thanks a lot!

Patricio Fuentealba Ortiz

Magdeburg, November, 2020.

Agradecimientos

Ante todo, me gustaría alabar y agradecer a Dios por todas sus bendiciones. Nunca podría haber hecho este trabajo de tesis sin ti, mi Señor.

Quiero agradecer a mis supervisores de tesis. En primer lugar, agradezco a Prof. Dr. Frank Ortmeier por brindarme la oportunidad de trabajar en el equipo "Chair of Software Engineering," Universidad Otto-von-Guericke, Magdeburgo, y por su orientación y asistencia a lo largo de este trabajo. Asimismo, quiero agradecer a mi supervisor, Dr. Alfredo Illanes, por su ayuda, tiempo, asistencia, paciencia y motivación durante todo el proceso involucrado en este trabajo de tesis. No es frecuente en la vida conocer a una persona con tal capacidad científica y cualidades humanas.

Mi agradecimiento especial a mi esposa por su amor y apoyo, y gracias a mi familia y amigos por todo.

Este trabajo de tesis ha sido financiado por la República de Chile a través de la Comisión Nacional de Investigación Científica y Tecnológica CONICYT, Programa Nacional de Becas para Estudios de Posgrado de Chile.

Dedico este trabajo de tesis a mi amada esposa María, a mis padres Claudio y Francisca, a mis hermanos en Cristo, y a todos mis amigos.

¡Muchas gracias!

Patricio Fuentealba Ortiz

Magdeburgo, Noviembre, 2020.

Contents

1 Introduction and Motivation

1.1 Motivation

Labor and delivery corresponds to the ending period of the pregnancy, which concludes with the expulsion of the neonate. It is a stressful procedure for the fetus as well as for the mother. During this step, a fetus can repeatedly suffer from decreased oxygen supply, which is normal, but fetuses with weakened defense mechanisms could develop metabolic acidosis. As a consequence, it can lead to neuro-development disability, cerebral palsy, or in some cases, even death resulting from a prolonged oxygen insufficiency. During labor, in order to prevent potential fetal adverse outcomes, and at the same time avoid excessive obstetric intervention, it is necessary a good understanding of the fetal status (Nageotte, 2015). In this context, there are different methodologies for fetal welfare assessment, such as cardiotocography (CTG), fetal blood sampling, pulse oximetry, fetal electrocardiogram, fetal phonocardiography, and fetal magnetocardiography.

In current clinical practice, the most used method for fetal monitoring is the non-invasive CTG, which provides continuous information of the fetal heart rate (FHR) synchronized with the maternal uterine contraction (UC) activity over time (Nageotte, 2015). In general, clinicians base the decision by observing and assessing patterns in the FHR signal, such as baseline rate, variability, decelerations, accelerations as well as the UC stimuli. Nowadays, these morphological patterns and their properties are defined in CTG guidelines proposed for fetal monitoring (American College of Obstetricians and Gynecologists and others, 2010; Artal and O'toole, 2003; Ayres-de Campos and Bernardes, 2010; Ayres-de Campos et al., 2015; Macones et al., 2008; Parer and Ikeda, 2007; Rooth et al., 1987; Royal College of Obstetricians and Gynaecologists (Great Britain). Clinical Effectiveness Support Unit, 2001; Spilka et al., 2014b), which systematically assist practitioners in assessing the fetal status based on morphological CTG patterns and the manner that these patterns are related to the fetal health condition.

However, the proposed CTG guidelines have been surrounded by great controversies, since they lack consensus in many key aspects. The absence of an objective definition of some FHR patterns and the disparity between different guidelines are often pointed out in the literature. Likewise, concerning CTG standardization, definitions, and recommendations for the intrapartum management can differ from one country to another or even from one

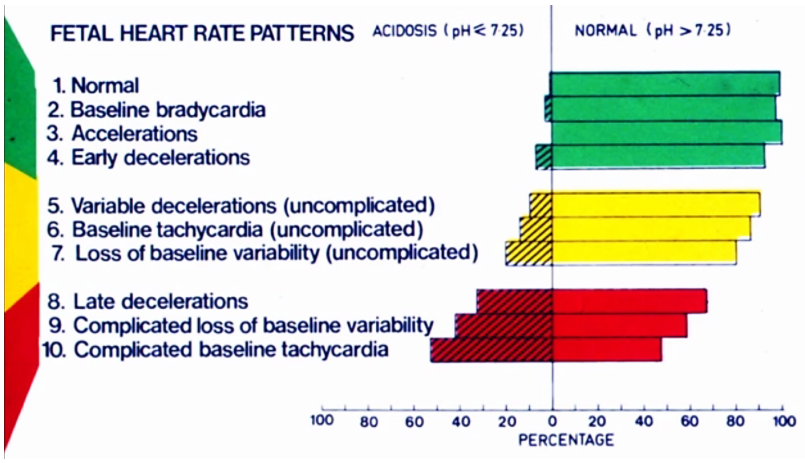


Figure 1.1: Fetal heart rate (FHR) signal patterns assessed by clinicians compared to the outcome evaluation based on artery pH (pH) (Beard et al., 1971).

clinical center to another (Parer and Ikeda, 2007; Santo et al., 2017; Ugwumadu, 2014). In consequence, the interpretation of CTG has shown to suffer from wide intra- and inter-observer disagreement, leading to a poor interpretation reproducibility (Ayres-de Campos and Bernardes, 2010; Devoe, 2016; Nageotte, 2015; Santo and Ayres-de Campos, 2012; Sartwelle et al., 2019). According to the literature, most of the CTG abnormalities do not cause acidosis, i.e., during labor, most of the CTG patterns considered as abnormal are related to a healthy fetus. This phenomenon is represented by the percentage of specificity of each FHR pattern, as presented in Fig. 1.1 (Beard et al., 1971). Here, we can observe that less than 50% of late decelerations or complicated loss of baseline variability (considered as abnormal patterns) lead to a pathological fetus. Likewise, recent works (Ayres-de Campos and Bernardes, 1999; Ayres-de Campos et al., 2004; Bernardes et al., 1997; Hruban et al., 2015; Spilka et al., 2014b) have demonstrated that observer disagreements are mainly related to the interpretation of FHR decelerations and variability.

Besides, an often debated concern is that the CTG interpretation problems can be partly one of the leading causes of the significant increase of cesarean section (CS) in the last twenty years (Alfirevic et al., 2017; Gibbons et al., 2010; Visser et al., 2018). In particular, a pathological CTG and auscultation are the second main documented indicator for CS in Germany (20.8%) after CS due to previous cesarean (23.6%) (Kolip et al., 2012). Hypothetically, CTG interpretation problems can lead to making decisions that are the less risky for the fetus, so in cases of doubts in interpretation, CS is decided as an option. This bias in the decision increases with the lack of experience of obstetricians in dealing with complicated and spontaneous deliveries. In consequence, in developed countries, even if cerebral palsy prevalence has remained nearly constant, the CS has dramatically increased, as shown in Fig. 1.2 (Clark and Hankins, 2003).

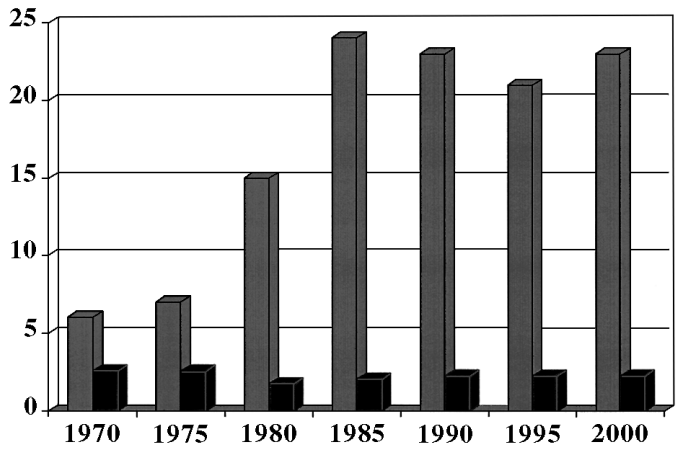


Figure 1.2: Prevalence of cerebral palsy (black bars) versus cesarean section rate (dark gray bars). This information is based on data gathered from Sweden, Australia, Canada, Scotland, Denmark, England, United States, Norway, and Ireland, as presented in Clark and Hankins (2003).

1.2 Research questions, hypothesis, and goals

On the one hand, all of the above leads us to consider whether the CTG data involve sufficient information for a correct fetal welfare assessment. Likewise, whether the conventional analysis, based on morphological signal patterns, makes that this technology has significant intra- and inter-observer disagreements and very low specificity. On the other hand, as will be explained in detail in Chapter 3, the published literature related to CTG analysis reveals that, despite the fact that a considerable number of approaches have been proposed, the results are not sufficient to be applied in clinical practice so far (Alfirevic et al., 2017; Haran and Everett, 2017). This phenomenon can be attributed to a low level of reliability in fetal assessment represented by the metrics of specificity and sensitivity achieved by the proposed approaches. Consequently, it brings up to the question of whether the research in fetal surveillance based on CTG analysis, as performed nowadays, takes into account all the essential fetal physiological characteristics involved in the labor process.

In this context, recent advances in clinical research suggest that a correct assessment of fetal hypoxemia demands a good understanding of the compensatory mechanisms modulated by the fetal autonomic nervous system (ANS). Each fetus has a particular behavior and control, and its condition depends on how the fetus is compensating and defending itself over time after a perceived oxygen insufficiency (Garabedian et al., 2017; Ugwumadu, 2014). At the same time, experienced clinicians attempt to consider the temporal evolution when interpreting CTGs by considering the variation of a specific FHR characteristic throughout a particular temporal window (e.g., 30 minutes) in order to assess the CTG patterns as described in guidelines (Ayres-de Campos et al., 2015). On the one hand, the FHR decelerations are one of the main patterns represented in the FHR signal that result from the fetal compensatory mechanism activity. These patterns correspond to FHR variations

produced as a consequence of the ANS modulation after a perceived oxygen insufficiency episode (Freeman et al., 2012). However, they are considered as one of the most significant and challenging patterns to interpret by the clinical staff (Hruban et al., 2015; Spilka et al., 2014b). On the other hand, the risk of fetal condition to evolve from one health state to another, frequently change in the course of labor (Coletta et al., 2012; Devoe, 2016; Parer and Ikeda, 2007). In the sequel, as a global concept, we denote these time-varying physiological characteristics involved in the fetal behavior during labor as the short-term *progressive fetal response*. In other words, we can define the progressive fetal response as progressive dynamical changes involved in the behavior of the ANS modulation.

With this concept in mind, the CTG classification into categories by using conventional clinical guidelines, as usually performed nowadays, could be not helpful to assess the fetal condition, because the interpretation is based only on a snapshot of the complete non-stationary process. Besides, methods that do not integrate the concept of progressive fetal response could not be appropriate for a correct CTG analysis.

In summary, considering all the information explained above, and based on preliminary results that we have published in peer-reviewed scientific journals and presented in international conferences (see Section 1.4), we firmly believe that the time-variant fetal response as a result of the fetal ANS modulation is an important phenomenon to be considered for fetal distress assessment during labor.

Interestingly, although clinical research in fetal monitoring has suggested assessing this physiological phenomenon during labor, it has not been considered in CTG guidelines or within a computer-based CTG signal analysis so far. Indeed, the main idea of this thesis work is to investigate these time-variant characteristics in the CTG recording, and thereby integrate them into a computer-based analysis.

As a result, the main research questions that naturally arise are the following:

- **Is it possible to identify and track modulated time-variant dynamics involved in the FHR signal that are related to the fetal health condition?**
- **Could the dynamics involved in FHR decelerations, analyzed independently from the complete signal, improve the automatic CTG signal analysis?**
- **Are these studied FHR signal dynamics linked to the fetal compensatory mechanisms modulated by the ANS?**
- **Could this approach based on a time-variant analysis of the fetal response improve the automatic computer-based CTG classification compared with the conventional CTG analysis?**

The main hypothesis of this thesis work is that the FHR response modulated by the ANS describes important time-variant characteristics, whose behavior over time can help to assess fetal distress. In other words, we postulate that FHR variations resulting from the

complex modulation of the ANS vary progressively in time and describe different dynamics associated with the fetal response, whose changes and their progression over time can be related to the fetal condition. Therefore, this approach could improve the interpretation and subsequent classification of CTG recordings during labor. The objectives of this thesis work are summarized as follows:

General objective

- Estimate the fetal condition during labor based on the characterization of the progressive fetal response described by the time-varying FHR dynamics as a result of the compensatory mechanism activity modulated by the ANS.

Specific objectives

- Identify and investigate the FHR signal dynamics involved in the fetal response as a result of the ANS modulation.
- Based on those dynamics, define and extract time- and spectral-based features that integrate the concept of time-variant fetal response.
- Establish the evaluation methodology to investigate the physiological connection of the extracted information with the fetal health condition.
- Evaluate the approach based on the performance of the proposed features in CTG classification.

1.3 Main contribution of the thesis

Recent clinical research indicates that the progressive fetal response is a fundamental characteristic to be considered in CTG assessment. However, most of the approaches for CTG analysis proposed in the area of signal processing focus on extracting information in different domains based only on a snapshot of the complete non-stationary process, without taking into account the concept of progressive fetal response. Considering that the ANS activity is an important phenomenon related to the fetal compensatory mechanism, this work focuses on studying this phenomenon based on the progressive dynamical changes reflected in the FHR signal as a result of this modulation. Therefore, the originality of this work lies in the fact that we propose a new methodology for CTG analysis based on the characterization of the time-variant FHR dynamics resulting from the ANS modulation. Indeed, the analysis of CTG from a progressive perspective becomes to be an innovative method to assess fetal distress during labor and delivery, whose concept has not been previously investigated by a computer-based approach. For more detailed information concerning the fetal compensatory mechanism and proposed solutions for the CTG analysis, see Chapter 2 and 3, respectively.

The main difference of this work with the time-varying CTG analysis already proposed in the literature is the generation of *modal-spectral* features, which aim to characterize the progressive fetal response over time. These features allow a better understanding of the fetal physiological characteristics involved in the ANS modulation and integrate them into a computer-based CTG analysis.

This proposed methodology opens perspectives for its implementation in real scenarios in order to improve the reliability of CTG analysis and make its interpretation a more straightforward task for clinicians. In consequence, the most important benefit is to get a better performance in clinical practice concerning the assistance and management of the patient during labor by a reliable and timely fetal evaluation. Thereby, it would be possible to standardize the fetal welfare assessment procedure and its protocol in clinical practice, which is the ultimate goal (currently no achieved) of most of the proposed research approaches related to fetal monitoring.

1.4 List of publications

Part of the content presented in this thesis work has been already published in peer-reviewed scientific journals and presented in international conferences. The list of the publications directly related to this thesis work are the following:

Journal publications

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2019c). Cardiotocographic signal feature extraction through ceemdan and time-varying autoregressive spectral-based analysis for fetal welfare assessment. *IEEE Access*

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2019e). Independent analysis of decelerations and resting periods through ceemdan and spectral-based feature extraction improves cardiotocographic assessment. *Applied Sciences*, 9(24):5421

Conference publications

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2017a). Analysis of the foetal heart rate in cardiotocographic recordings through a progressive characterization of decelerations. *Current Directions in Biomedical Engineering*, 3(2):423–427

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2017b). Progressive fetal distress estimation by characterization of fetal heart rate decelerations response based on signal variability in cardiotocographic recordings. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2018a). Foetal heart rate signal spectral analysis by using time-varying autoregressive modelling. *Current Directions in Biomedical Engineering*, 4(1):579–582

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2018b). Spectral-based analysis of progressive dynamical changes in the fetal heart rate signal during labor by using empirical mode decomposition. In *2018 Computing in Cardiology (CinC)*, pages 1–4. IEEE

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2019a). Automatic fetal distress assessment during labor based on a progressive analysis of the cardiotocographic recording. In *1.Doktorandentagung am 28.& 29. Januar, Magdeburg, Germany, 2019*. Faculty of Computer Science, Otto-von-Guericke University

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2019b). Cardiotocograph data classification improvement by using empirical mode decomposition. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2019d). Foetal heart rate assessment by empirical mode decomposition and spectral analysis. *Current Directions in Biomedical Engineering*, 5(1):381–383

- Fuentealba, P., Illanes, A., and Ortmeier, F. (2019f). A study on the classification performance of cardiotocographic data vs. class formation criteria. In *2019 International Student Conference (ISC) IEEE EMBS, Magdeburg, Germany*. IEEE

2

Fetal welfare assessment

This chapter first provides an insight into the fetal control mechanisms activity as a response to oxygen insufficiency during labor (Section 2.1). Then, Section 2.2 presents the main diagnostic tools used for fetal monitoring, which can be classified into two main groups: invasive (Fetal Blood Sampling, Pulse Oximetry, and Invasive Fetal Electrocardiogram) and non-invasive (Fetal Phonocardiography, Non-invasive Fetal Electrocardiogram, Fetal Magnetocardiography, and CTG) methods. Likewise, this section presents the procedure and management performed as the outcome evaluation. Finally, in Section 2.3, we discuss about fetal welfare assessment by using cardiotocographic data, which is the main issue of analysis in this thesis work.

2.1 Fetal control mechanisms during labor

Labor and delivery is a very stressful period for both the mother and the fetus, who is affected by the behavior of the mother and internal environment conditions. The way that the fetus responds to different stimuli could provide valuable information about its condition. During this process, a fetus can suffer from intermittent events of decreased oxygen insufficiency, which is a natural phenomenon well tolerated by healthy fetuses. However, fetuses with weakened defense mechanisms could develop a *hypoxemia*, *hypoxia*, or *asphyxia*, which can lead to severe consequences such as cell dysfunction, organ failure, neuro-development disability, cerebral palsy, or in some cases, even death (Freeman et al., 2012).

Hypoxemia corresponds to an initial step of oxygen deficiency in arterial blood. In this step, the peripheral tissues and central organs are intact, and enough oxygen is provided to maintain aerobic metabolism. However, a fetus presenting hypoxemia during labor has less ability to react adequately because of the restriction of energetic reserves.

Hypoxia represents a second step of oxygen deficiency when the peripheral body tissues are affected. This phenomenon describes a reduction in oxygen supply in relation to the oxygen demand of the tissues. In order to ensure the aerobic metabolism in central organs, the blood flows mainly to such organs. This situation may last for hours without causing any severe damage to the fetus.

Asphyxia is the most critical step because the oxygen is depleted, and the main fetus's organs such as the heart, lungs, liver, gut, and kidneys make use of anaerobic metabolism. During this step, hypoxic damage may occur in these organs, but brain damage is the most significant consequence that affects the fetus. The final step of asphyxia involves the brain and heart failure, and it may produce irreversible damage.

A reliable assessment and diagnosis in fetuses during labor is of significant importance since it allows obtaining information about the fetal condition in order to prevent all the consequences explained above (Romano et al., 2016b).

According to the literature (Ugwumadu, 2014), a fetus presents particular FHR dynamics over time that can involve an essential relation with its condition. Indeed, as well evidenced in medical research for adults, heart rate dynamics are modulated by the ANS by the cardioregulatory centers located in the brain stem, which regulate the sympathetic and parasympathetic stimuli to the heart (see Fig. 2.1) (Ferreira Jr and Zanesco, 2016). As represented in Fig. 2.2, both systems have opposite effects on heart activity. The sympathetic system prepares the body for intense activity after a perceived threat. In contrast, the parasympathetic relaxes the body and inhibits many energy functions and control the heart response while at rest. Sympathetic and parasympathetic impulses, originated in the brain stem, are carried to the heart via the cervical sympathetic fibers and over the vagus nerves, respectively (Freeman et al., 2012). The cardiovascular centers receive information from a set of physiological sensors, which correspond to afferent nerve endings called chemoreceptors and baroreceptors.

On the one hand, chemoreceptors, subdivided into central and peripheral, sense the concentration of blood oxygen. Central chemoreceptors locate in the brain stem, whereas peripheral chemoreceptors locate at the aortic arch and the carotid. On the other hand, the baroreceptors, located mainly in the carotid sinuses and aortic arch, sense the arterial stretch proportional to changes in arterial pressure (Garabedian et al., 2017; Wehrwein and Joyner, 2013). When receptors detect a change of blood artery pH (pH) or artery pressure, the cardiovascular centers modulate the FHR by sending a nerve impulse. For example, when the chemoreceptors detect a reduced blood oxygen concentration, the cardiovascular centers decrease the parasympathetic and increase the sympathetic stimulation and vice versa (Freeman et al., 2012).

2.2 Fetal monitoring

As explained above, the sympathetic and parasympathetic ANS modulation can reflect specific characteristics in the fetal cardiovascular activity that may help to identify fetal oxygen insufficiency or cerebral dysfunction. This modulation produces small variations in the FHR signal, which could have a close relationship with fetal oxygenation as an indicator of the fetal condition (Bursa and Lhotska, 2017; Garabedian et al., 2017). According to that information, ideally, fetal monitoring during labor should be based on the analysis of such

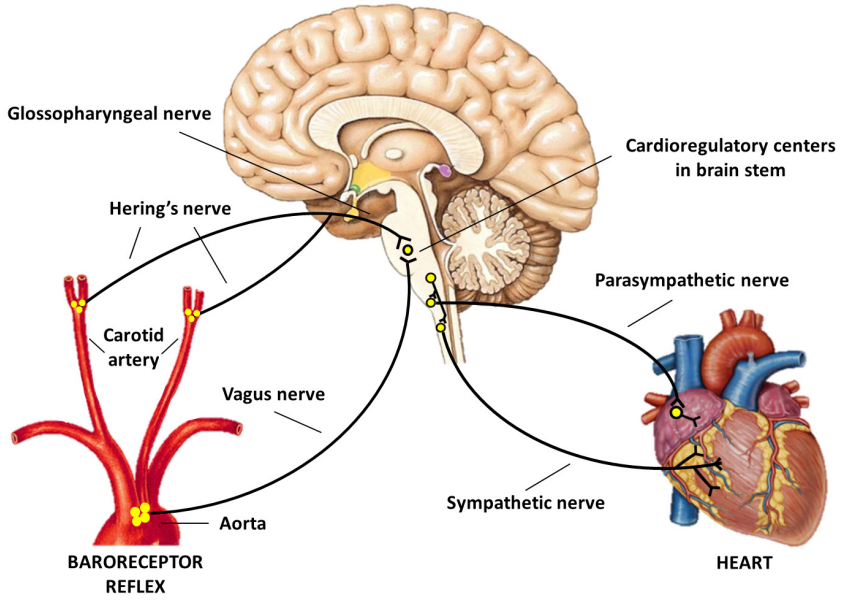


Figure 2.1: Neural communication pathways between baroreceptors and heart rate control modulated by the autonomic nervous system (ANS) (Ferreira Jr and Zanesco, 2016).

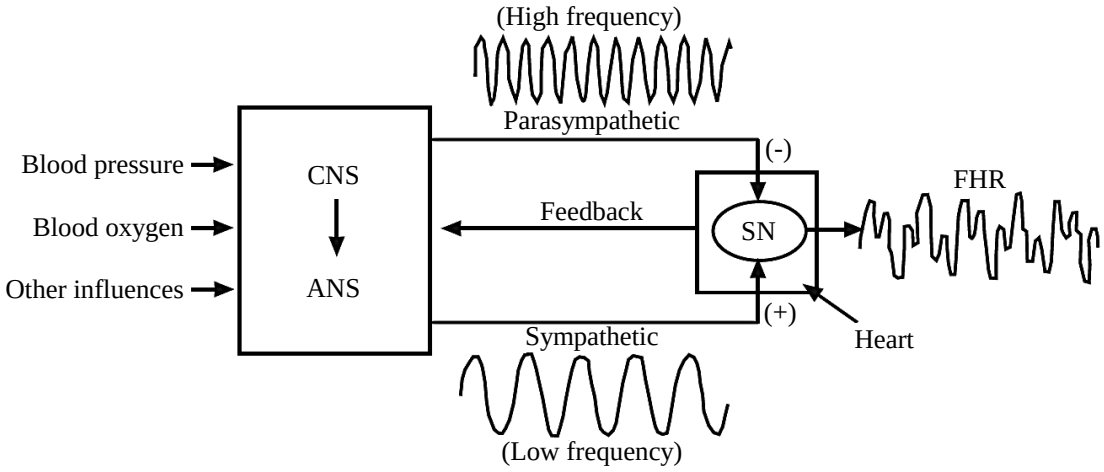


Figure 2.2: Representation of fetal regulation mechanisms modulated by the ANS; CNS: central nervous system; SN: sympathetic nervous (modified from Schönauer et al. (2008)).

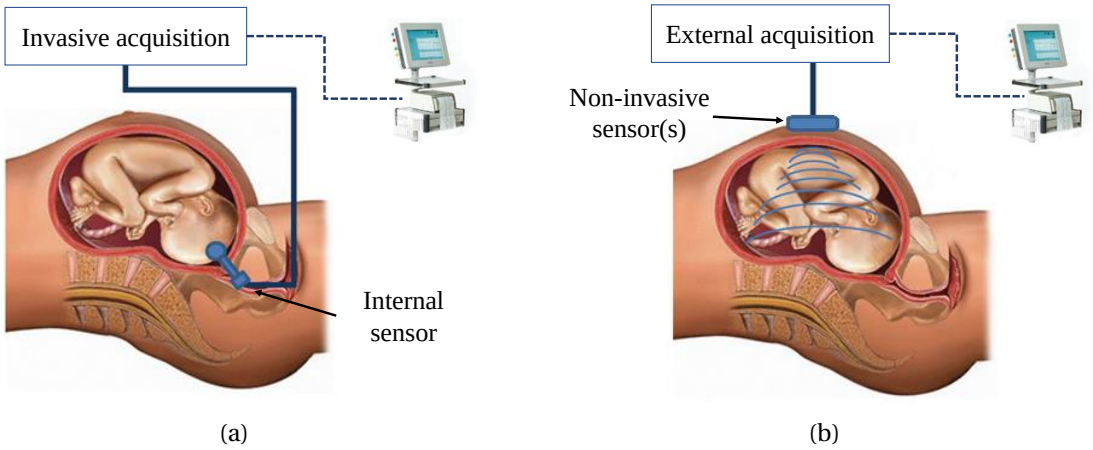


Figure 2.3: Representation of invasive (a) and non-invasive (b) techniques for fetal monitoring (modified from Martinek and Žídek (2012)).

characteristics resulting from the continuous, simultaneous, and antagonistic action of the mentioned physiological regulation mechanisms. However, considering the complexity involved in the labor and delivery process, the described physiological phenomenon can not be easily understood. In this context, different methods have been proposed for fetal monitoring, which can be roughly classified into invasive and non-invasive techniques. Invasive techniques involve the use of an internal sensor to extract information directly from the fetus, usually attached to its head (see Fig. 2.3(a)). In contrast, non-invasive techniques allow monitoring by using an external sensor(s) (see Fig. 2.3(b)). Most of these techniques are based on extracting information related to the fetal heart activity, whereas other alternatives or complementary techniques estimate the concentration of blood oxygen.

Table 2.1 provides an overview of methods proposed for intrapartum evaluation, where the first three correspond to invasive, and the last four correspond to non-invasive techniques. This table includes the minimal gestational age for their application and comments related to their main characteristics. In Section 2.2.1 and 2.2.2, we provide a brief explanation of the invasive and non-invasive techniques for fetal monitoring, respectively. Besides, Section 2.2.3 explains the analysis commonly performed as fetal outcome evaluation. The concept of intrapartum corresponds to the period spanning childbirth, whereas outcome evaluation corresponds to an examination of the baby performed just after delivery.

2.2.1 Invasive methods for intrapartum evaluation

Fetal blood sampling

Fetal blood sampling (FBS) involves the task of obtaining a small blood sample from the fetal scalp capillary. The pH level decreases when the acidity of the body fluids increases as a consequence of the accumulation of lactic acid due to metabolic acidosis (Parer, 1980).

Table 2.1: Overview of the main methods proposed for fetal monitoring during labor. Information extracted mainly from Behar et al. (2016), Clifford et al. (2014), and Peters et al. (2001).

Method	System	Gestational age	Comments
FBS	Fetal blood sampling	At delivery	- Invasive; blood sample from the fetal scalp - A good estimation of pH value (although not as precise as umbilical blood sample) - Uncommonly in an everyday clinical setting - Time-consuming; used only as complementary information to non-invasive methods
PO	Pulse Oximetry	At delivery	- Estimation of oxygen saturation based on light reflection - Invasive, not routinely recommended - Low specificity for fetal acidosis - External PO is under research
I-FECG	Invasive fetal electrocardiogram	At delivery	- Acquisition of ECG and FHR of the fetus - Single-channel only - Invasive, carrying a small additional risk, not routinely recommended - High signal-to-noise ratio (SNR)
PCG	Fetal phonocardiography; acoustic recording from abdomen	$\geq 28-30$ weeks	- Manual auscultation is usually employed, but electronic devices are sometimes used - Requires expert to locate the fetal heart - Low SNR; prone to acoustic noise (e.g., gastrointestinal activity and external noise) - Not routinely employed for continuous fetal monitoring
NI-FECG	Non-invasive fetal electrocardiogram	≥ 20 weeks	- Cheap and easy to handle - Continuous monitoring possible - FHR and possibly morphological analysis - Low SNR, improvements for this technology is under research
FMCG	Fetal magnetocardiography	≥ 20 weeks	- Requires skilled personnel - Morphological analysis of the FMCG easier than NI-FECG because of higher SNR - No long term monitoring possible to date because of the device size, cost, and complexity of the required instrumentation
CTG	Cardiotocography; ultrasound and pressure-sensitive transducers	≥ 20 weeks	- Contraction through pressure transducer - Non-invasive acquisition of FHR signal - No beat-to-beat data, cardiac function descriptor limited to FHR - Prone to artifacts and signal noise - Not passive; ultrasound irradiation

Since the metabolic acidosis is a key indicator of asphyxia, a low level of pH indicates fetal distress and a risk of permanent perinatal brain damage.

In general, it is used together with other methods to obtain complementary information about the fetal status, because the pH value is considered as an accurate indicator for the diagnosis of metabolic acidosis. For example, in the case of a non-reassuring prediction obtained by an external method (see Section 2.2.2), FBS might be performed in order to acquire a value of pH from the fetal blood. Unfortunately, FBS is considered as a time-consuming procedure and uncommonly performed in an everyday clinical setting (Bhide et al., 2016). Besides, the pH value obtained by the fetal scalp blood can differ from the value obtained by the umbilical artery (as a gold standard outcome measure). Therefore, the assessment performed by FBS is not always reliable (Kuehnle et al., 2016).

Pulse oximetry

Pulse Oximetry (PO) is a method developed to estimate the measure of oxygen saturation (FS_pO_2) of a fetus in-utero, using the principle of light reflection. Emitted photons travel through fetal tissue layers, then the light is dispersed, absorbed, and reflected, whose intensity depends on the oxygen saturation in the blood. However, as indicated by Steer (2008), low oxygen saturation has poor specificity for fetal acidosis. Also, there is no evidence that monitoring based on PO has improved the overall cesarean section rate compared with the conventional analysis (East et al., 2014). Therefore, the application of PO has not shown a significant contribution to fetal welfare assessment.

Although in clinical practice, this method is usually performed invasively, there is scientific research focused on trans-abdominal fetal PO (Böttrich et al., 2015; Zourabian et al., 2000). It consists of a technique performed externally through the abdomen of the mother. However, these works have concluded that the distance between the source and detector significantly affects the signal composition, and several factors related to the interference in fetal PO are not yet characterized. Therefore, more realistic models are needed to use the results for implementation purposes in real scenarios.

Invasive fetal electrocardiogram

Fetal Electrocardiogram (FECG) signal can be obtained either invasively (I-FECG) or non-invasively (NI-FECG). For a description of the NI-FECG, please refer to Section 2.2.2.

The I-FECG technique consists of acquiring the electrocardiogram signal of the fetus by introducing and attaching an electrode to its head. Then, the complete ECG curve is obtained to examine and evaluate its morphological characteristics. For that operation, it is necessary the rupture of membranes, which usually occurs at the second stage of labor.

This technique is commonly referred to as ST Analysis (STAN), which is suitably established in detecting and monitoring myocardial insufficiency in adults. Thus, the development of STAN of FECG has based on such experience and knowledge. The fetal brain

and heart are equally sensitive to changes in oxygen content; therefore, myocardial function serves as an indirect measurement of brain condition. During the signal acquisition, it displays the continuous ECG and allows the automatic computation of important markers of ECG.

The STAN, as an invasive technique, is not intended to be used autonomously but rather as an additional tool for standard external methods (see Section 2.2.2). It serves as a source of complementary information validating or invalidating the hypothesis of fetal condition and behavior assessed by other methods. However, recent studies show that the use of STAN during labor, as a complement to standard technologies such as CTG, does not improve perinatal outcomes or decrease the cesarean delivery rate (Saccone et al., 2016).

2.2.2 Non-invasive methods for intrapartum evaluation

Fetal phonocardiography

Auscultation is a concept that generally refers to listening for particular sounds produced within the body. Using this method, the FHR can be monitored either intermittently (at regular intervals during labor) or continuously. For this purpose, FHR auscultation guidelines have been proposed in the literature (Lewis et al., 2015), which help to assess the fetal condition through intermittent monitoring.

The main advantage of this methodology is that it allows the monitoring of the fetal heart in different positions and favors the mother mobility during labor. However, it takes time to develop clinical expertise, and even for the most experienced clinicians, it is impossible to recognize some subtle features involved in the cardiac activity, such as the FHR variability (Lewis et al., 2015).

The most basic tool for auscultation is the Pinard stethoscope, which allows the clinicians to monitor the fetal heart tones. However, it only provides a short term discrete evaluation of the cardiac function. As a consequence, the use of the stethoscope only focuses on assessing the extreme limits of the normal FHR baseline. In order to deal with this disadvantage, recent approaches have proposed automated microphone-based auscultation techniques, which introduce the concept of continuous auscultation-based monitoring (Jiménez-González and James, 2009; Jimenez-Gonzalez and James, 2010; Khokhlova et al., 2016; Zhdanov et al., 2015), whose technique is commonly referred to as the phonocardiography (PCG).

Although electronic fetal PCG has been investigated and proposed in the literature, the acquisition of PCG involves a very low signal-to-noise ratio (SNR). A higher signal quality can be obtained when the back of the fetus is in contact with the abdomen of the mother, but unfortunately, this position is not always possible and it can significantly change during labor. Therefore, this technology is almost unemployed in the current clinical practice (Kovács et al., 2011).

Non-invasive fetal electrocardiogram

The non-invasive fetal electrocardiogram (NI-FECG) is an alternative monitoring method to standard Doppler ultrasound (US) techniques. As explained in Section 2.2.1, the invasive FECG requires the rupture of membranes to attach the scalp in the fetal head and record the ECG signal. In the case of NI-FECG, the operation is performed externally based on US through the maternal abdomen. However, the acquisition of ECG signal by using this technique is still challenging, because it involves several artifacts and noise generated mainly by the heart of the mother, and abdominal and uterine muscles (Kahankova et al., 2019; Li et al., 2016). For a better signal acquisition, the US electrodes should be located as close as possible to the fetal head, but considering the complexity of the process, this condition is not always possible.

Fetal magnetocardiography

The fetal magnetocardiography (FMCG) is a non-invasive technique for studying fetal cardiac activity. This method employs a superconducting quantum interference device (SQUID) to facilitate a 3-dimensional analysis of the micromagnetic field (commonly 10^{-12} tesla) generated by phenomenal electric activities of the heart (Fukushima et al., 2011). The FMCG measures the magnetic field produced by the electrical activity of the heart.

Recent applications of FMCG have provided information about different physiological phenomena such as fetal behavioral states (Vairavan et al., 2016), acquired QT prolongation syndrome (Cuneo et al., 2013), supraventricular tachycardia (Campbell et al., 2006), assessment of fetal neurodevelopment (Wakai, 2004), and fetal arrhythmias (Stingl et al., 2013).

The main advantage of using the FMCG for the FHR analysis is that it is possible to obtain accurate beat-to-beat data non-invasively. However, the main disadvantage of the FMCG is the size, cost, and complexity of the required instrumentation, nevertheless smaller and portable devices may become available (Li and Farajidavar, 2015). Likewise, the FMCG signals may often involve maternal cardiac components and background noise. Consequently, FMCG observations need to be preprocessed in order to reconstruct uncontaminated fetal signals that may be more suitable for clinical application.

Cardiotocography

Fetal monitoring based on CTG involves the joint recording of FHR and UC signals during intrapartum as well as peripartum. In current clinical practice, these data are the primary source of information used for fetal welfare assessment, obtained through a non-invasive instrument known as cardiotocograph. This tool provides continuous information about changes in the FHR and their temporal relationship to UC stimuli, which may be examined during labor or subsequently if required.

This technique, as a non-invasive method, is usually used before the rupture of membranes, based on an external ultrasound probe and a transducer to acquire FHR and UC pressure, respectively. It is important to note that after the rupture of membranes, the acquisition of FHR data (extracted until this point externally by CTG) might be performed directly by an electrode attached to the fetal head. Then, the FHR can be computed from ECG R-R intervals acquired by an invasive method such as the I-FECG (explained in Section 2.2.1). Also, the UC might be obtained using either an external strain gauge or an invasive intrauterine pressure catheter.

Nowadays, although CTG is the most used technology for fetal monitoring in clinical practice, CTG analysis has been surrounded by significant controversies. Likewise, the interpretation of CTG has shown to suffer from very low specificity and wide intra- and inter-observer disagreement. Indeed, this thesis work focuses on extracting more significant information form the CTG signal that allows a better fetal assessment during labor. Considering that the CTG analysis is the subject of study in this thesis work, in Section 2.3, we provide a more in-depth explanation about its clinical operation and signal interpretation procedures.

2.2.3 Outcome evaluation

Acid-base analysis

The neonatal outcome assessment commonly relies on the analysis of blood gases in the umbilical cord of the baby immediately after delivery. This analysis provides information about the neonate condition in order to guide the requirements for its care.

The procedure to obtain the blood sample starts just after delivery by doubly clamping the cord, and taking a sample of both vein's blood and artery's blood. These samples allow calculating the values of blood gases. The main biochemical markers are the umbilical artery pH, the base excess, and the base deficit of extracellular fluid.

The pH measurement is the most common outcome parameter used to determine the presence of respiratory and metabolic acids in neonates. According to the literature (Kumar et al., 2016), the normal range of umbilical cord pH is 7.40 ± 0.20. As explained in Georgieva et al. (2013) and Yeh et al. (2012), pH is the most robust indicator to assess potential adverse outcomes. Although the relationship between the pH value and adverse outcomes may be relatively weak, it is considered as a gold standard outcome measure (Spilka et al., 2013). Therefore, pH values is the most used parameter for outcome evaluation in the current clinical practice.

Another biochemical marker is the base excess (BE), which is often used in clinical practice as a sign of metabolic hypoxia. The BE is defined as the amount of strong acid that must be added to each liter of fully oxygenated blood to return the pH to 7.40 at a temperature of 37°C and a partial pressure of carbon dioxide (pCO_2) of 40 mmHg. The main disadvantage of this measurement is that its values are often false positive (Rosén et al., 2007).

The last main biochemical marker is the base deficit of extracellular fluid (BDecf), which is an opposite measure than BE and obtained from the pH and pCO_2 level (Garibaldi et al., 1999). This measurement may estimate the cause of a low pH between the distinct physiological conditions of respiratory acidosis, due to a short-term accumulation of CO_2, and metabolic acidosis as a result of the lactic acid from a longer-term oxygen deficiency. According to Rosén et al. (2007), this parameter is a better measure of metabolic hypoxia than BE. However, pH remains the most robust measure, and according to Georgieva et al. (2013), it remains as the most informative.

Metabolic acidosis (phenomenon related to significant increase of possibility of cerebral palsy) in umbilical cord arterial blood at birth is commonly defined by pH < 7.00 (GOLDABER et al., 1991) (or < 7.05) and a Base deficit of extracellular fluid (BDecf) ≥ 12.0 mmol/L (Kumar et al., 2016; MacLennan, 1999). Likewise, several authors have pointed out different pH levels that can indicate pathological fetal acidemia such as 7.10 (Cahill et al., 2012; Georgoulas et al., 2006; Yeh et al., 2012), 7.15 (Di Tommaso et al., 2013), and 7.17 (Victory et al., 2004).

Apgar score

Virginia Apgar proposed this method in 1953, initially developed in order to ascertain the effects of obstetric anesthesia on neonates. Nevertheless, nowadays, this method is widely used in the outcome assessment.

The Apgar scale is determined by evaluating five parameters of the neonate examined at 1, 5, and 10 minutes after delivery. Each parameter is evaluated on a scale from 0 to 2, and then all values are summed up, giving the score at the particular instants. Therefore, the maximum score corresponds to 10 points. It has been reported that 5-min Apgar score had a higher concordance with asphyxia during labor compared with the other instances (Manganaro et al., 1994). Likewise, Eun et al. (2016) have concluded that low Apgar scores are significant perinatal risk factors for infantile seizures, especially in full-term and normal-birth weight infants. In general, the level of Apgar score ≤ 7 at 5 minutes is considered as an indicator of metabolic acidosis (Doria et al., 2007; Svenvik et al., 2015; Westerhuis et al., 2007) and cerebral palsy (MacLennan, 1999). However, the assessment of child is subjective and a significant variability among health care providers has been reported by Gupta et al. (2016).

2.3 Fetal monitoring by using CTG

2.3.1 CTG operation

As explained in Section 2.2.2, in current clinical practice, fetal monitoring during labor is commonly based on the assessment of FHR and UC signals, which are acquired electroni-

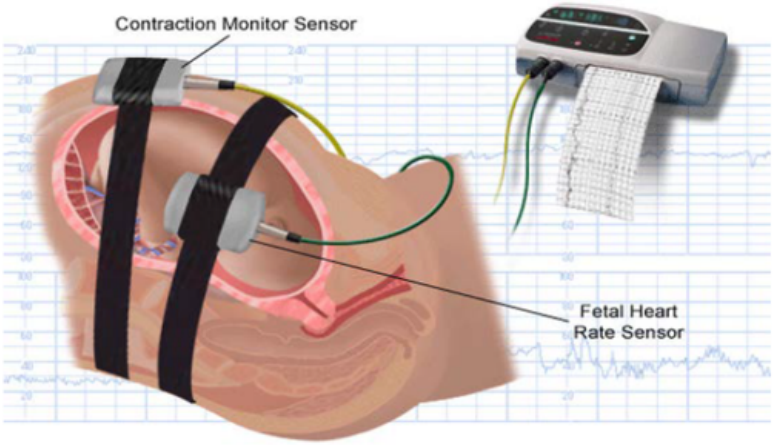


Figure 2.4: Example of an external cardiotocographic (CTG) setup; extracted from Abdulhay et al. (2014).

cally by CTG. The signal recording may be externally or internally acquired, depending on the process stage and conditions of the clinical procedure during labor.

On the one hand, the external CTG involves the use of a Doppler ultrasound transducer and a pressure transducer, for the acquisition of the FHR and a UC signal, respectively. The external CTG requires the mother to wear a belt across her abdomen while monitoring is being conducted (see Fig. 2.4). The ultrasound signals contain the information representing the fetal heart movements, so in order to obtain the fetal pulsations, an algorithm calculates the time interval between the loudest points in the cardiac cycle and displays the heart rate. This device uses frequencies around 1.5 MHz, and based on the reflected ultrasound signal, the fetal heart movement is detected within the same transducer. The signals are usually acquired continuously, although occasionally, they are captured intermittently, depending on the clinician's criteria.

On the other hand, an alternative to external CTG monitoring is internal CTG, which consists of attaching an electrode directly to the head of the fetus. This form of monitoring requires a ruptured amniotic sac (either spontaneously or artificially). The main disadvantage of the internal CTG is that it is an invasive method, and it also restricts the woman's mobility considerably. Therefore, external CTG remains as the most used technique in the current clinical practice, whose main advantage is its non-invasibility for working. However, CTG monitoring suffers from a significant disadvantage, which is related to the complexity of the FHR interpretation, whose procedure is explained in the following.

2.3.2 CTG interpretation by guidelines

As explained in Section 2.2, the sympathetic and parasympathetic ANS modulation can reflect particular characteristics in the fetal cardiac activity, whose variations could have a close relationship with fetal oxygenation as an indicator of the fetal condition (Bursa and Lhotska, 2017; Garabedian et al., 2017). In the current clinical practice, the fetal welfare

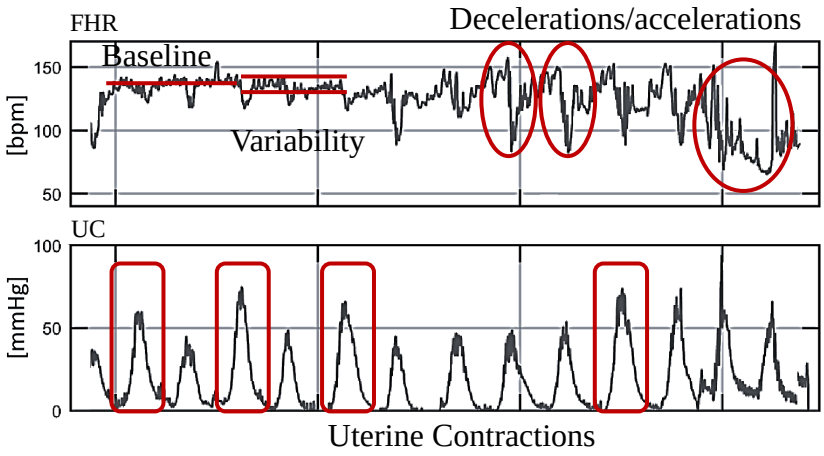


Figure 2.5: Conventional CTG patterns defined in guidelines.

assessment based on CTG involves a visual interpretation that attempts to assess those characteristics. This operation depends on the analysis of several morphological signal patterns from the FHR and UC signals, such as the FHR baseline, FHR variability, accelerations, decelerations, and UC events. Fig. 2.5 shows a graphic representation of these signal patterns recognized from real CTG data.

Nowadays, these morphological patterns and their physiological connection are defined in guidelines proposed for CTG fetal monitoring (American College of Obstetricians and Gynecologists and others, 2010; Artal and O'toole, 2003; Ayres-de Campos and Bernardes, 2010; Ayres-de Campos et al., 2015; Macones et al., 2008; Parer and Ikeda, 2007; Rooth et al., 1987; Royal College of Obstetricians and Gynaecologists (Great Britain). Clinical Effectiveness Support Unit, 2001; Spilka et al., 2014b), where the FIGO (International Federation of Gynecology and Obstetrics) guidelines proposed by Rooth et al. (1987) were the first guidelines for CTG interpretation. According to CTG guidelines, these signal patterns can provide representative information to clinicians about the fetal response that can help to assess the fetal health condition. As shown in Table 2.2, these patterns involve particular characteristics that can be associated with normal, suspicious, and pathological fetuses. For example, a FHR baseline lower than 100 bpm, or reduced variability (< 5 bpm) are patterns that indicate a fetus with a high probability of hypoxia. On the contrary, a FHR baseline between 110-160 bpm, or variability in the range of 5-25 bpm, indicates a normal fetal condition (see more details in Table 2.2). This information helps clinicians to interpret how the fetus reacts to stress and thus indicate timely intervention before the risk of consequences increases (Ayres-de Campos and Bernardes, 2010; Hruban et al., 2015).

The main purpose of the CTG guidelines is to *standardize* the interpretation and classification of the CTG recordings into different categories, trying to avoid subjective interpretation. Basically, clinicians observe by naked eye morphological time-domain CTG patterns and associate them with a particular situation defined in guidelines. Then, according to the guidelines recommendations for this particular situation, they commonly classify the

	Normal	Suspicious	Pathological
Baseline	110-160 bpm	Lacking at least one characteristic of normality, but with no pathological features	< 100 bpm
Decelerations	No repetitive[a] decelerations	Lacking at least one characteristic of normality, but with no pathological features	Repetitive[a] late or prolonged decelerations during > 30 min or 20 min if reduced variability, or one prolonged deceleration with > 5 min
Variability	5-25 bpm	Lacking at least one characteristic of normality, but with no pathological features	Reduced variability, increased variability, or sinusoidal pattern
Interpretation	Fetus with no hypoxia/acidosis	Fetus with a low probability of having hypoxia/acidosis	Fetus with a high probability of having hypoxia/acidosis
Clinical management	No intervention necessary to improve fetal oxygenation state	Action to correct reversible causes if identified, close monitoring or additional methods to evaluate fetal oxygenation	Immediate action to correct reversible causes, additional methods to evaluate fetal oxygenation, or if this is not possible expedite delivery. In acute situations (cord prolapse, uterine rupture, or placental abruption) immediate delivery should be accomplished

[a] Decelerations are repetitive in nature when they are associated with more than 50% of uterine contractions.

Table 2.2: Summary of FIGO (International Federation of Gynecology and Obstetrics) guidelines for cardiotocographic (CTG) classification, extracted from Ayres-de Campos et al. (2015).

fetal condition as normal, suspicious, or pathological in order to decide about the next step of the management (Sweha et al., 1999).

However, guidelines have been surrounded by great controversies, because they still lack consensus in many key aspects. The absence of objective criteria to define some of the FHR patterns, together with the disparity of definitions between different guidelines, make the diagnosis a difficult task. Moreover, guidelines are considered too complex, making them prone to rapid memory decline (Santo and Ayres-de Campos, 2012). Using guidelines leads to good consensus in extreme fetal cases (normal and severe fetal asphyxia) but no clear recommendation for cases between these two extremes, which represent at least 50% of all the intrapartum cases (Parer and Ikeda, 2007; Ugwumadu, 2014). As a consequence, CTG interpretation has demonstrated a wide intra- and inter-observer disagreement, lack of objectivity, and poor interpretation reproducibility (Ayres-de Campos and Bernardes, 2010; Freeman et al., 2012; Ghi et al., 2016; Haritopoulos et al., 2016; Santo and Ayres-de Campos, 2012; Santo et al., 2017).

In this context, clinical research has focused mainly on improving conventional CTG guidelines (Clark et al., 2017; Evans et al., 2018; Pinas and Chandraharan, 2016; Thellesen et al., 2015). Likewise, more advanced guidelines have been proposed, such as the 5-tier classification system proposed by Parer et al. (2006) and Parer and Ikeda (2007), which empirically defines different patterns of fetal health state and risk of evolution to another more severe pattern with a higher risk of acidemia. The 5-tier system had been examined in several publications (Coletta et al., 2012; Elliott et al., 2010; Parer, 2014; Parer and Hamilton, 2010), and they have found that achieved levels of agreement and sensitivity are much higher compared with the previously reported guidelines. However, this system is not commonly used in the current clinical practice, which can be explained by the fact that their analysis, based on human interpretation, is considered more complex than other proposed guidelines. Besides, the mentioned literature emphasizes that a correct interpretation and a significant intra- and inter-observer agreement probably will never be achieved by using guidelines because this methodology still requires the clinician to visually measure several factors, grade them, and then consider many possible combinations of them. In consequence, the lack of consensus of CTG guidelines and interpretation problems are still unsolved so far.

2.3.3 Computer-based systems support

In order to improve the CTG interpretation, several CTG training simulators and expert systems have been proposed and implemented.

On the one hand, training simulators (3B Scientific®; iSimulate CTGi; K2 Medical Systems™ PTP; PHC NORDIC) can be used to assist clinical staff to get expertise in the interpretation of CTG in a risk-free and controllable environment, since training can contribute to an increased quality of care, and improve CTG competence and clinical management (Fritz et al., 2017; Pehrson et al., 2011). However, regular education is very resource demanding,

and it usually sustained only for a limited period, after which the additional interpretive skills deteriorate (Hruban et al., 2015; Pehrson et al., 2011).

On the other hand, expert systems (Central Station Monitor MF-7400 (Maeda et al., 2012), HUNTLEIGH - Arjo (Pardey et al., 2002), INFANT® (Keith and Greene, 1994), IntelliSpace Perinatal® (Devoe et al., 2000), K2 INFANT-Guardian® (K2 Medical Systems™, 2020), Omniview-SisPorto® (Ayres-de Campos et al., 2008), PeriCALM™ (Hamilton and Kimanani, 1994), Trium CTG Online (Schiermeier et al., 2008)) can assist clinicians with the interpretation of CTG. Most of them are based on conventional patterns defined in CTG guidelines, which are employed used as a knowledge database in order to emit information and visual alerts to the observer related to changes in the CTG recording. However, recent literature shows that there is little or no evidence that expert systems have an effect on the incidence of fetal acidemia, cesarean delivery, or perinatal mortality, compared to normal CTG analysis (Lutomski et al., 2013; Nunes and Ayres-de Campos, 2016).

Although training simulators and expert systems are under development and evaluation, they have not shown superior effectiveness in reducing newborn acidemia without excessive obstetric intervention (Ugwumadu et al., 2016).

In summary, in the last thirty years, many efforts have been made in order to reduce the problems of CTG intra- and inter-observer disagreement and interpretation. Concerning the solutions related to clinical support, visual analysis by using CTG guidelines makes the diagnosis a difficult task. As a possible solution, intensive and continuous clinicians training in CTG may, in some way, improve the competence and clinical performance, however, the interpretive skills seem to decrease faster than theoretical knowledge. At the same time, regular education is very resource-demanding (Hruban et al., 2015; Pehrson et al., 2011), which is an important disadvantage considering the time that doctors need to invest on training activities. Concerning the proposed computer-based support, there is almost no evidence that it can improve the fetal welfare assessment compared with the conventional CTG analysis. This phenomenon can be explained by the fact that most of these systems still depend on human factors and emulate what clinicians visually do with the CTG signal, i.e., these systems assume that currently used time-domain patterns fulfill the essential requirements for a correct interpretation. Conclusively, the methods proposed in the clinical research area have not shown satisfactory results on improving the CTG analysis so far, and a reliable CTG interpretation is still challenging.

3

State of the art

This chapter presents the state of the art in CTG analysis focused on the field of biomedical signal processing and classification. In order to identify better the scope of related works presented in this chapter, in Fig. 3.1, we show a block diagram that provides an overview about the main proposed solution in both clinical and engineering research areas. In Chapter 2, we have already explained the main context, problematic behind the CTG interpretation, and clinical solutions focused on improving the interpretation and use of CTG guidelines. Likewise, we have explained the proposed computer-based solutions such as training simulators and expert systems. In this chapter, we present a review of recent scientific works proposed in biomedical engineering for CTG signal analysis (see right block of Fig. 3.1), which are focused on extracting more representative information from the FHR signal in order to improve the automatic CTG assessment and classification. These works can be roughly divided into three main groups: signal feature extraction, CTG classification, and CTG signal modeling approaches, which are explained in detail in the following sections.

3.1 Signal feature extraction

In recent years, several approaches have been proposed for FHR signal feature extraction, which can be subdivided into two main categories: stationary and time-variant techniques (Haritopoulos et al., 2016).

3.1.1 Stationary techniques

CTG analysis based on stationary techniques commonly involves the extraction of signal features in different domains such as time-domain, frequency-domain and nonlinear features.

Time-domain features

Features in the time-domain are usually based on statistical indicators computed from short- and long-term analyses, which mainly aim to describe the standard morphological patterns described in CTG guidelines (Agostinelli et al., 2017; Cesarelli et al., 2009, 2011;

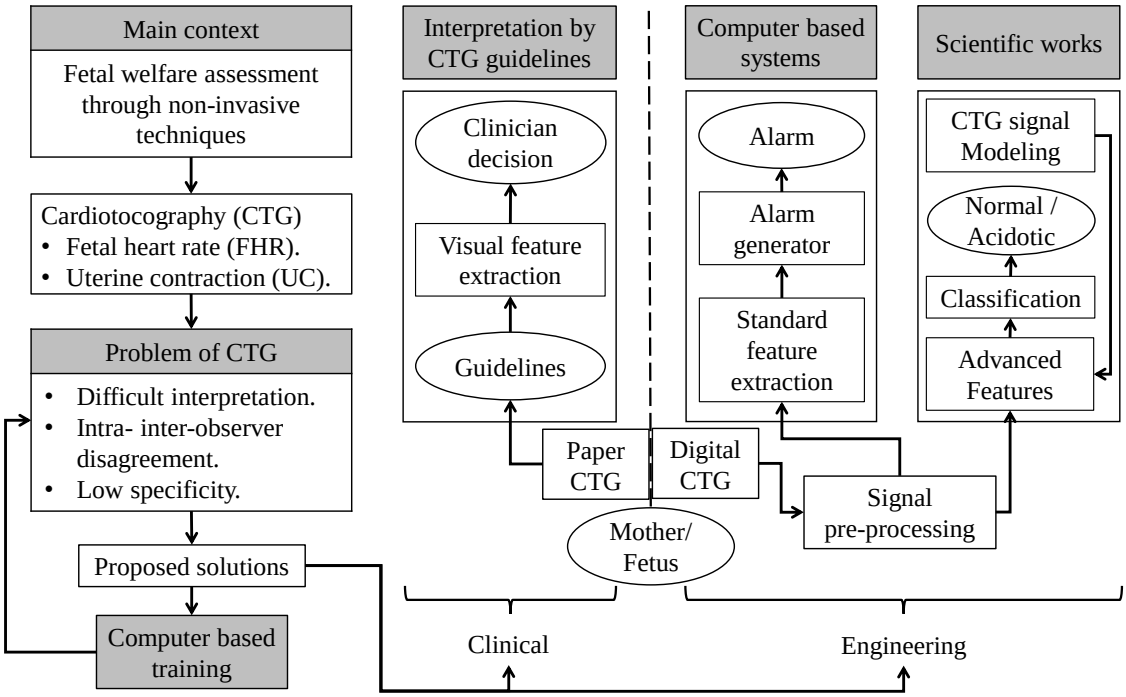


Figure 3.1: Overview of main proposed solutions for CTG analysis in clinical and engineering research areas.

Das et al., 2011; Fuentealba et al., 2017a; Wei et al., 2012). Besides, these type of features has been employed to develop statistical models in order to reproduce the CTG interpretation by clinicians (Cao et al., 2006).

Frequency-domain features

As established in the literature (Gonçalves et al., 2018, 2006; Kwon et al., 2012; Signorini et al., 2003; Van Laar et al., 2010), different frequency bands have been investigated in the FHR signal because it is assumed that variations in particular spectral bands can be related to the fetal health condition (Siira et al., 2013). The studied bands in the FHR signal include a zero-frequency component (DC), a very low frequency (VLF) band ($0-0.03$ Hz), a low frequency (LF) band ($0.03-0.15$ Hz), a medium frequency (MF) band ($0.15-0.5$ Hz), and a high frequency (HF) band ($0.5-1.0$ Hz). These FHR frequency bands and their associated physiological phenomenon can be summarized as presented in Table 3.1.

The DC component corresponds to a long-term average of the FHR signal values. The VLF band is associated with thermoregulatory processes and slow control regulating mechanisms of peripheral vessels, which can present nonlinear characteristics. The LF involves information mainly associated with the fetal sympathetic and parasympathetic ANS modulation. The MF depends on maternal breathing and fetal movements. The HF is related to the parasympathetic ANS modulation and fetal breathing. It is important to note that the sympathetic branch develops earlier than parasympathetic one, which can achieve a

Table 3.1: FHR frequency bands commonly investigated in the literature.

Frequency band	Abbrev.	Associated phenomenon
Zero-frequency	DC	Component describing the average of the FHR.
0 – 0.03 Hz	VLF	Slow control mechanisms; presents nonlinear contributions.
0.03 – 0.15 Hz	LF	Stimuli of the fetal autonomic nervous system (ANS).
0.15 – 0.5 Hz	MF	Maternal breathing and fetal movements.
0.5 – 1.0 Hz	HF	Parasympathetic ANS and fetal breathing.

complete development after birth (Signorini and Magenes, 2016). For this reason, the association between the LF and HF bands with the parasympathetic ANS activity can not be assured, and therefore usually omitted in the literature.

The description concerning the FHR frequency bands presented above provides important information, which is employed for the analysis of results, as will be explained in Chapter 7.

In order to analyze the information described above, different approaches based on frequency analysis have been proposed. Most of these stationary frequency-based approaches depend on operations computed by the Fast Fourier Transform (Cesarelli et al., 2007; Kwon et al., 2012; Schneider et al., 2009; Siira et al., 2005; Takatani et al., 2018; Van Laar et al., 2011, 2010, 2009, 2008), and AR spectral-based analysis (Cazares et al., 2001; Illanes and Haritopoulos, 2015; Sankhe and Desai, 2016; Signorini and Magenes, 2016; Signorini et al., 2020; Warrick and Hamilton, 2012). The AR modeling offers a quantitative spectral description of the signal based on a small number of model parameters, which requires only a fraction of samples compared to the classical Fourier analysis (Salamalekis et al., 2006).

Nonlinear features

In order to study the nonlinear characteristics involved in the FHR signal, several approaches have investigated the connection between nonlinear features and the fetal health condition. In this perspective, approximate and sample entropy have been employed to characterize and analyze the complexity of the FHR and UC signals (Marques et al., 2020). Mutual information has been employed to study nonlinear temporal FHR dynamics (Granero-Belinchon et al., 2017), and assess the UC and FHR signals coupling (Warrick and Hamilton, 2015). Multiscale entropy has been used to investigate the complexity (Costa et al., 2014; Ferrario et al., 2009) and regularity (Ferrario et al., 2006) of the FHR signal. Multivariate analysis based on linear and nonlinear features has been applied to distinguish between healthy and intrauterine growth-restricted fetuses (Magenes et al., 2014; Signorini et al., 2020). Linear and nonlinear parameters have been used to assess their evolution of throughout pregnancy (Gonçalves et al., 2018), and study the impact of fetal age and gender on FHR signals (Spyridou et al., 2018). Nonlinear parameters (namely entropy and compression) have been also evaluated as labor predictors based on traces

corresponding to one or two weeks before labor (Monteiro-Santos et al., 2020). Fractal dimension, entropy, and complexity based features have been extracted and evaluated in CTG intrapartum classification by using Naive Bayes, Support Vector Machine (SVM), and Decision Tree (Spilka et al., 2012). Besides, as explained in Section 4.2, recent approaches have used empirical mode decomposition (EMD) to characterize and extract information from the FHR signal dynamics.

3.1.2 Time-variant techniques

Considering that the FHR dynamics modulated by the ANS describe characteristics highly variant in time, some works have focused on studying these non-stationary characteristics. These works are based mainly on time-varying frequency methods such as short-time Fourier transform (Romano et al., 2006b), quadratic time-frequency distributions (Dong et al., 2014), and time-varying AR modeling (Fuentealba et al., 2018a; Romano et al., 2006a; Signorini et al., 2003). Besides, discrete (Cattani et al., 2006; Spilka et al., 2014a; Zhang et al., 2019) and continuous wavelet transform (Bursa and Lhotska, 2017) has been applied for spectral-based characterization of the of FHR and UC signals.

3.2 Signal classification approaches

In CTG classification, several approaches have been proposed. Most of them, as performed by approaches explained above, also employ computerized FHR features extracted from different domains such as time-domain, frequency-domain, and nonlinear features. Nevertheless, they mainly focus on tuning computer-based algorithms to optimize their model, thereby improving their performance in CTG classification. In this context, SVM have been used to identify normal fetuses or at risk of developing metabolic acidosis (Georgoulas et al., 2006; Krupa et al., 2011). Hidden Markov Models (HMM) (Georgoulas et al., 2004), Sparse SVM (Spilka et al., 2017), Least Squares SVM (Cömert et al., 2018; Georgoulas et al., 2017), non-parametric Bayesian models (Yu et al., 2017), fuzzy inference systems (Signorini et al., 2000), and convolutional neural networks (Bursa and Lhotska, 2017; Marques et al., 2019; Petrozziello et al., 2018) have been proposed for automatic classification of FHR recordings.

Some classification methods attempt to consider the phenomenon involved in the fetal response by studying the relationship between the UC and FHR as an input/output system. In this context, Warrick et al. (2010) have proposed a system identification approach based on power spectral density computed from an AR-based model. Likewise, statistical methods based on a Wilcoxon rank-sum test (Warmerdam et al., 2016a) and SVM classifiers (Warmerdam et al., 2016b, 2018) have been employed to examine whether FHR features during UC and rest periods, analyzed separately, improve the CTG classification.

Also, many approaches focusing on performance evaluation and comparison of different machine learning methods for CTG classification have been published in the literature

(Bhatnagar and Maheshwari, 2016; Cömert et al., 2016; Czabanski et al., 2008; Huang and Hsu, 2012; Jezewski et al., 2007; Nagendra et al., 2017; Ocak, 2013; Sahin and Subasi, 2015; Sundar et al., 2012; Yılmaz and Kılıkçıer, 2013).

3.3 CTG signal modeling

Since the CTG interpretation is still a challenging issue, in the last years, it has emerged an increasing interest in developing more advanced and complex computerized methods created for better understanding the fetal response during labor. In this perspective, more sophisticated analyzing algorithms and modeling approaches have been proposed, which mainly focus on generating synthetic CTG signal patterns. These approaches can be roughly divided into two main categories: physiological- and qualitative-based models.

On the one hand, physiological-based models consist of closed-loop control systems, which consider the fetal response and ANS modulation. This type of model aims to reproduce mathematically synthetic CTG signal data and offer quantitative information of certain emulated patterns such as late (Jongen et al., 2016a; van der Hout-van et al., 2013b), variable (Jongen et al., 2016b; van der Hout-van et al., 2013a), and early FHR decelerations (van der Hout-van et al., 2012), and FHR variability (Jongen et al., 2017). Likewise, some approaches have proposed simulated models to describe the spontaneous evolution of UC events (Bastos et al., 2012), and the relationship between UC and FHR signals by using a non-parametric approach to describe the involved dynamics in terms of an impulse response function (Warrick et al., 2009).

On the other hand, qualitative-based models have been designed to consider the characteristics of the CTG signal as observation of the process but without considering the fetal physiological phenomena. These simpler mathematical models aim to simulate particular temporal events present in the CTG recording (Illanes et al., 2015; Improta et al., 2014; Martinek et al., 2016). Likewise, synthetics CTG signals have been used for a comparison of baseline estimation algorithms (Kupka et al., 2006).

3.4 Summary of the state of the art

In the last thirty years, many efforts have been made in order to reduce the problems of CTG intra- and inter-observer disagreements and interpretation.

As explained in Section 2.3.2, clinical research has focused mainly on improving the conventional CTG guidelines. However, although more advanced guidelines have been proposed, their dependency on visual analysis still makes the diagnosis a difficult task, and the CTG interpretation problems are still unsolved so far. In this context, computer-based support has been proposed, such as training simulators and expert systems (see Section 2.3.3). On the one hand, concerning training simulators, intensive and continuous clinicians training in CTG could, in some way, improve the competence and clinical practice. How-

ever, the interpretive skills seem to decrease faster than theoretical knowledge, and at the same time, regular education is very resource demanding. On the other hand, concerning the computer-based support by expert systems, there is practically no evidence that it has improved the fetal welfare assessment compared to the conventional CTG analysis. This phenomenon can be explained by the fact that most of these systems still depend on human factors and emulate what clinicians visually assess by using CTG guidelines.

In the biomedical engineering area, several approaches have been proposed for CTG feature extraction and classification. In general, they reveal that, by using an optimal set of signal features and tuning a particular computer-based classifier properly, the CTG classification performance can improve compared with the conventional CTG analysis. However, although several efforts have been made, the results are not satisfactory enough for their use in clinical practice (Haritopoulos et al., 2016; Lutomski et al., 2015; Nunes and Ayres-de Campos, 2016; Ugwumadu et al., 2016). This phenomenon can be explained by the fact that most of the automatic classification techniques proposed in the literature base the operation only on a snapshot of the complete process without considering the the nonlinear and non-stationary phenomenon involved in the ANS activity (Garabedian et al., 2017; Romano et al., 2016b). Certainly, the fetal compensation system modulated by the ANS triggers variations in the beat-to-beat FHR, whose time-varying dynamics can involve significant information related to the fetal health status (Cesarelli et al., 2010; Romano et al., 2006a). Consequently, conventional processing techniques that do not integrate these physiological characteristics may not be appropriate for a correct CTG assessment.

In that context, time-variant based approaches have been proposed in order to study these implicit characteristics. However, they have certain limitations because they are mainly focused only on studying the fetal reactivity based on the analysis of the FHR as a response to the UC stimulus. Besides, as presented in Section 3.3, physiological- and qualitative-based models have been proposed to describe the implicit phenomenon related to the time-variant fetal behavior. These approaches are based on mathematical CTG models focused on generating synthetic CTG signal, which have been designed as a training tool for the educational purpose of gynecologists. Therefore, as they are designed nowadays, they can help to understand the physiological process, but they are not designed to characterize the fetal physiological phenomena from real CTG data.

Conclusively, the described solutions proposed in biomedical engineering in the area of signal processing, are practically not used in real scenarios. Several algorithms have been proposed for signal feature extraction and classification, but the clinical staff can not directly assess them. This perspective brings up to the question of whether approaches based only on the extraction of signal feature and machine learning classification are appropriate for a correct CTG signal interpretation.

In our opinion, a computer-based CTG analysis must integrate the described fetal physiological phenomenon, and provide significant and understandable information to clinicians (in Section 7.3 we address in-depth this important matter).

Although the CTG interpretation problems are still unsolved so far, all the advances in biomedical engineering research presented above have paved the way towards an automatic CTG analysis based on the progressive fetal response over time. In this context, the study of the time-variant FHR dynamics involved in the fetal compensatory mechanism is essential, which is the main direction of this thesis work.

4

Signal processing techniques: Theoretical background

This chapter provides a comprehensive review of the principles behind the signal processing methods employed in this work, such as the Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) and time-varying autoregressive (TV-AR) modeling. First, we present an introductory section (Section 4.1), which explains the main concepts behind the proposed time-variant spectral analysis by comparing the employed techniques with standard signal processing methods. Then, in Section 4.2 and 4.3, we present the theoretical background and main concepts related to ICEEMDAN and TV-AR modeling, respectively.

4.1 Basic concepts in frequency analysis

This section explains the essential concepts that are commonly involved in a stationary and a time-variant spectral analysis, and how the properties of the signal data (e.g., sampling frequency, SNR, and length of available data) can affect the spectral estimation. These concepts are essential for selecting an appropriate method for the analysis of the CTG signal.

For a better understanding of these concepts, we generate three examples, which are presented in Section. 4.1.1, 4.1.2, and 4.1.3, respectively. The first example corresponds to stationary frequency analysis, whereas the second and third examples consist of time-variant frequency analysis, and signal decomposition, respectively.

4.1.1 Example 1: Stationary frequency analysis

For the first example, we make use of a simulated signal (Sim1) resulting from the sum of four sinusoidal components of 60 s length and a sampling frequency $fs = 10$ Hz, with constant amplitude (A) and frequency (f) (see Fig. 4.1). As can be observed in Fig. 4.1(a-d), the four components correspond to sine curves of $[A, f] = [1.0, 0.15$ Hz$]$, $[A, f] = [0.8, 0.5$ Hz$]$, $[A, f] = [0.6, 1.0$ Hz$]$, and $[A, f] = [0.4, 1.5$ Hz$]$, respectively. The resulting signal Sim1, used for this example, is plotted in Fig. 4.1(e). The frequencies involved in the simulated signal

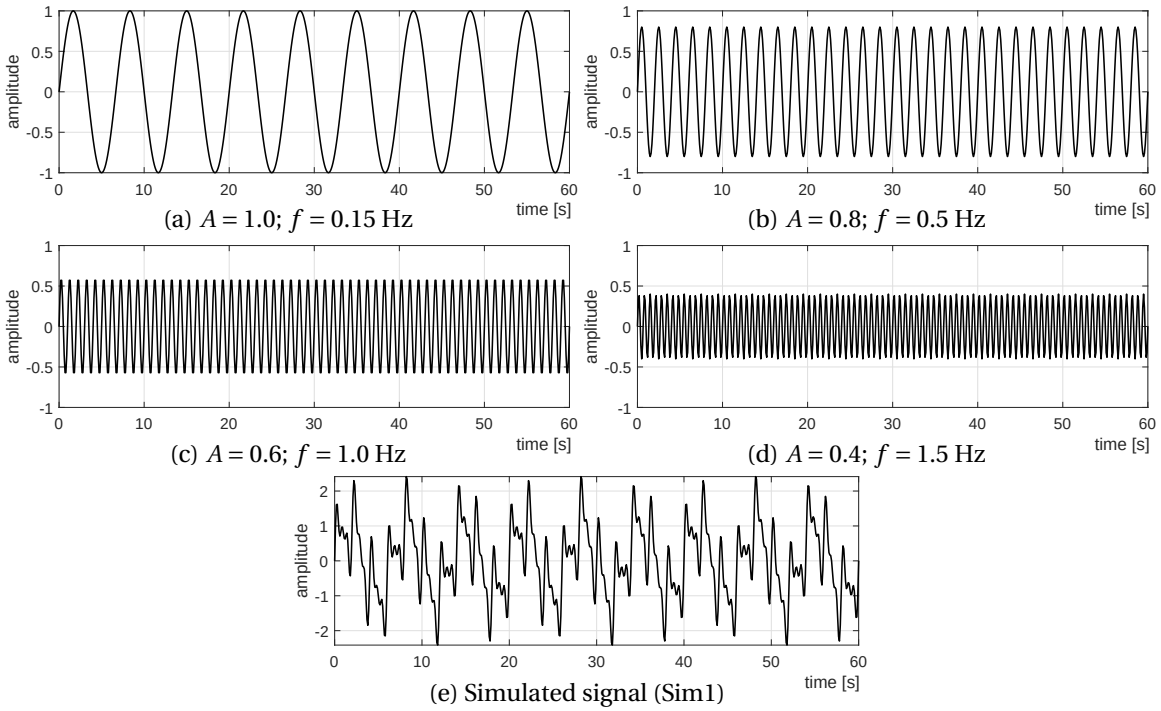


Figure 4.1: (a-d) Sinusoidal components of amplitude (A) and frequency (f), sampled at 10 Hz; (e) Simulated signal 1 (Sim1) resulting from the sum the sinusoidal components (a-d).

were selected arbitrarily from the frequency range usually studied for FHR signal analysis (see Section 3.1.1).

Considering that Sim1 involves the known frequency components described above, this example aims to show how the estimation of the power spectral density (PSD) depends on the properties of the signal data. For this purpose, we compute the PSD based on two methods, AR modeling and discrete Fourier transform (DFT). In this example, for the AR spectrum estimation, we employ a model order $p = 10$, whose coefficients are computed by using the Yule-Walker method. For the DFT spectrum computation, we use a Hamming window of length equal to signal length. The idea is to compare the performance of these two methods when considering different sampling frequencies (fs), signal lengths (L), and the addition of white zero-mean noise of amplitude e.

The proposed example of frequency analysis is displayed in Fig. 4.2, where the data in time-domain are plotted in black, and under each time-domain data, the corresponding PSD, estimated by AR modeling and DFT, are plotted in blue and red color, respectively.

In Fig. 4.2(a-c), as a first experiment, we vary L from 60 to 20 s, maintaining the value of fs, and without adding noise. Here, we can observe that the PSD computed from Sim1 ($L = 60$ s; $fs = 10$ Hz) is well defined, and both the AR and DFT spectrum show a very similar frequency resolution. However, by considering a shorter L (Fig. 4.2(c)), the PSD computed by AR modeling (blue) shows a better frequency resolution than DFT, which is represented by thinner frequency peaks, mainly for higher frequency components.

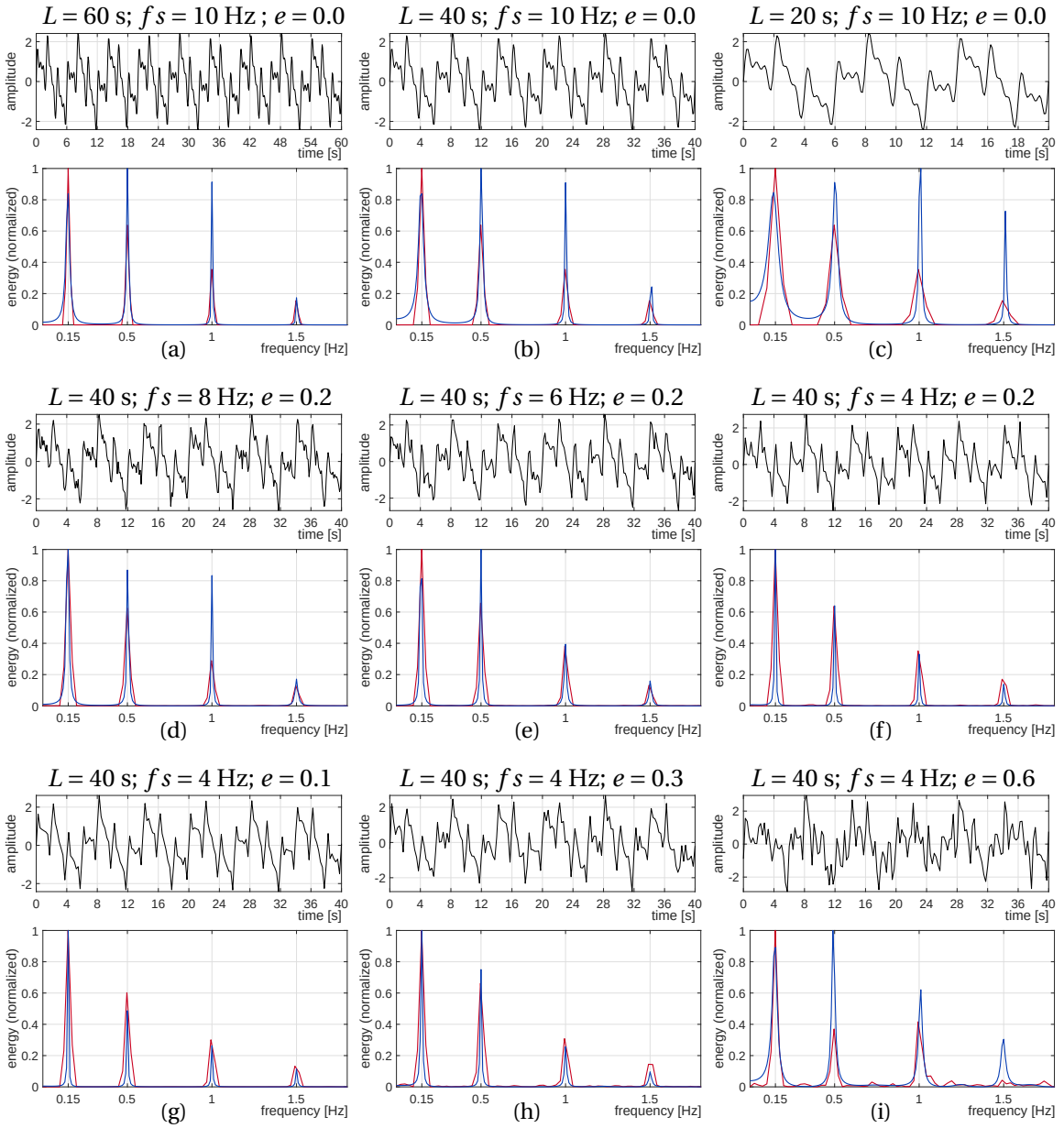


Figure 4.2: Example 1: Frequency analysis based on fast Fourier transform (FFT) (red) and autoregressive (AR) modeling (blue) applied to the simulated signal Sim1 presented in Fig. 4.2.

As a second experiment, in Fig. 4.2(d-f), we add white noise of amplitude $e = 0.2$, considering a signal length $L = 40$ s, and we decrease fs from 8 to 4 Hz. As a result, we can observe that the spectral estimation also depends on the fs parameter. A lower value of fs leads to less frequency resolution in the DFT spectrum (red). Nevertheless, the AR spectrum (blue) shows a more robust estimation while decreasing fs.

The last experiment (see Fig. 4.2(g-i)) aims to show how white noise added to the signal can affect the spectral estimation. For this operation, we consider a signal length $L = 20$ s and $fs = 4$ Hz, while e is increased from 0.1 to 0.6 in amplitude. As can be observed, the added noise strongly affects the spectral estimation performed by the DFT, represented by *false* frequency peaks that appear along the frequency axis. Particularly, when adding white noise of amplitude 0.6 (see Fig. 4.2(i)), we can observe that, for the DFT spectrum (red), the component of $f = 1.5$ Hz has practically disappeared. In contrast, the AR spectrum keeps its shape and still estimate correctly all the frequency components involved in Sim1.

In summary, the example presented in Fig. 4.2 indicates that the characteristics of the signal can considerably affect the spectral estimation, and thus the performed frequency analysis. Nevertheless, when varying L, fs, and SNR, the spectral representation based on AR modeling is more robust than the DFT.

4.1.2 Example 2: Time-variant frequency analysis

The example presented in Fig. 4.3 shows a frequency analysis performed on a signal involving time-variant characteristics. For this purpose, we generate a simulated signal (Sim2) from four sinusoidal components of 20 s length, sampled at 4 Hz, concatenated in series. Similarly to the previous example, these four components correspond to sine curves of $[A, f] = [1.0, 0.15$ Hz$]$, $[A, f] = [0.8, 0.5$ Hz$]$, $[A, f] = [0.6, 1.0$ Hz$]$, and $[A, f] = [0.4, 1.5$ Hz$]$, respectively. In order to evaluate the performance of the methods used for spectral estimation, we add white zero-mean noise of amplitude $e = 0.1$ and 0.2.

Fig. 4.3(a), (b), and (c) show the signal Sim2 with the white noise of amplitude 0 (without noise), 0.1, and 0.2, respectively. The second row shows the stationary spectrum based on AR modeling (blue) and DFT (red) (explained in Section 4.1.1), computed for each signal. The third and fourth rows display the time-variant spectrum computed by short-time Fourier transform (STFT) and TV-AR modeling, respectively. The STFT spectrum is computed with a Hamming window of 8 s, overlapping of 7 s, and nfft points set to 256. In this example, the TV-AR spectrum is estimated by using the same window length (8 s) and overlapping (7 s), and a model order $p = 10$, whose coefficients are computed by using the Yule-Walker method. It is important to note that for the spectral analysis applied to the FHR signal, we use different AR model parameters set according to the requirements for the FHR signal analysis, as will be explained in see Section 5.5.

As shown in Fig. 4.3(d), (e), and (f), AR modeling allows better spectral estimation compared with DFT. In the AR spectrum (blue), even for the case with the noise of amplitude

0.2, all the frequency components involved in Sim2 are correctly estimated, whereas, in the spectrum computed by DFT (red), only the components of $f = 0.5$ and $f = 1.0$ Hz can be recognized.

A stationary spectrum can illustrate the frequency components involved in the signal, but it does not provide information concerning the instant that these components occur. Therefore, in the presence of non-stationary signal characteristics, a stationary frequency analysis is no longer applicable. In order to analyze these characteristics, a time-variant spectral estimation is required. A time-variant spectral representation of the studied signals is displayed in the third and fourth rows of Fig. 4.3, for the STFT and TV-AR spectra, respectively. In each figure, the spectral information vs. time is represented on the axis of the ordinates and abscissa, respectively.

In this display, we can observe the behavior of the frequency components over time, where their time and duration can be easily recognized. Here, the advantages of TV-AR modeling compared with STFT, are clearly visualized. TV-AR offers a considerably higher spectral resolution, represented by a thinner yellow band describing the frequency components involved in the signal Sim2. Besides, as can be observed in Fig. 4.3(k) and (l), this method is more robust to noise compared with the STFT spectrum. Particularly, Fig. 4.3(i) shows that in the presence of noise of amplitude $e = 0.2$, the component of $f = 1.5$ Hz has practically disappeared in the STFT spectrum; nevertheless, in the TV-AR spectrum, it can be clearly identified.

4.1.3 Example 3: Signal decomposition

As explained in Chapter 3, the FHR signal can involve nonlinear and non-stationary characteristics associated with the physiological regulation mechanisms modulated by the ANS contributing to the FHR activity (Garabedian et al., 2017; Romano et al., 2016b). Therefore, the proposed spectral analysis applied to the FHR signal must consider these important characteristics, whose behavior can involve significant information of interest.

For a better understanding of this phenomenon, we present a basic example of time-variant frequency analysis and decomposition applied to a signal that involves components of non-stationary frequencies. For this purpose, we generate a signal (Sim3) of 80 minutes length sampled at 8 Hz, resulting from the sum of two components corresponding to cosine signals that increase their instantaneous frequency quadratically and white zero-mean noise of 0.1 of amplitude. The first component is generated with frequencies from 0.1 to 1.2 Hz, whereas the second one is generated with frequencies from 0.7 to 1.9 Hz, in the course of 80 s. Fig. 4.4(a), (b), and (c) show the first component, the second component, and resulting signal Sim3, respectively. In order to visualize the spectral information of the signal Sim3, we compute the TV-AR spectrum (explained in Section 4.1.2), which is displayed in Fig. 4.4(d). Here, we can observe the non-linear behavior of the frequency components involved in Sim3.

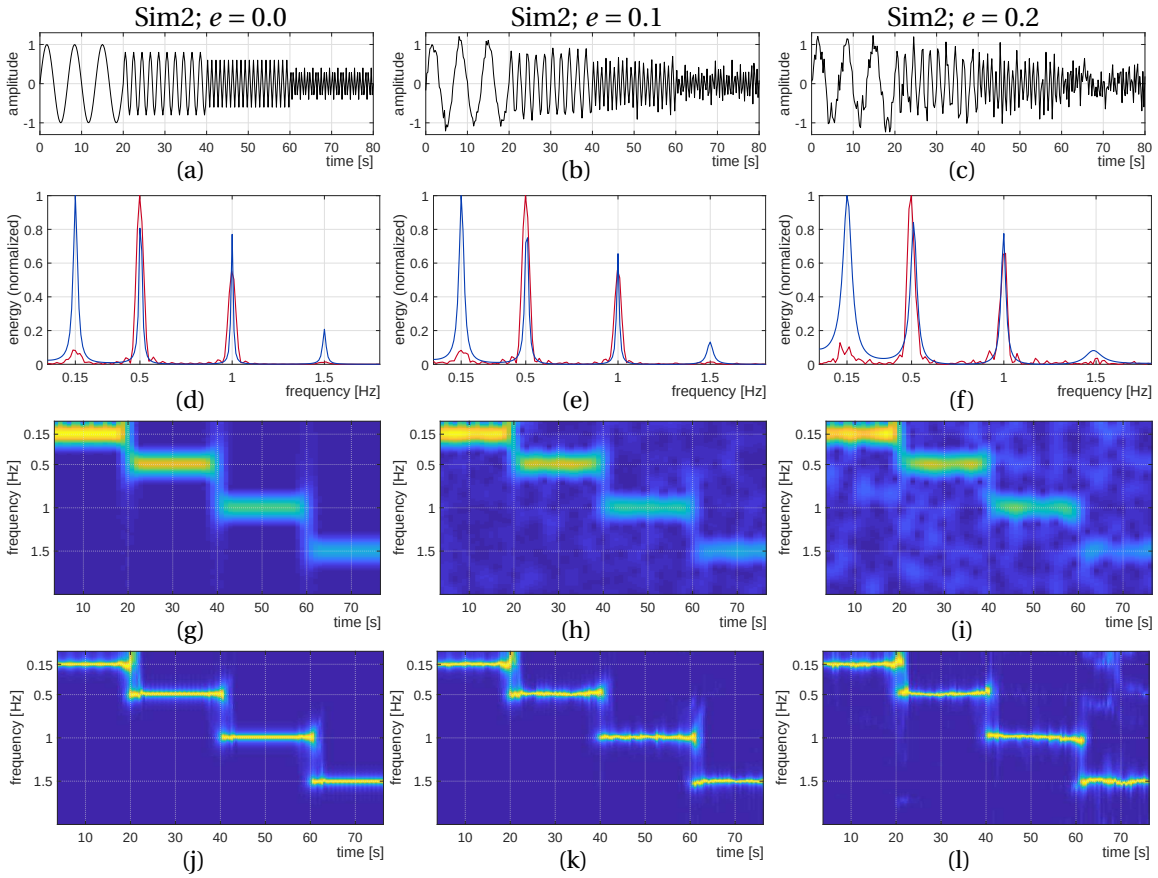


Figure 4.3: Example 2: Time-variant frequency analysis applied to a simulated signal (Sim2) resulting from four concatenated sinusoidal components of 20 s length; (a) Signal Sim2; $(0-20$ s): $[A, f] = [1.0, 0.15$ Hz$]$, $(20-40$ s): $[A, f] = [0.8, 0.5$ Hz$]$, $(40-60$ s): $[A, f] = [0.6, 1.0$ Hz$]$, and $(60-80$ s): $[A, f] = [0.4, 1.5$ Hz$]$; (b) Sim2 + noise $e = 0.1$; (c) Sim2 + noise $e = 0.2$. (d-f) Stationary spectrum computed by DFT and AR modeling from (a-c), respectively; (g-i) Time-variant spectrum computed by short-time Fourier transform (STFT) from (a-c), respectively; (j-l) Time-variant spectrum computed by TV-AR modeling from (a-c), respectively.

Table 4.1: Comparison between empirical mode decomposition (EMD), wavelet, and Fourier transform. Information extracted from Huang (2014b).

	Fourier	Wavelet	EMD
Basis	*a priori*	*a priori*	adaptive
Frequency	convolution: global uncertainty	convolution: regional uncertainty	differentiation: local certainty
Nonlinear	no	no	yes
Non-stationary	no	yes	yes
Theoretical base	theory complete	theory complete	empirical

For tracking and analysis of this type of nonlinear components, conventional signal processing methods no longer can be used. For example, a linear frequency filter can be suitable for separating the components involved in the signal of example 2 (see Fig. 4.3), because each component involves only one frequency invariant in time. However, this type of filter is not applicable to the signal Sim3, because, as can be observed in Fig. 4.4(d), both components involve frequencies in common (in the range between 0.7 and 1.2 Hz). For this reason, as a solution, before the time-variant spectral analysis, we perform a signal decomposition, which allows separating these signal components. For the signal decomposition, three techniques are commonly used, such as Fourier, wavelet, and EMD. The main characteristics of these decomposition techniques are presented in Table 4.1 (extracted from Huang (2014b)). As can be observed in this table, a suitable technique for the decomposition of nonlinear and non-stationary signals is the EMD. Besides, EMD does not require a priori known basis as the case of other traditional methods such as wavelet and Fourier transform-based decomposition (Huang, 2014b; Huang et al., 1998; Krupa et al., 2009, 2011; Wu and Huang, 2004).

Therefore, in this work, we decided to use EMD, particularly, a more advanced version called as ICEEMDAN. It is an adaptive method that allows decomposing time series into a finite number of intrinsic mode functions (IMFs). It provides information on the signal dynamics in the time-domain based on local properties of the signal data itself by the direct extraction of the energy associated with signal oscillations (Colominas et al., 2014). In this work, and following previous contributions (Fuentealba et al., 2019b,c,e), the IMFs are obtained by using a noise standard deviation Nstd= 0.03, and both the maximum number of sifting iterations (NI) and number realizations (NR) are set to 50.

An example of the performed decomposition operation is presented in Fig. 4.4(e) and (f), which show the IMFs decomposed from the signal Sim3, whose TV-AR spectra are shown in Fig. 4.4(g) and (h), respectively. Here, we can observe that the components involved in Sim3, can be separated into two IMFs, one corresponding to each component. Then, as

represented by their spectral information, these nonlinear frequency components can be easily tracked, represented by only one main frequency peak over time.

Conclusively, according to the information presented in examples 1, 2, and 3, as appropriate methods for the proposed time-variant analysis applied to the FHR signal, we select ICEEMDAN in combination with TV-AR modeling. A more in-depth explanation of the mathematical background and parameters involved in these methods is presented in Section 4.2 and 4.3, respectively. Besides, their application on real FHR data is performed in Section 5.5.

4.2 ICEEMDAN

The first version of this technique (EMD) have been initially proposed by Huang et al. (1998). It is an adaptive method that allows decomposing nonlinear and non-stationary time series. This method decompose the signal into a finite number of non-overlapping monotonic components termed as intrinsic mode functions (IMF) of different time scales. The main advantage of this method is that it depends on the local properties of the signal data itself by the direct extraction of the energy associated with signal oscillations. In consequence, EMD does not require a priori information as in the case of traditional methods such as wavelet and Fourier transform-based decomposition (Huang et al., 1998; Krupa et al., 2011; Wu and Huang, 2004).

The EMD has been used to analyze data in different fields of biomedical engineering (Krupa et al., 2011). Particularly, in CTG data preparation, EMD has been utilized for FHR estimation from Doppler Ultrasound signals by using EMD-kurtosis method (Al-Angari et al., 2017), and enhancement of the CTG signal by reducing signal artifacts (Krupa et al., 2009). For CTG analysis, EMD has been employed in several approaches such as assessment of the high-frequency information of FHR in different conditions of fetal activity (Ortiz et al., 2005); estimation of FHR baseline with analysis of fetal movements by using EMD and Kohonen neural network (Lu et al., 2014); investigation of temporal dynamics and complex interaction between FHR and UC signals (Saleem et al., 2019); study of the FHR signal components in order test the reliability of the EMD performance by using simulated FHR signals (Romano et al., 2016a); and FHR signal feature extraction and classification by using EMD and SVM (Krupa et al., 2011).

Since EMD was proposed by Huang et al. (1998), more sophisticated EMD methods has been developed in order to overcome different problems associated with the signal decomposition. The main drawback of the original EMD is the *mode mixing*, i.e., more than one mode of oscillation may contribute in one IMF, or one mode can spread across different IMFs. This phenomenon can occur when a mode has a very disparate amplitude, or more than one mode has very similar oscillations. In order to solve this problem, the Ensemble Empirical Mode Decomposition (EEMD) (Wu and Huang, 2009) was proposed. This method performs the EMD over an ensemble of the original signal plus white Gaus-

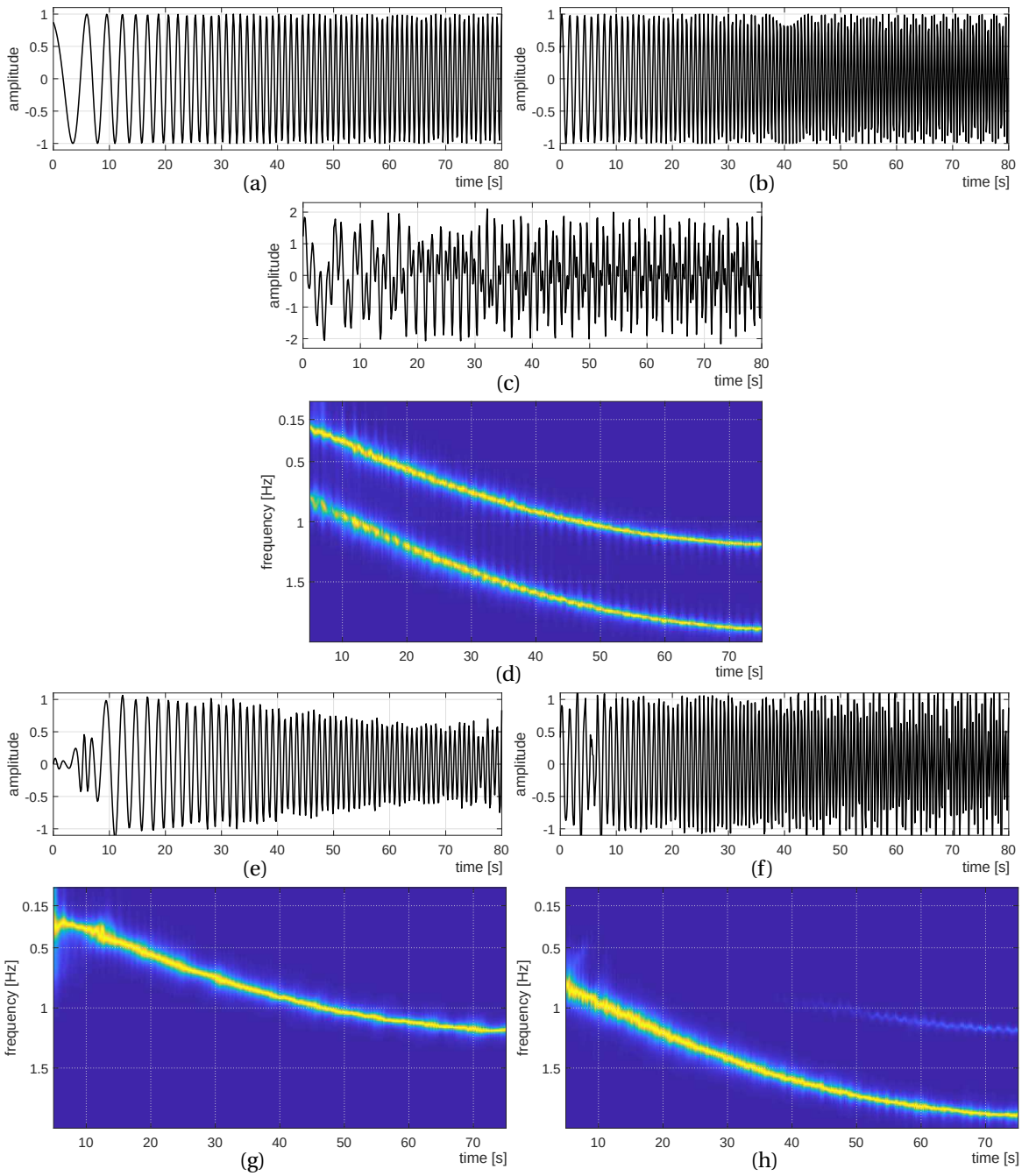


Figure 4.4: Example of time-variant spectral analysis and signal decomposition applied to a simulated signal (Sim3) involving two components of variable frequency; (a-b) Quadratic swept-frequency cosine signals (chirp signals) of 80 s length with frequencies in the range of $[0.1-1.2]$ and $[0.7-1.9]$ Hz, respectively; (c) Simulated signal Sim3 resulting from the sum of the components presented in (a) and (b); (d) TV-AR spectrum computed from Sim3; (e-f) Intrinsic mode functions (IMFs) decomposed by ICEEMDAN from signal Sim3. (g-h) TV-AR spectra computed from the IMFs plotted in (e) and (f), respectively.

sian noise, adding a different white noise series for each decomposition. Finally, the corresponding IMFs are averaged to obtain the final decomposition. The addition of this noise solves the mode mixing problems; however it generates other drawbacks. The final reconstructed signal contains residual noise, and also different realizations of signal plus noise may produce a different number of modes (Torres et al., 2011).

As an improved version of EEMD, the Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) method has been proposed by Torres et al. (2011). This method considers a unique residue for each mode independently from the noise realizations. The first CEEMDAN mode is computed exactly as in EEMD, so the first residue is obtained by the detrending operation between the original signal and this mode. Then, the first EMD mode is computed from an ensemble of the first residue plus different realizations of a particular noise. The second CEEMDAN mode is computed by averaging these EMD modes. Then, the second residue is obtained by the detrending operation between the first residue and the second CEEMDAN mode. This procedure repeats for each mode computation until the stopping criterion is reached.

The CEEMDAN method allows a negligible signal reconstruction error and solves the problem of the different number of modes, however it still has problems to be improved: (a) the CEEMDAN modes contain some residual noise, and (b) some *spurious* modes in the first stages of the decomposition, which makes that the signal information appear *later* than in EEMD (Colominas et al., 2014). In order to deal with these drawbacks, the ICEEMDAN method has been proposed by Colominas et al. (2014). This method allows a reduced amount of mode residual noise by computing the mode as the difference between the current residue and the average of the local means of each realization of signal plus noise. Likewise, it allows reducing the problem of the spurious mode by not considering the direct use of white noise for the extraction of the first IMF. Therefore, in this work, the ICEEMDAN method is used for the FHR signal decomposition.

For a more in-depth explanation of the ICEEMDAN method, please refer to Colominas et al. (2014).

4.3 Time-varying AR spectrum computation

In order to explain the concepts involved in the TV-AR spectrum computation, we first introduce the stationary version of AR modeling, and then the TV-AR modeling version.

4.3.1 Stationary AR modeling

As presented in Kay and Marple (1981) and Pardey et al. (1996), an AR model assumes that the current signal value $y[n]$ at sample number n in a data sequence, $y[1], y[2], \ldots, y[N]$, can be modeled as a linear combination of the p most recent sample values, $y[n-1], y[n-2], \ldots, y[n-p]$ and a white zero-mean noise $e[n]$ of variance σ^2, as described in Eq. (4.1).

In this equation, p is the model's order, whose value is generally much smaller than the sequence length N, $a_k\{k = 1,2,\ldots,p\}$ are the AR parameters, and n is the discrete-time index.

$$y[n] = -\sum_{k=1}^{p} a_k y[n-k] + e[n] \tag{4.1}$$

Considering that the AR model is computed for each IMF, Eq. (4.1) can be expressed as presented in Eq. (4.2), where i represents the index number of the IMFs.

$$\mathrm{IMF}_i[n] = -\sum_{k=1}^{p} a_{i,k}\mathrm{IMF}_i[n-k] + e[n] \tag{4.2}$$

The z-transform can be applied to Eq. (4.2) and the AR model transfer function can be represented by:

$$H_i[z] = \frac{IMF_i[z]}{E[z]} = \frac{1}{1+\sum_{k=1}^{p} a_{i,k}z^{-k}} \tag{4.3}$$

In Eq. (4.3), $IMF_i[z]$ and $E[z]$ are the z-transforms of the $\mathrm{IMF}_i[n]$ and the noise $e[n]$, respectively. Then, the AR spectrum can be computed by evaluating $H_i[z]$ around the unit circle in the complex plane, i.e., $z = e^{j2\pi f}$. Finally, the stationary AR spectrum representation is described by Eq. (4.4).

$$SAR_i(f) = \frac{\sigma^2}{\left|1+\sum_{k=1}^{p} a_{i,k}e^{-j2\pi fk}\right|^2} \tag{4.4}$$

4.3.2 Time-varying AR modeling

As explained above, the FHR signal dynamics of interest involve highly complex characteristics mainly associated with nonlinearities as a result of the physiological regulation mechanisms modulated by the ANS, which is highly variant in time. As explained in the example presented in Section 4.1.2, in the presence of such signal dynamics, the classical stationary version of AR modeling is no longer suited for their analysis. Therefore, in our proposed approach, we use the time-varying version of AR modeling, which can be appropriate for the FHR signal analysis since its parameters are now time-dependent. The TV-AR parametric model can be represented by Eq. (4.5), where the AR parameters $a_{i,k}(n)\{k = 1,2,...,p\}$ are now the time-dependent coefficients, which correspond to a set of values $a_{i,k}$ that is updated sample-by-sample n.

$$\mathrm{IMF}_i[n] = -\sum_{k=1}^{p} a_{i,k}(n)\mathrm{IMF}_i[n-k] + e[n] \tag{4.5}$$

From Eq. (4.5), analogously to Eq. (4.3), the TV-AR model transfer function represented by:

$$H_i[z,n] = \frac{IMF_i[z,n]}{E[z,n]} = \frac{1}{1+\sum_{k=1}^{p} a_{i,k}(n)z^{-k}} \tag{4.6}$$

Then, the TV-AR spectrum can be expressed as:

$$SAR_i(f,n) = \frac{\sigma^2}{\left|1+\sum_{k=1}^{p} a_{i,k}(n)e^{-j2\pi fk}\right|^2} \tag{4.7}$$

This last equation (Eq. (4.7)) allows performing a spectral time-variant analysis since the computed spectrum not only depends on the frequency-domain but also on the time-domain.

For a more in-depth explanation of the AR modeling method, please refer to Abramovich et al. (2007), Kay and Marple (1981) and Pardey et al. (1996).

All the signal processing techniques presented above will be employed for the analysis of the modulated dynamics involved in the FHR signal, as explained in the next chapters.

5

Studying the FHR dynamics by ICEEMDAN and TV-AR: Concepts

In order to answer the proposed research questions and reach the objectives of this thesis work (see Section 1.2), we propose to investigate the FHR signal dynamics resulting from the ANS modulation (signal dynamics of interest) and analyze their connection with the fetal health condition. The main idea is to track these dynamics of interest as observation of the process involved in the activity of the physiological behavior of the fetal compensatory mechanisms (physiological phenomenon explained in detail in Section 2.1). For this purpose, we propose the signal feature extraction strategy presented in Fig. 5.1. As can be observed in this diagram, firstly, a preprocessing step is applied to the FHR and UC signals. This step involves both signal outliers removal and signal interpolation in order to deal with artifacts and sensor contact interruption involved in the CTG signal acquisition. Secondly, we identify the FHR deceleration episodes based on a *progressive baseline* and UC events. As explained in Section 1.1, the FHR decelerations are considered as one of the main patterns that can involve significant information about fetal distress. However, at the same time, they are the most difficult patterns to interpret by the clinical staff (Hruban et al., 2015; Spilka et al., 2014b). Therefore, we propose to identify and study the deceleration episodes together as well as independent from the complete FHR signal to investigate their contribution to CTG assessment. Thirdly, the FHR signal is detrended and decomposed by ICEEMDAN, whose extracted IMFs are the input of the last signal processing step corresponding to the TV-AR spectral estimation. Finally, as shown in Fig. 5.1, the feature extraction operation is performed based on information in the time-domain as well as in the frequency-domain. All of these steps are described in this chapter, which explains the foundations behind the proposed strategy and its significance for our proposed approach. For a better explanation of the concepts involved in the proposed strategy, we explain it step-by-step making use of the CTG recording shown in in Fig. 5.2, which corresponds to the recording no. 1179 extracted from the CTU-UHB database (Chudáček et al., 2014).

5.1 CTG signal preprocessing

In signal processing, mainly in the area of biomedical signal processing, a preprocessing step is essential because an appropriate signal analysis depends significantly on the quality

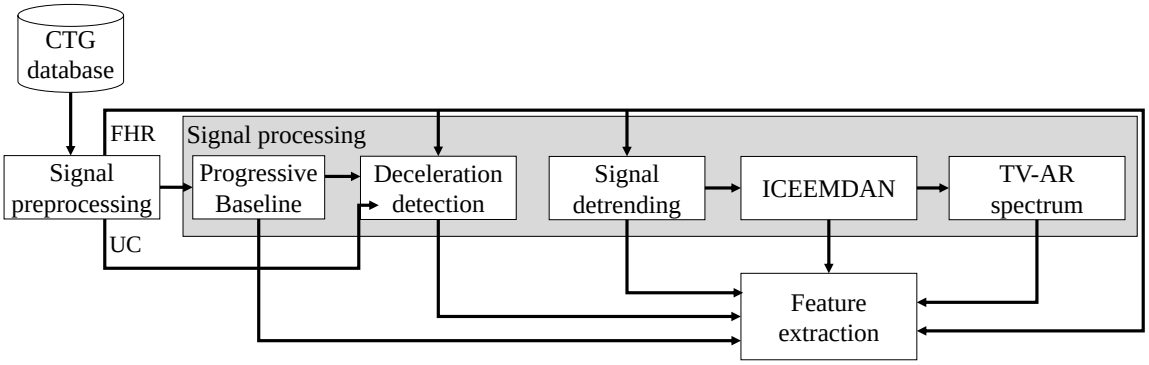


Figure 5.1: Block diagram of the strategy proposed for the CTG signal feature extraction.

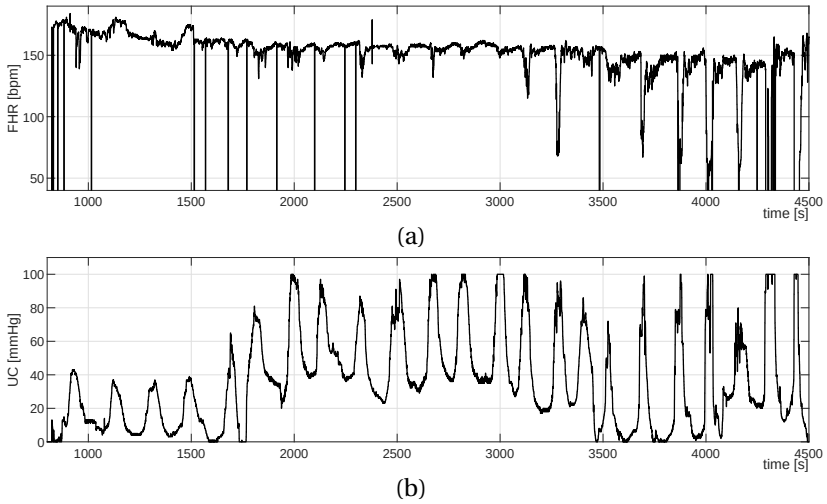


Figure 5.2: CTG recording no. 1179 extracted from the CTU-UHB database. (a) Raw FHR signal; (b) Raw UC signal.

of the input data. Particularly, the CTG signal acquisition usually involves different types of artifacts such as loss of data and signal outliers, mainly produced by the mother's movements and loss of sensor's contact that can temporarily interrupt the signal acquisition. In order to deal with these acquisition problems, we apply a preprocessing step for both the FHR and UC signals to prepare them for the subsequent analysis.

On the one hand, for the FHR signal, the artifact rejection method proposed by Spilka et al. (2013) is applied, which consists of two main steps: outliers removal and signal interpolation. In the outliers removal step, FHR signal values considered physiologically inconsistent in amplitude, i.e., values outside the range between 50 bpm and 210 bpm, are removed from the signal. Then, in the FHR signal interpolation step, loss of signal data corresponding to segments of length equal or less than 75 s are interpolated by using a Hermite spline method.

On the other hand, for the UC signal, loss of data less than 25 s are interpolated, and then the UC signal is filtered by a moving average filter of 15 s windows length. Note that this filtered UC signal is used for the decelerations characterization, as explained in Section 5.2.

Fig. 5.3 shows an example of the resulting preprocessed FHR and UC signals computed from the CTG signals presented in Fig. 5.2.

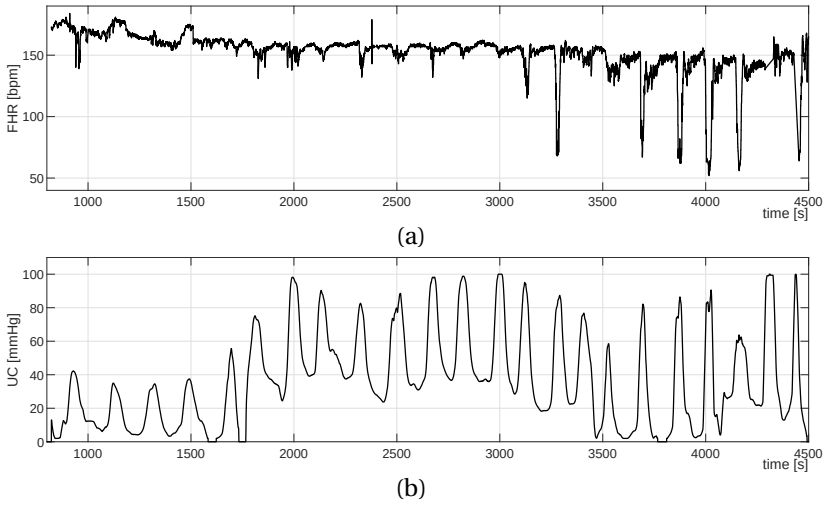


Figure 5.3: Preprocessed FHR and UC signals computed from the CTG recording no. 1179 presented in Fig. 5.2(a) and (b), respectively.

It is important to note that the obtained preprocessed FHR signal is used only for the FHR baseline estimation and identification of FHR deceleration episodes (explained below). The signal detrending and subsequent signal processing operations are performed by using the raw FHR signal only after the outliers removal step.

5.2 Progressive FHR baseline estimation

According to CTG guidelines (Ayres-de Campos et al., 2015), the FHR baseline is considered as the mean level of the most horizontal and less oscillatory FHR segments. In other words, the baseline is an imaginary trace representing a basal FHR value in the absence of accelerations or decelerations. In traditional CTG analysis, this pattern is particularly important because the assessment of other time-domain patterns, such as decelerations, rely on it (Agostinelli et al., 2017). Several algorithms have been proposed for the baseline estimation (Das et al., 2011; Kupka et al., 2006); however, considering that there is no precise definition of how to identify this pattern, none of them is optimal or can guarantee a correct FHR baseline assessment.

In this work, we propose to estimate a progressive baseline (PBL). For this purpose, first, a virtual baseline (VBL) is extracted by filtering the FHR signal using a nonlinear median filter for baseline wander extraction (Sameni et al., 2008). Based on a previous work (Fuentealba et al., 2017a), this filter is computed over a sliding window of 400 s length. Then, this VBL is used to define the low (T_L) and high (T_H) traces, which are represented by Eq. (5.1) and (5.1), respectively, where n is the sample number, and $\triangle$FHR is set to 10 bpm following Agostinelli et al. (2017). Fig. 5.4 shows the T_L and T_H traces plotted in dotted lines, which delimit the range of interest that will be considered for the PBL computation.

$$T_L(n) = \mathrm{VBL}(n) - \triangle\mathrm{FHR} \tag{5.1}$$

$$T_H(n) = \mathrm{VBL}(n) + \triangle\mathrm{FHR} \tag{5.2}$$

Then, the FHR data used for PBL estimation (FHR_{LH}) is described by Eq. (5.3), whose graphic representation is presented in Fig. 5.4 in black.

$$\mathrm{FHR}_{LH}(n) = \begin{cases} T_H(n) & \mathrm{FHR}(n) > T_H(n) \\ \mathrm{FHR}(n) & T_L(n) \leq \mathrm{FHR}(n) \leq T_H(n) \\ T_L(n) & \mathrm{FHR}(n) < T_L(n) \end{cases} \tag{5.3}$$

Finally, the PBL is computed by using the same nonlinear median filter used for the VBL extraction, but considering only data from the FHR_{LH} signal. The PBL is plotted in Fig. 5.4 in red.

5.3 Identification of FHR deceleration episodes

As explained in Section 1.1, the FHR signal decelerations are considered one of the most significant and challenging patterns to interpret by the clinical staff (Hruban et al., 2015;

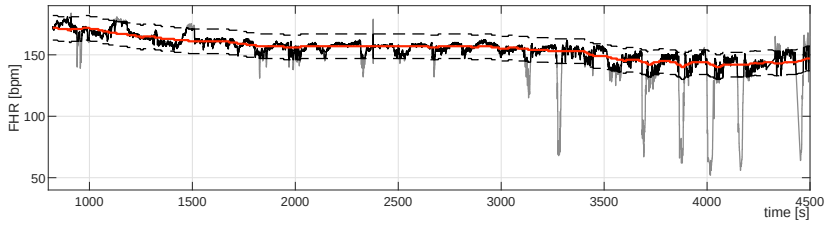


Figure 5.4: Preprocessed FHR signal (gray), FHR_{LH} (black), T_L and T_H traces (dotted lines), and Progressive baseline (PBL) (red).

Spilka et al., 2014b). Indeed, these patterns correspond to FHR variations produced as a consequence of the ANS modulation after a perceived oxygen insufficiency episode (Freeman et al., 2012). Therefore, we propose to identify and study the deceleration episodes together and independent from the complete FHR signal to investigate their contribution in CTG assessment.

In order to explain the strategy proposed for characterizing the FHR decelerations, we make use of the CTG recording of Fig. 5.5, which corresponds to a segment of 1500 s length arbitrary extracted from the preprocessed CTG signals presented in Fig. 5.3.

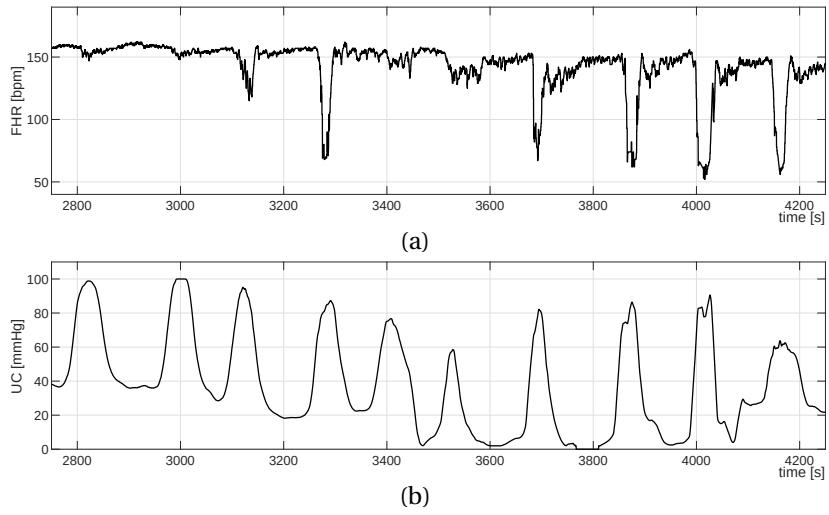


Figure 5.5: CTG segment of 1500 s length extracted from the preprocessed FHR and UC signals presented in Fig. 5.3.

In this work, following the analysis proposed in Fuentealba et al. (2017b), the deceleration identification is performed into two steps.

The first step allows the detection of *evident* deceleration episodes making use of the computed PBL and a *floating-line*. This floating-line corresponds to a signal trend that allows the tracking of the morphological shape of decelerations (see Fig. 5.6(a) in blue color), whose behavior involves nonlinear characteristics (Haritopoulos et al., 2016; Krupa et al., 2009). The floating-line is extracted by using the same nonlinear median filter applied for the PBL computation (see Section 5.2), but with a length of sliding window set to 10 s, whose value was determined as follows:

1. From the CTU-UHB database, we randomly selected ten CTG recordings.

2. For each recording, the FHR signal was filtered by the employed nonlinear filter using different sizes of the sliding window between 6 and 12 s length.

3. The computed preliminary floating-lines (seven for each recording) were superimposed on the corresponding FHR signal in order to examine which one tracks better the morphological characteristics of decelerations.

4. After a visual comparison, we concluded that a sliding window of 10 s length allows better tracking of the examined patterns.

The PBL is used as a threshold over the floating-line in such a way that floating-line segments of more than 15 bpm in amplitude and more than 15 s length (Ayres-de Campos et al., 2015) are identified as evident decelerations. Fig. 5.6(b) shows the PBL and the evident deceleration episodes, in red and black color, respectively.

The second step considers the UC events as stimuli in order to detect the episodes that were not recognized as evident decelerations but are certainly a response to a UC event. Those episodes (called as *UC-segments*) are identified according to the criteria proposed by Romano et al. (2006a), where the UC-segments start 7 s before the UC apex and end 50 s after it. For this purpose, first, UC apexes of significant intensity ($\geq 30\ mmHg$) are detected (see Fig. 5.6(c)). Then, when a UC event is detected, but an evident deceleration episode has not been identified in the first step, its corresponding UC-segment is considered, as can be observed in Fig. 5.6(d) (D01, D02, D05, and D06).

5.4 FHR signal detrending

According to the FIGO guidelines (Ayres-de Campos et al., 2015), the FHR variability corresponds to the average of the amplitude of fast FHR signal oscillations evaluated in stable signal segments. The FHR signal variability, together with the FHR decelerations, can involve relevant information related to the ANS activity as a result of potential hypoxia/acidosis. Therefore, they are considered as significant patterns to estimate the fetal condition. However, they are also considered the most difficult to assess by clinicians.

In this work, the estimation of such signal oscillations is based on the computation of a *detrended FHR signal*. This detrended signal is obtained by the subtraction operation between the preprocessed FHR signal and floating-line, which are plotted in Fig. 5.7(a) in black and blue color, respectively. Fig. 5.7(b) shows the detrended FHR signal, which represents the signal oscillations in relation to the floating-line.

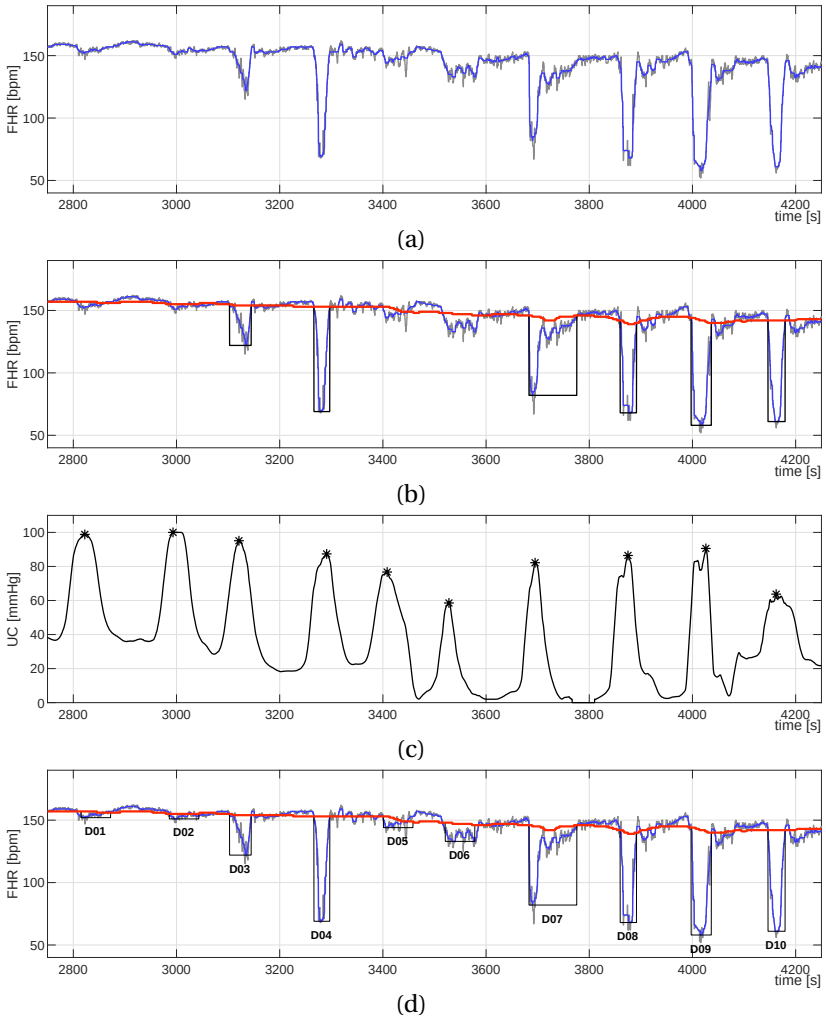


Figure 5.6: Characterization of deceleration episodes. (a) Preprocessed FHR signal (gray) and floating-line (blue); (b) PBL (red) and evident deceleration episodes (black); (c) Preprocessed UC signal and UC signal apexes (black dots); (d) Identified deceleration episodes (D01...D10) (black).

5.5 ICEEMDAN and TV-AR modeling implementation

As explained in Chapter 3, the FHR signal involves highly complex characteristics mainly associated with nonlinearities as a result of the physiological regulation mechanisms contributing to the fetal cardiac activity modulated by the ANS (Garabedian et al., 2017; Romano et al., 2016b). Under this concept, we postulate that if such characteristics are not previously decomposed or demodulated, the FHR signal can still involve high complexity dynamics, which are difficult to analyze. In order to study these characteristics, conventional signal processing methods no longer can be used because they cannot describe many processing conditions involved in this non-stationary physiological phenomenon.

In this perspective, we propose a new approach for processing the FHR signal by combining two methods: ICEEMDAN (Colominas et al., 2014) and TV-AR modeling (Kay and Marple, 1981), which are described in Section 4.2 and 4.3.1, respectively.

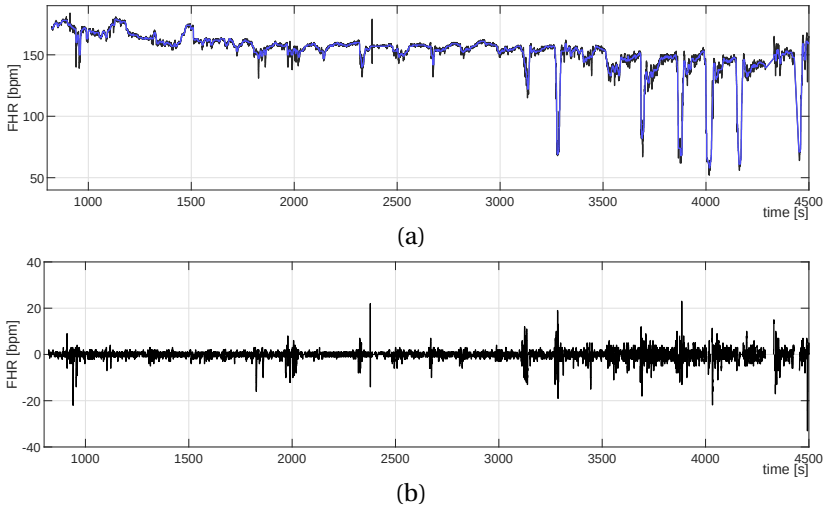


Figure 5.7: Signal detrending operation. (a) Preprocessed FHR signal presented in Fig. 5.3(a) and its computed floating-line, in black and blue colors, respectively; (b) Detrended FHR signal obtained by the subtraction operation of signals presented (a).

On the one hand, ICEEMDAN is an adaptive method that allows decomposing nonlinear and non-stationary time series into a finite number of IMFs. This method provides information on the signal dynamics in the time-domain based on local properties of the signal data itself by the direct extraction of the energy associated with signal oscillations. Considering that this decomposition operation has similar principles of signal demodulation, this method can be appropriate for extracting information about the FHR dynamics modulated by the ANS activity. In consequence, the main advantage of ICEEMDAN is its dependency on the data-driven mechanism, and it does not require a priori known basis as the case of other traditional methods such as wavelet and Fourier transform-based decomposition (Huang et al., 1998; Krupa et al., 2009, 2011; Wu and Huang, 2004).

On the other hand, we postulate that the dynamics described by each IMF can involve significant information not only in time-domain but also in the spectral-domain. For this reason, in order to study such information, we use TV-AR spectral-based analysis. In the spectral-domain, this information of interest is represented by the spectral density peaks, whose dynamical behavior describes the contribution of the frequency components variation over time. With this in mind and considering that the sampling frequency of the studied FHR signals is the only 4 Hz, the AR modeling can be appropriate for the spectral analysis of the extracted IMFs.

The classical stationary AR modeling (Kay and Marple, 1981) can effectively describe the peaks of a narrow-band power spectrum (Huang, 2014a) and requires only a fraction of the signal samples that are required by standard methods, such as the FFT, in order to obtain the same spectral resolution. It is a technique for time series analysis in which a mathematical model is fitted to a sampled signal. Therefore, it has several advantages for the FHR signal analysis since this parametric modeling provides a signal description simpler to analyze by a few model parameters. Also, it allows the extraction of quantitative spectral

parameters versus time, which are better suited for quantitative spectral analysis (Illanes and Haritopoulos, 2015).

Consequently, ICEEMDAN allows obtaining a less complex signal by decomposing its dynamics into the IMFs, which are better suited for parametrical modeling such as TV-AR modeling. Therefore, the time-variant frequency characteristics are easier to track because the spectral analysis is now based on tracking only one main frequency component of interest for each IMF over time.

In this work, and following Romano et al. (2006a), the model coefficients $a_k(n)$ are estimated by using a recursive least squares algorithm with a forgetting factor (λ). This forgetting factor corresponds to a model parameter related to the memory time considered in the studied phenomenon. Therefore, considering that our approach involves the analysis of the same physiological phenomenon performed by Romano et al. (2006a) (analysis of fetal ANS activity over time), we decided to employ the same forgetting factor set to .99. This value of λ is considered appropriate to describe not only the faster but also the slower signal dynamics of the frequency band of interest.

In order to choose an optimal AR model order p, we considered the characteristics of the IMFs and their expected spectral behavior. By definition, each IMF consists of a non-overlapping function, which is modulated in both amplitude and frequency (Colominas et al., 2014). Thus, we assume that in the spectral-domain, only one main frequency component, represented by one absolute maximum peak, contains the information of interest, which needs to be represented in the AR spectrum updated at each sample n.

With this in mind, for the selection of the model order p, we analyzed a set of 30 IMFs (selected randomly from different FHR signals). For each IMF, the AR spectrum was computed by using different AR model orders p in the range of 4 and 10. Then, we examined which model order offers a better spectral representation of the frequency component of interest, taking into account that only one and marked component should be distinguished. After this visual analysis, and following Fuentealba et al. (2018b), we decided to use a 6th AR model order p.

Fig. 5.8 shows an example of the decomposition operation of the detrended FHR signal plotted in Fig. 5.7(b) by using ICEEMDAN method. Each row depicts one IMF together with its corresponding TV-AR spectrum at the left and right sides, respectively. In this figure, we can observe the dynamics of the IMFs not only in the time-domain but also their behavior in the frequency-domain. This frequency information is represented by the TV-AR spectrum, where the x-axis is the time axis in seconds, and the y-axis is the frequency axis in Hz. The spectral energy is represented by a color map, where the blue and yellow colors correspond to the lowest and highest energy level, respectively. In the graph, the energy values are normalized between 0 and 1 for each sample n for better visualization of the frequency dynamics. It is important to note that the proposed features (presented in Section 5.6) are based on this spectral representation. For more example of the decomposition and TV-AR spectrum computation see Appendix A.1.

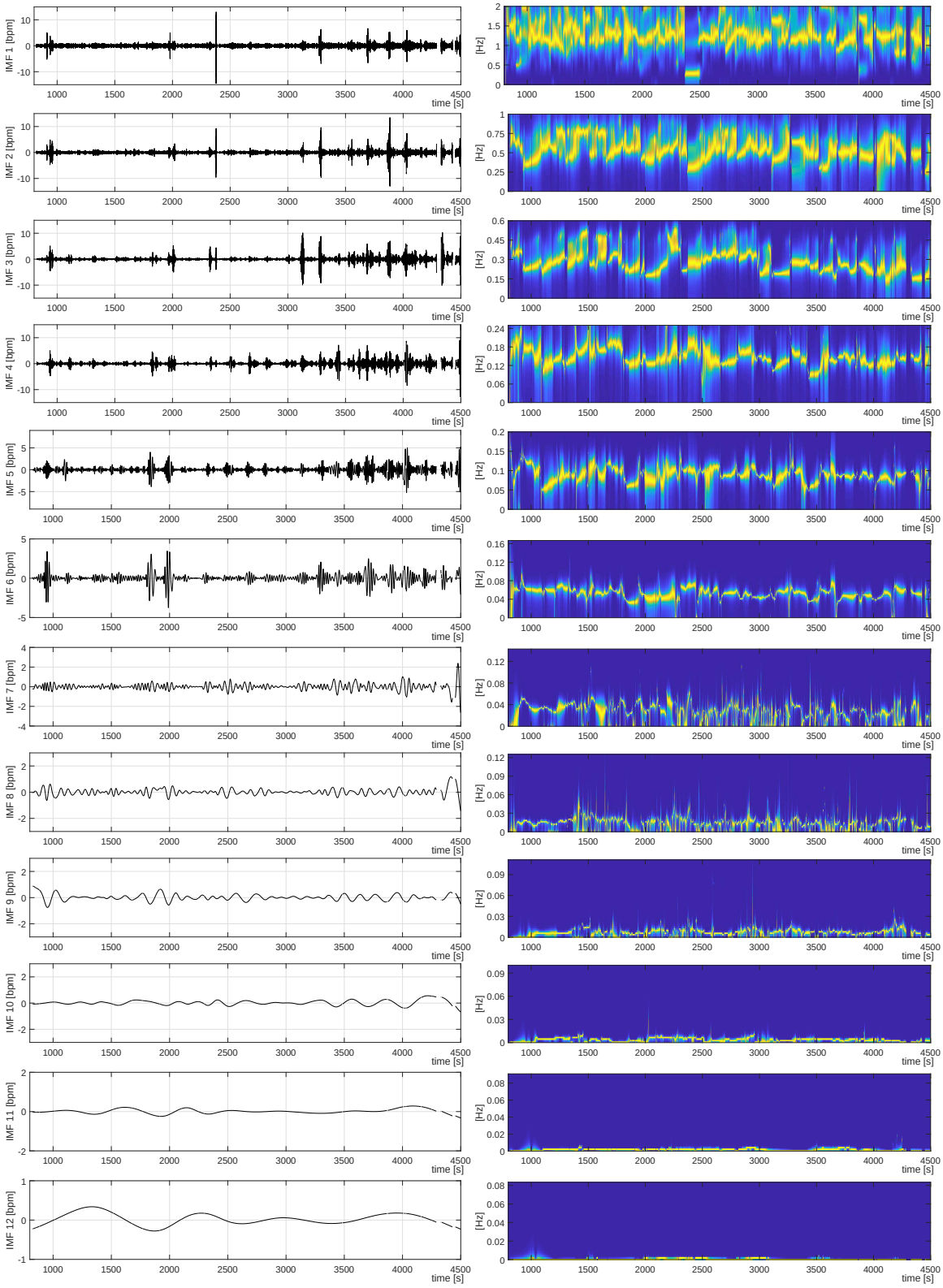


Figure 5.8: ICEEMDAN decomposition and TV-AR spectrum of FHR recording no. 1179; (left column) IMFs computed from the detrended FHR signal presented in Fig. 5.7(b); (right column) TV-AR spectrum of each IMF.

5.6 Definition of proposed features

This section presents a qualitative analysis related to the foundations on which the feature extraction operation is based. For a comprehensive explanation of the feature extraction operation, we make use of two representative cases extracted arbitrarily from the selected CTG dataset described in Section 6.2.1 (four more representative cases are presented in Appendix A.2). These cases are shown in Fig. 5.9, where the left and right columns correspond to examples of normal and acidotic fetal conditions, respectively. The first row shows the raw FHR signals for each case, which belong to the recordings no. 1427 and 2009 from the CTU-UHB database, respectively. The second row plots the preprocessed FHR signal, PBL, floating-line, and decelerations episodes, which are plotted in gray, red, blue, and black, respectively. The third and fourth rows show the detrended FHR signal and IMF_6, respectively. For this example, the IMF_6 was arbitrarily selected from those IMFs that involve spectral information in the frequency band between 0.03 and 0.15 Hz (frequency analysis of IMFs explained in Section 6.2.2 and represented by Fig. 6.2), which is mainly related to the fetal sympathetic ANS activity (as explained in Section 3.1.1). The fifth row exhibits the TV-AR spectrum computed from the IMF_6, whose spectral energy values were normalized between 0 and 1 for each sample n for better visualization of the frequency dynamics (instead of components energies). Finally, the sixth row shows an energy indicator (E) computed from the TV-AR spectrum, which allows examining the spectral energy changes involved in the studied signal dynamics over time. This indicator is calculated from the total frequency band ($0-2$ Hz) of the AR spectrum for each sample n as described in Eq. (5.4). In Fig. 5.9(e-h,k-l), the signal segments corresponding to deceleration episodes are highlighted in gray color, in order to differentiate them from resting periods.

It is important to note that for this qualitative analysis, only one IMF was considered. Nevertheless, for the subsequent feature extraction evaluation explained in Section 6, the first ten IMFs are analyzed, thereby considering the entire frequency band of interest (see Section 3.1.1).

$$E[n] = \sum_{f=0}^{2\,Hz} (S_{AR}[f,n]) \tag{5.4}$$

Results presented in Fig. 5.9 show that the representative cases exhibit spectral dynamics strongly variant in time (see the TV-AR spectrum in the fifth row). The spectral energy behavior associated with these dynamics differs between the normal and the acidotic case. As can be observed in Fig. 5.9(k-l), for the acidotic case, the spectral energy is generally lower than the normal case. Besides, the normal case exhibits important energy variations, whose amplitude is considerably higher than the acidotic case. This phenomenon can be related to the capacity of the fetal response modulated by the sympathetic ANS, which may decrease for acidotic fetuses (Romano et al., 2006a). In other words, the sympathetic path corresponding to a normal fetus might reflect higher activity in the FHR than the activity of an acidotic fetus.

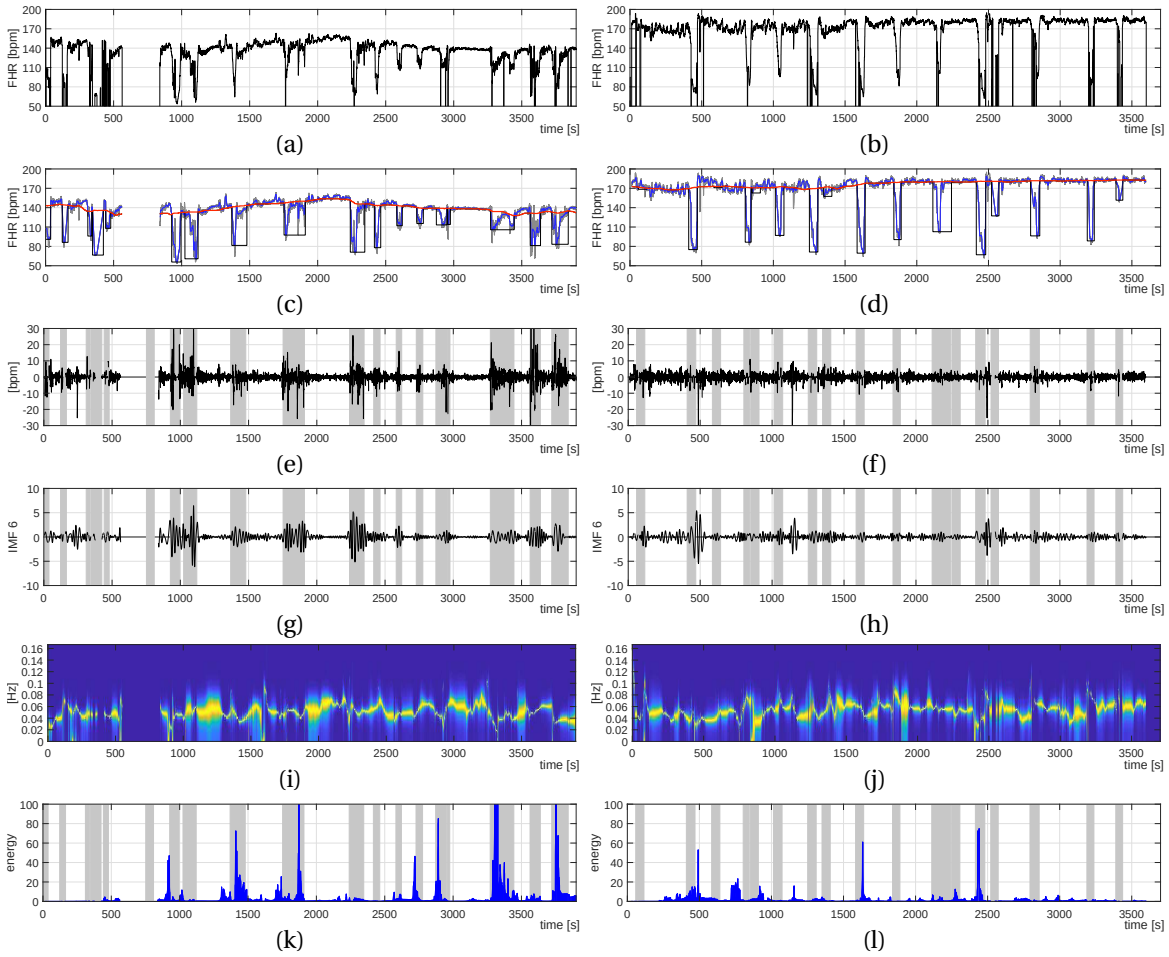


Figure 5.9: Representative cases; (a) raw FHR signal of recording no. 1427, pH = 7.37, BDecf = 1.34; (b) raw FHR signal of recording no. 2009, pH = 6.96, BDecf = 20.34; (c-d) preprocessed FHR signal (gray), PBL (red), floating-line (blue), and deceleration episodes (black); (e-f) detrended FHR signal; (g-h) IMF_6; (i-j) TV-AR spectrum computed from IMF_6; (k-l) E from (i-j).

Table 5.1: Significance (p-value) of the spectral features E_S, E_D, and E_R, obtained by the Wilcoxon rank-sum test. Values presented in second and third columns correspond to the 50th[25th−75th] percentiles of the data, and the fourth column shows the computed p-value for each feature.

Feature	Normal cases	Acidotic cases	Significance (p-value)
$E_{S\mu}$	5.09[3.14 − 7.66]	3.50[2.14 − 4.56]	.0213
$E_{S\sigma}$	2.06[1.18 − 3.78]	1.03[0.74 − 2.42]	.0079
$E_{D\mu}$	4.87[2.85 − 8.35]	2.67[1.52 − 4.22]	.0016
$E_{D\sigma}$	1.69[0.95 − 3.03]	0.72[0.39 − 1.64]	.0015
$E_{R\mu}$	4.76[2.98 − 7.31]	3.63[1.90 − 5.31]	.0708
$E_{R\sigma}$	1.77[1.01 − 3.42]	1.08[0.82 − 2.11]	.0277

In order to study whether the observed characteristics in these two examples are related to the fetal health condition, we examine their significance in the complete dataset. For this purpose, we calculate the features E_μ and E_σ, which correspond to the average and standard deviation computed from the indicator E.

According to the literature (Hruban et al., 2015; Spilka et al., 2014b), the CTG interpretation problems are mainly associated with the assessment of FHR decelerations, which certainly correspond to a cardiac response of the fetus modulated by the ANS (Freeman et al., 2012). Besides, recent research in biomedical engineering indicates that the UC activity has a graded effect on the FHR response (Feng et al., 2019; Sletten et al., 2016).

Therefore, we postulate that the spectral energy E computed in the complete signal (E_S), during decelerations (E_D), and during resting periods (E_R), describes different dynamics, whose behavior differs between a normal and an acidotic fetal condition. Therefore, the previous operation was performed independently for these three *terms*, i.e., we calculate the average and standard deviation for these three indicators. It is important to note that in this qualitative analysis, only those two features were computed. Nevertheless, as explained in Chapter 6, the evaluation of our approach, based on quantitative analysis, involves a larger set of features.

Once the features are extracted, a Wilcoxon rank-sum test (Gibbons and Chakraborti, 2011) is applied (for more information about the Wilcoxon rank-sum test, please refer to Section 6.2.3). This stochastic test allows us to examine the discriminant capability of the studied features under the hypothesis that in the feature dimension, the median value of the normal class data significantly (p-value < .05) differs from the median value of the acidotic class data. It is applied independently for E_S, E_D, and E_R. Table 5.1 and Fig. 5.10 show the statistical results and corresponding boxplots, respectively.

As can be observed in Table 5.1, most of the tested features are statistically significant (p-value < .05), i.e., in the feature dimension, the median value of the normal class data is significantly higher than the median value of the acidotic class data. These results open

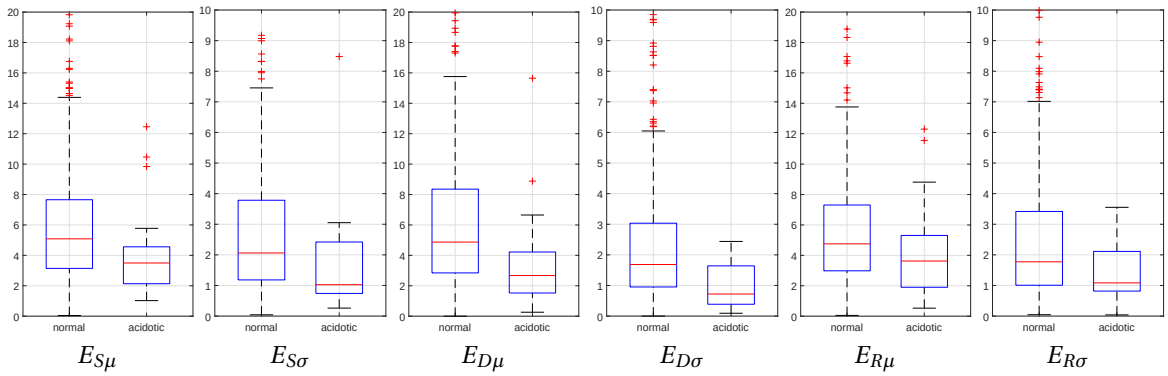


Figure 5.10: Boxplots of features presented in Table 5.1. The box borders correspond to the 25th and 75th percentiles of the feature data. The red line represents the median value of the data, and red crosses are the outliers.

perspectives for the extraction of a larger set of features considering all the IMFs involved in the complete frequency band of interest, and employing more statistical coefficients. This analysis is presented in Chapter 6, where the features are evaluated based on their performance in CTG classification.

6

CTG feature selection: Evaluation strategy design

After the feature extraction step, in this chapter, we proceed to evaluate the proposed features based on their performance in CTG classification. First, Section 6.1 presents the selection criteria and demographic information of the CTU-UHB database, which corresponds to the CTG data used in this work. Then, Section 6.2 explains in-depth the steps involved in the proposed evaluation strategy, such as the class formation criteria, feature computation, feature elimination, and feature classification performance. In summary, this chapter provides a detailed explanation of the concept and computer-based techniques behind the proposed feature evaluation strategy, whose results are presented in Chapter 7.

6.1 CTG database

For the CTG signal analysis and evaluation proposed in this work, the CTU-UHB Intrapartum Cardiotocography Database (Chudáček et al., 2014) is used. This database, from the Czech Technical University (CTU) in Prague and the University Hospital in Brno (UHB), is the first open-access database for research purposes on CTG signal analysis. It consists of 552 CTG recordings, which provide FHR and UC signals, both sampled at 4 Hz. These 552 recordings were selected from 9164 recordings collected between 2010 and 2012 at the UHB. The selection criteria used for this database were based mainly on the following parameters:

- Singleton pregnancy.
- No a priori known developmental defects.
- Gestational age > 36 weeks.
- Duration of stage 2 of labor ≤ 30 minutes.
- FHR signal quality: > 50% of FHR data were available in each segment of 30 minutes length.
- Available information on the pH biochemical parameter obtained from the umbilical arterial blood sample.
- Majority of vaginal deliveries (only 46 cesarean section deliveries included).

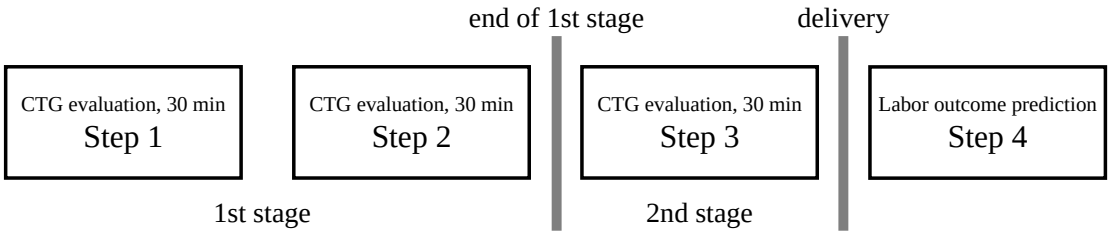


Figure 6.1: Workflow of the annotation procedure performed by steps, as presented by Hruban et al. (2015).

For all recordings, the CTU-UHB database includes physiological information concerning to the maternal data as well as fetal data, which are listed in the following:

- Maternal data: age; parity; gravidity.
- Delivery data: type of delivery (vaginal; operative vaginal; cesarean section); duration of delivery; meconium-stained fluid; type of measurement (i.e., ultrasound or direct scalp electrode).
- Fetal data: sex; birth weight.
- Fetal outcome data: analysis of umbilical artery blood sample (pH; pCO_2, partial pressure of oxygen (pO_2); BE and BDecf); Apgar score; neonatology evaluation (i.e., need for O_2 and seizures).

In addition to the information described above, the CTU-UHB database provides the evaluation of nine experienced obstetricians, four annotations for each CTG recording. As can be observed in Fig. 6.1, the first and second annotations correspond to the evaluation of two windows of 30 minutes length in the first stage of labor. The third annotation corresponds to a window in the second stage of labor. After that, in the fourth annotation, clinicians predict the labor outcome classifying their prediction into three possible ranges of pH values. For all the steps, they classify the recording as normal, suspicious, pathological, or uninterpretable.

Table 6.1 provides the main demographic information and statistics of the CTU-UHB database, which include physiological data of the mother as well as the outcome parameters of the newborn. For further information about the CTU-UHB database, please refer to Chudáček et al. (2014). It is available at https://physionet.org/physiobank/database/ctu-uhb-ctgdb/.

Table 6.1: Demographic information and statistics of the CTU-UHB database.

Parameter	mean	min	max	50th[25th−75th] percentiles
pH	7.23	6.85	7.47	7.25[7.17 − 7.30]
BDecf (mmol/l)	4.60	-3.40	26.11	4.13[2.49 − 6.15]
BE	-6.38	-26.8	-0.20	-5.60[-8.03−-4.00]
Apgar 1	8.26	1	10	8[8 − 9]
Apgar 5	9.06	4	10	10[9 − 10]
Maternal age (years)	29.8	18	46	30[27 − 33]
Parity	0.43	0	7	0[0 − 1]
Gravidity	1.43	1	11	1[1 − 1]
Gestational age (weeks)	40	37	43	40[39 − 41]
Neonate's weight (g)	3408	1970	4750	3410[3100 − 3700]
Neonate's sex (F/M)	259/293			

6.2 Feature evaluation strategy

6.2.1 Class formation criteria for dataset selection

The evaluation of our proposed features is based on their performance in CTG classification making use of data extracted from the CTU-UHB database presented above. From this database, we select a data subset based on a formation criteria involving two classes: normal and acidotic conditions, which allows obtaining comparable results with related studies (as will be explained in Section 7.3). According to the literature (Kuehnle et al., 2016; Maso et al., 2015; Spilka et al., 2013; Zhao et al., 2018; Zourabian et al., 2000), arterial blood gas sampling is considered as a gold standard measure for outcome evaluation. In umbilical cord test sampling, at birth, values of pH < 7.05 and BDecf ≥ 12 commonly indicate a fetal metabolic acidosis, whereas values of pH > 7.20 and BDecf < 12 indicate a normal fetal health condition (Kumar et al., 2016). In this work, those outcome parameters are used to assign the fetal condition into one of the two classes. Under these class formation criteria, 372 recordings were selected from the database, 354 labeled as examples of normal fetuses, and 18 labeled as examples of acidotic fetuses. Hence, the features are evaluated based on the CTG classification performance computed according to this class formation criteria. As shown in Table 6.1 for the complete database, Table 6.2 shows the main demographic information and statistics of the CTG dataset selected according to the mentioned class formation criteria.

6.2.2 Feature computation

The feature computation performed in this section is based on the concepts explained in Section 5. There, we had explained the foundation of the proposed features by an example based on the IMF_6. In this section, the feature extraction operation involves the study of a larger set of features. The idea is to include not only the IMF_6 but all the IMFs involving

Table 6.2: Demographic information and statistics of the selected CTG data used in this work.

Parameter	mean	min	max	50th[25th−75th] percentiles
pH	7.28	6.85	7.47	7.28[7.24 – 7.32]
BDecf (mmol/l)	3.77	-3.40	26.11	3.32[1.88 – 4.83]
BE	-5.27	-26.8	-0.20	-4.60[-6.00−-3.40]
Apgar 1	8.62	2	10	9[8 – 9]
Apgar 5	9.30	4	10	10[9 – 10]
Maternal age (years)	29.7	18	43	30[27 – 33]
Parity	0.48	0	5	0[0 – 1]
Gravidity	1.46	1	11	1[1 – 1]
Gestational age (weeks)	39.96	37	43	40[39 – 41]
Neonate's weight (g)	3396	1970	4750	3450[3100 – 3700]
Neonate's sex (F/M)	181/191			

dynamics in the FHR frequency band of interest (see Section 3.1.1). In order to determine these IMFs of interest, we first compute the TV-AR spectrum for each FHR signal of the database (operation explained in Section 5.5). Second, for all IMFs, we compute the frequency of the main component (f^c), which corresponds to the frequency value at which the spectrum presents the maximal energy (energy of the main component) level for each sample n. This indicator is computed as described in Eq. (6.1), where SAR is the TV-AR spectrum, and E^c is the energy of the main component.

$$E^c[n] = SAR[f, n], f = f^c \tag{6.1}$$

Then, making use of the f^c indicator, we calculate the average of the main frequency involved in each IMF, performed on the complete database. The results of this operation are presented in Fig. 6.2, where each point corresponds to the average of f^c for each IMFs and each FHR recording. As can be observed in the abscissa axis, the number of studied IMFs is 12, which corresponds to the minimum number of IMFs computed from the employed CTG database, thereby ensuring that all the recordings are considered for this analysis. For better visualization of the frequency information, we employ a semi-log chart, where the ordinate axis is in a logarithmic scale of base ten. Considering that the complete FHR frequency bands involve values between 0 and 1.0 Hz, and according to the information presented in the figure, we arbitrarily selected the first 10 IMFs.

Once the number of the studied IMFs has been determined, we compute the TV-AR spectrum for each IMF. For the feature computation performed in this section, in addition to the indicator of total spectral energy (E) (Eq. (5.4)) explained in Section 5.6, we compute the energy of the main component (E^c) and frequency of the main component (f^c) (see Eq. (6.1)). Then, for each indicator, we calculate seven statistical coefficients that have been commonly used in FHR signal feature extraction. This set of coefficients corresponds to the arithmetic mean (μ), median (M), standard deviation (σ), mean absolute deviation (mad), root mean square (RMS), sample entropy (SampEn) (Richman and Moorman, 2000),

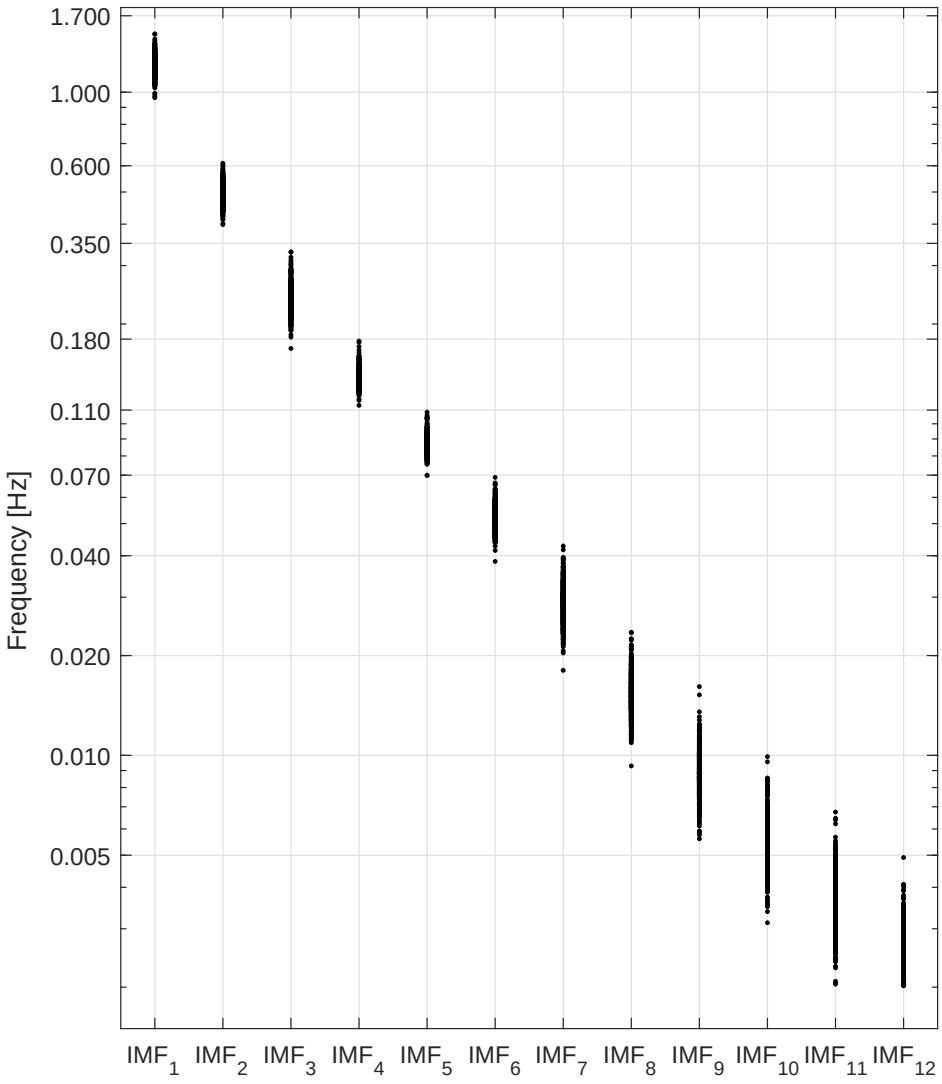


Figure 6.2: Average of frequencies involved in the IMFs computed on the entire CTG database.

and approximate entropy (ApEn) (Pincus, 1995). These last two coefficients allow quantifying the signal data complexity, which, according to Pincus and Viscarello (1992), might be related to the interaction of the sympathetic and parasympathetic ANS activity of the fetus. Besides, features of entropy have shown better performance in classification compared with the conventional CTG feature extraction (Fergus et al., 2018; Ferrario et al., 2006; Spilka et al., 2012). For the computation of the features of entropy, following the approaches proposed by Pincus (1995), Spilka et al. (2009) and Spilka et al. (2014c), we employ an embedding dimension $m = 2$ and a tolerance $r = 0.2 \times \sigma$ (σ corresponds to the standard deviation).

In addition to the features described above (in the sequel denoted as the *modal-spectral* features), we propose to include another set of features in the time-domain (denoted as the *conventional* features), which have shown relatively satisfactory results in CTG classification (Fuentealba et al., 2019b; Krupa et al., 2011). For the extraction of these conventional features, we employ the same seven statistical coefficients described above but now computed from the IMFs in the time-domain, raw FHR signal, PBL, and detrended FHR signal.

It is important to note that the detrended FHR signal and the IMFs in the time-domain are indicators whose data are centered in zero (see Section 5.4 and 5.5, respectively). Therefore, central tendency measures are not informative for their analysis; in consequence, the arithmetic mean, and median coefficients are not considered for such indicators. Besides, as explained in Section 5.6, the main idea is to extract features from the complete FHR signal (S-features), during decelerations (D-features), and resting periods (R-features), independently. However, considering that the deceleration episodes can range from 15 s (60 samples) length, the SampEn and ApEn coefficients are not appropriate for the analysis during decelerations because their computation is applicable only for segments longer than 100 samples (Pincus, 1995), or 200 samples (Yentes et al., 2013). Therefore, the entropy coefficients are not considered for D-features.

In complete, S-features and R-features consist of 210 modal-spectral and 69 conventional features, independently, whereas S-features include 150 modal-spectral and 63 conventional features. After the feature extraction operation, the data in the feature dimension are standardized using z-score based normalization. Consequently, before the subsequent steps, the data of each feature have a mean value and a standard deviation equal to zero and one, respectively.

6.2.3 Feature elimination

The feature elimination step consists of removing the weakest features from the complete set of extracted features. As a result of this operation, only features presenting a significant discriminant capability are considered for the analysis.

In this work, following several approaches proposed for CTG analysis (Georgoulas et al., 2017; Rivolta et al., 2019; Spilka et al., 2016a; Yu et al., 2017), we apply a Wilcoxon rank-sum test (Gibbons and Chakraborti, 2011) as an appropriate method for feature elimination. Its main advantage is that it does not require a priori assumption of data normality as required in the case of a t-test (Kwak and Park, 2019). This stochastic test allows us to examine the discriminant capability of the extracted features and determine the statistically significant features (p-value < .05). It is applied independently for each feature under the hypothesis that in the feature dimension, the median of the acidotic data distribution significantly differs from the median of the normal data distribution. Therefore, all the features that present a statistically significant difference (p-value < .05) are selected, and the others are excluded from the analysis.

6.2.4 Feature classification performance

For the evaluation of the proposed features, following Georgoulas et al. (2006), they are first grouped by their category, denoted in the sequel as the *feature sets*. Each feature set consists of the statistical coefficients computed for each indicator (e.g., FHR signal, PBL, and $\text{IMF}_1 \ldots \text{IMF}_{10}$), and separately for S-features, D-features, and R-features.

Then, the feature sets performance is evaluated by using the proposed strategy presented in the diagram of Fig. 6.3 (Fuentealba et al., 2019c). It consists of two main loops (*loop A* and *loop B*). Loop A is based on 5-folds stratified cross-validation using a computer-based classifier. Loop B calculates the classification performance based on the average across across 300 iterations of the complete process involved in Loop A. First, the data in the feature dimension are randomly split into five non-overlapping folds and keeping the original distribution of normal and acidotic classes. Then, for each iteration of Loop A, one different fold is used as a testing data subset, while the other remaining four folds are used as a training data subset. In order to deal with the redundant information that the studied features may convey, for the training subset, we apply a feature preprocessing step based on principal component analysis (PCA). This statistically based technique allows obtaining attributes that are considerably less correlated to each other and preserving as much of the significant information as possible (Georgoulas et al., 2006). In order to attenuate the potential bias generated by the imbalanced input data to the computer-based classifier, we employ the adaptive synthetic sampling (ADASYN) technique computed using 5 k-neighbors. In view that the selected CTG dataset consists of 354 and 18 observations for the normal and acidotic classes, respectively (see Section 6.2.1), we apply an oversampling of the minority class by using a factor of 19. The main advantage of ADASYN compared with conventional methods is that it can generate synthetic data in the feature dimension for oversampling the imbalanced class by considering not only the data minority class but also the nearest data of the majority class. For more in-depth information about the ADASYN technique, please refer to He et al. (2008).

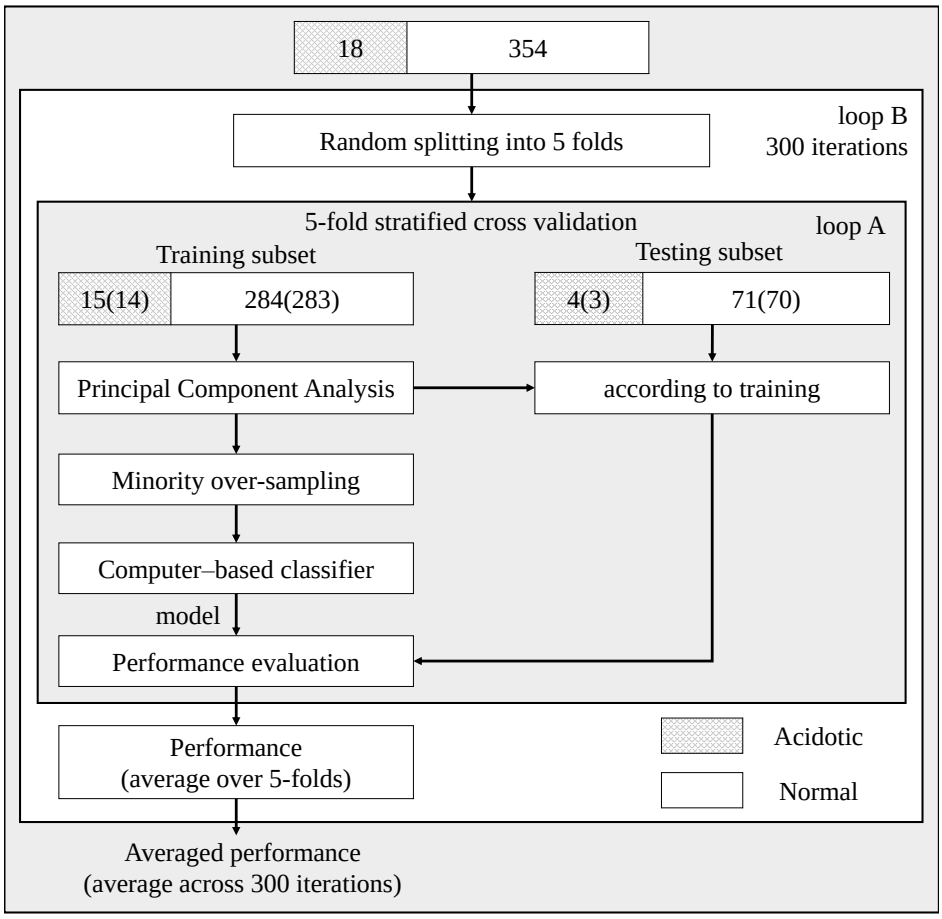


Figure 6.3: Proposed strategy for the features performance evaluation.

Note that the PCA is computed only from the training data subset, and then the testing data subset is transformed accordingly (see Fig. 6.3). Likewise, for a correct evaluation of the features, ADASYN is computed only in the training data subset, i.e., the testing subset does not include synthetic data, only real data.

Machine learning classifiers

For the evaluation of our proposed features, we selected three machine learning classifiers: SVM, Linear Discriminant Analysis (LDA), and k-Nearest Neighbors (k-NN), which have been commonly used in CTG classification (Georgoulas et al., 2017, 2006; Huang and Hsu, 2012; Krupa et al., 2011; Spilka et al., 2009, 2012). It is important to note that the main objective of this approach is not the selection of the best classifier, but rather the performance evaluation of our proposed features based on the combination of ICEEMDAN and TV-AR modeling. Therefore, these three classification methods, usually employed in CTG classification for discriminating two classes, can provide representative results that are comparable with related works that involve a similar type of classifiers.

The employed machine learning classifiers were implemented in Matlab® environment version 2018b. A brief explanation of their characteristics is presented below:

- A SVM classifier consists of a supervised learning model based on hyperplanes in the feature space, whose dimension depends on the number of features. It is based on dividing the space into two half subspaces, minimizing the empirical error and maximizing the margin between the nearest instances and the hyperplanes (Cristianini et al., 2000). The employed SVM classifier was implemented by using a linear kernel function and a penalty factor for classification $C = 1$. For more detailed information about SVM, please refer to Cristianini et al. (2000).

- LDA is a simple but powerful technique based on the difference between classes compared to a linear combination of the most significant features. This combination depends on the observations of predefined groups, which is used to predict the new observations. This widely used statistical method is based on a linear projection of the feature vector onto an optimal vector, whose classification is performed in this feature space. This operation allows minimizing the variance of each class while maximizing the class separation represented by the mean of each class. A more detailed explanation of the statistical operation involved in this technique can be found in Mika et al. (1999).

- k-NN is a nonparametrical method, whose classification technique is based on the k closest points or set of points in the feature space. The k-NN uses distance metrics, usually Euclidean distance, to find the optimal *neighborhood* of attributes in relation to the class labels of the training data. In this work, in order to determine an appropriate number of k nearest neighbors, we experiment with different values of k such as 1, 3, 5, 7, 9, and 11. Then, the k number achieving the highest classification performance is selected. As a result, we found that the highest classification performance was achieved by using 7-nearest neighbors (the classification results are explained in Chapter 7). For a detailed explanation of the k-NN classifier, please refer to Cover et al. (1967).

Performance evaluation

As explained in Section 6.2.1, the selected dataset consists of imbalanced data, i.e., the number of acidotic cases is considerably lower than than the normal cases. As explained above, in order to deal with the class imbalance, we employ ADASYN to create synthetic data for the minority class of the training data subset. However, the testing subset includes only real data, which is necessary for a correct feature evaluation; therefore, the testing subset still involves imbalance data. In this context, as a measure of classification performance, we employ the geometric mean Eq. (6.2). This statistical metric has been

commonly used as an indicator of performance in CTG classification because it is considered as an appropriate quality metric (QI) for imbalanced data (Georgoulas et al., 2006; Warmerdam et al., 2016b).

$$QI = \sqrt{Se \cdot Sp} \tag{6.2}$$

This quality metric is based on the statistical metrics of sensitivity (Se) and specificity (Sp), obtained as a result of the classification step. The metrics of Se and Sp are defined in Eq. (6.3) and (6.4), respectively. In these equations, TP and FP correspond to the number of true and false predicted as acidotic cases, respectively. Analogously, TN, and FN correspond to the number of true and false predicted as normal cases, respectively.

$$Se = \frac{TP}{TP + FN} \tag{6.3}$$

$$Sp = \frac{TN}{TN + FP} \tag{6.4}$$

The main idea behind the feature classification performance is to, firstly, select the feature set or a combination of them (denoted as the *combined set*) that achieves the highest QI as an indicator of classification performance. Then, based on that information, examine the selected feature sets in order to study their contribution according to their category. For this operation, we propose the selection strategy described in Fig. 6.4.

In the diagram of Fig. 6.4, N is the number of features sets to be tested, and the performance evaluation block represents a function that returns the set from the input sets that achieves the best classification performance. This function corresponds to the performance evaluation strategy described in Fig. 6.3. The proposed feature selection strategy first computes the classification performance of the N feature sets independently. Second, the feature set achieving the best performance is selected as the combined set 1 ($comb\ set_1$). Then, in the main loop, the $comb\ set_1$ is tested in combination with each remaining feature set independently, whose operation is described by the inner loop of N-k iterations, i.e., as a first iteration we have N-1 sets (*curr_sets*) to be tested. Now, the *curr_set* achieving the best performance is selected as the $comb\ set_2$. The $comb\ set_2$ is tested together with each remaining feature set independently (N-2 *curr_sets* to be tested), and the *curr_set* achieving the best performance is selected as the $comb\ set_3$. The described process repeats until all the combined sets have been tested.

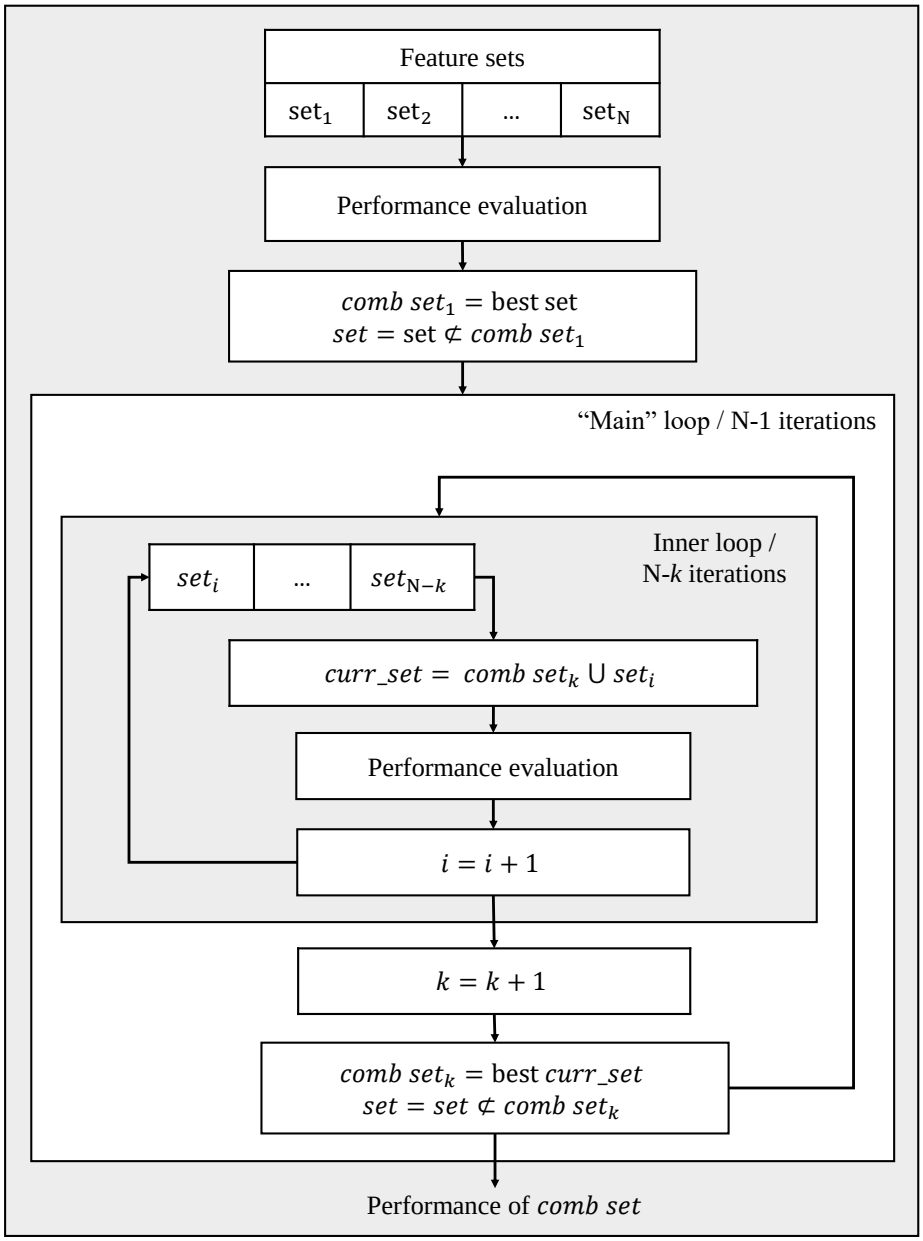


Figure 6.4: Proposed strategy for the feature sets selection based on their classification performance.

7

Results and discussion

This chapter explains in detail the main results obtained in this thesis work, based on the proposed methodology explained in Chapter 6. It is divided into three main sections. Section 7.1 focuses on evaluating the contribution of modal-spectral features compared with conventional features. Section 7.2 involves an independent analysis of FHR decelerations and resting periods in order to investigate their contribution in CTG classification compared with the analysis of the complete FHR signal segment. Finally, Section 7.3 concludes this chapter by providing a discussion concerning the significance of the obtained results, their contribution to CTG signal assessment, and their application in real scenarios.

7.1 Contribution of the modal-spectral vs. conventional features

This section focuses on studying the contribution of the modal-spectral features based on the performance evaluation strategy presented in Section 6.2. For this purpose, we make use of the selected CTG dataset defined in Section 6.2.1, and the features are extracted from the complete FHR segment (S-features) (feature extraction operation explained in Section 6.2.2). First, we define an optimal informative FHR signal segment, where the features can be extracted. Second, based on the strategy explained in Section 6.2.2 and Section 6.2.3, we perform the feature computation and feature elimination in order to obtain significant features, which are grouped into feature sets. Then, we evaluate the performance of the feature sets based on the strategy proposed in Section 6.2.4 by using the three machine learning classifiers presented in Section 6.2.4. It is important to note that the study performed in this section is based only on S-features. Nevertheless, in Section 7.2, we also include the features extracted during decelerations (D-features) and during resting periods (R-features) in order to evaluate their contribution in CTG classification.

7.1.1 Selection of an informative FHR signal segment

The feature evaluation usually involves the study of a particular CTG segment extracted from the recordings. Considering that the UC activity increases towards delivery, in the last minutes of labor, major changes in fetal condition can occur (Spilka et al., 2012). Therefore, it is commonly proposed in the literature the extraction of an informative segment, as close as possible to the delivery, where features can be computed (also called *epoch*). However,

there is no precise definition concerning an optimal epoch for the FHR signal analysis, whose selection depends mainly on the type of analysis (e.g., time-invariant, time-variant, short term, or long term analysis) (Haritopoulos et al., 2016).

For the evaluation of the studied modal-spectral features, we propose to estimate an optimal epoch length by exploring different FHR signal segments in the range of 20 and 60 minutes in steps of 5 minutes before delivery. The lower limit of the studied range (20 minutes) was selected according to the literature (Spilka et al., 2016b; Warrick and Hamilton, 2015) as the shortest FHR signal length to be considered for the analysis. The higher limit range was set to 60 minutes because it corresponds to the length of the shortest CTG recording included in the selected dataset.

For the selection of an optimal epoch, we calculate the number of statistically significant features (p-value < .05) for each epoch based on the feature elimination step presented in Section 6.2.3. Then, we select the epoch that allows obtaining the highest number of statistically significant features. In summary, the length of an optimal epoch for this analysis was determined as follows:

1. We compute the proposed features from different FHR segment lengths such as 20, 25, 30, 35, 40, 45, 50, 55, and 60 minutes just before delivery.

2. For each extracted feature, we apply a Wilcoxon rank-sum test (Gibbons and Chakraborti, 2011) under the hypothesis that in the feature dimension, the median value of the data significantly differs from one class to the other.

3. For each set of features (nine sets in total), we calculate the amount of statistically significant features (p-value < .05) that can be extracted.

4. The segment length that provides the highest number of statistically significant features is chosen as an optimal epoch to be used in this analysis.

Results obtained from the previous operation are presented in Fig. 7.1. On the abscissa axis, we have the studied epoch length, and on the ordinate axis, we have the number of significant features for each epoch. As can be observed in the graph, for epochs in the range between 20 and 35 minutes length, the number of significant features is increasing from 21 to 30. Then, for longer epochs (≥ 40 min), the number of significant features decreases. According to this information, a FHR segment of 35 minutes length before delivery can be selected as an informative epoch to perform the proposed analysis.

7.1.2 Performance evaluation of the modal-spectral features

As explained in Section 6.2.2, for the evaluation of our proposed features, we propose to include another set of conventional features, which have shown relatively satisfactory results in CTG classification (Fuentealba et al., 2019b; Krupa et al., 2011). The evaluation

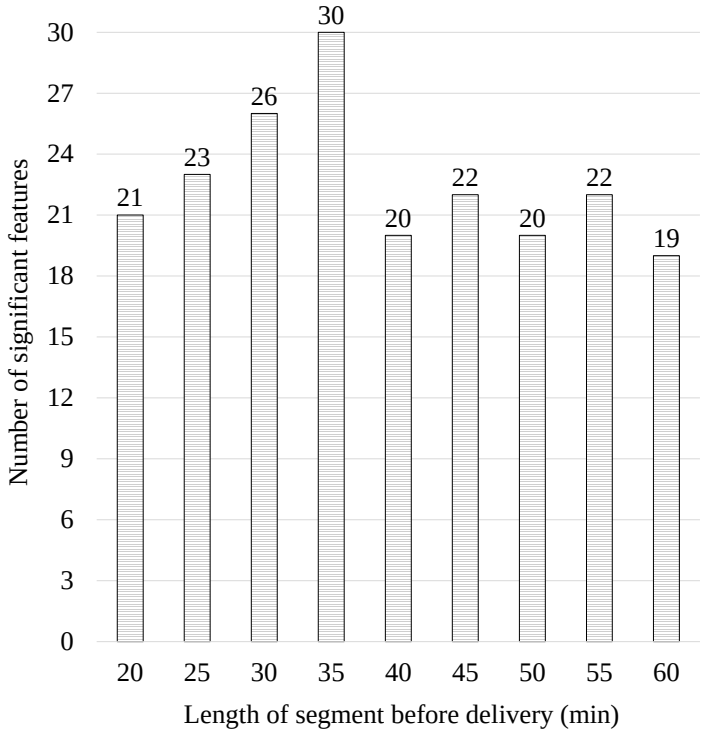


Figure 7.1: Number of significant features (p-value < .05) extracted from different FHR epochs.

performed in this section focuses on studying the contribution of the modal-spectral features in comparison to the conventional features.

From the feature extraction operation explained in Section 6.2.2, the complete set of modal-spectral and conventional features includes 210 and 69 features, respectively. Once the features are extracted, we apply an elimination step performed, as described in Section 6.2.3. This operation allows us to identify the significant features and exclude the non-significant ones for the analysis. As a result, from the complete set of extracted features, we obtain 38 significant features: 30 modal-spectral features and 8 conventional features, which are presented in Table 7.1. Here, we can observe the significant features divided by their category into 14 feature sets (procedure explained in Section 6.2.4). The first ten feature sets correspond to modal-spectral features, while the last four correspond to conventional features.

Once the significant features are obtained, and the feature sets are defined, we evaluate them following the strategy presented in Section 6.2.4. For this operation, we employ three machine learning classifiers (SVM, LDA, and k-NN), whose classification performance is evaluated based on the geometric mean Eq. (6.2) because it is considered an appropriate quality metric (QI) for imbalanced data (Georgoulas et al., 2006; Warmerdam et al., 2016b). Note that this study does not attempt selecting the best classifier, but rather it focuses on evaluating the proposed features extracted based on the combination of ICEEMDAN and TV-AR modeling. Therefore, these three computed-based classifiers, usually employed

Table 7.1: Feature sets and corresponding significant features.

Feature set	Significant features		Type
S_1	IMF_1-E	μ, M, RMS	modal-spectral
S_2	IMF_1-E^c	μ, M, RMS	
S_3	IMF_2-E^c	ApEn, SampEn	
S_4	IMF_4-E^c	ApEn, SampEn	
S_5	IMF_6-E	μ, σ, mad, RMS	
S_6	IMF_6-E^c	μ, σ, mad, RMS	
S_7	IMF_8-E	μ, σ, mad, RMS	
S_8	IMF_8-E^c	μ, σ, mad, RMS	
S_9	IMF_{10}-E	ApEn, SampEn	
S_{10}	IMF_{10}-E^c	ApEn, SampEn	
S_{11}	FHR signal	σ, mad, ApEn, SampEn	conventional
S_{12}	PBL	σ, mad	
S_{13}	IMF_6	ApEn	
S_{14}	IMF_7	ApEn	

classifiers in CTG analysis, are used in this work to provide representative results that are comparable to related works in FHR feature extraction (See Section 7.3).

Based on the previous operation, the results corresponding to the feature sets evaluation are presented in Fig. 7.2, 7.3, and 7.4 for the SVM, LDA, and k-NN classifiers, respectively. Likewise, Tables 7.2, 7.3, and 7.4 show the features automatically included in each combined set to achieve the highest quality (QI) (see Section 6.2.4). In both figures and tables corresponding to this display of results, for better visualization of the contribution of the modal-spectral and conventional features, they are illustrated in gray and white colors, respectively.

Table 7.2: Features automatically selected by using the Support Vector Machine (SVM) classifier, corresponding to the first four combined sets presented in Fig. 7.2.

Combined set k	Features			QI
1	raw FHR		σ	76.4%
	raw FHR		mad	
	raw FHR		ApEn	
	raw FHR		SampEn	
2	IMF_4	E^c	ApEn	78.6%
	IMF_4	E^c	SampEn	
3	PBL		σ	80.7%
	PBL		mad	
4	IMF_{10}	E^c	ApEn	81.7%
	IMF_{10}	E^c	SampEn	

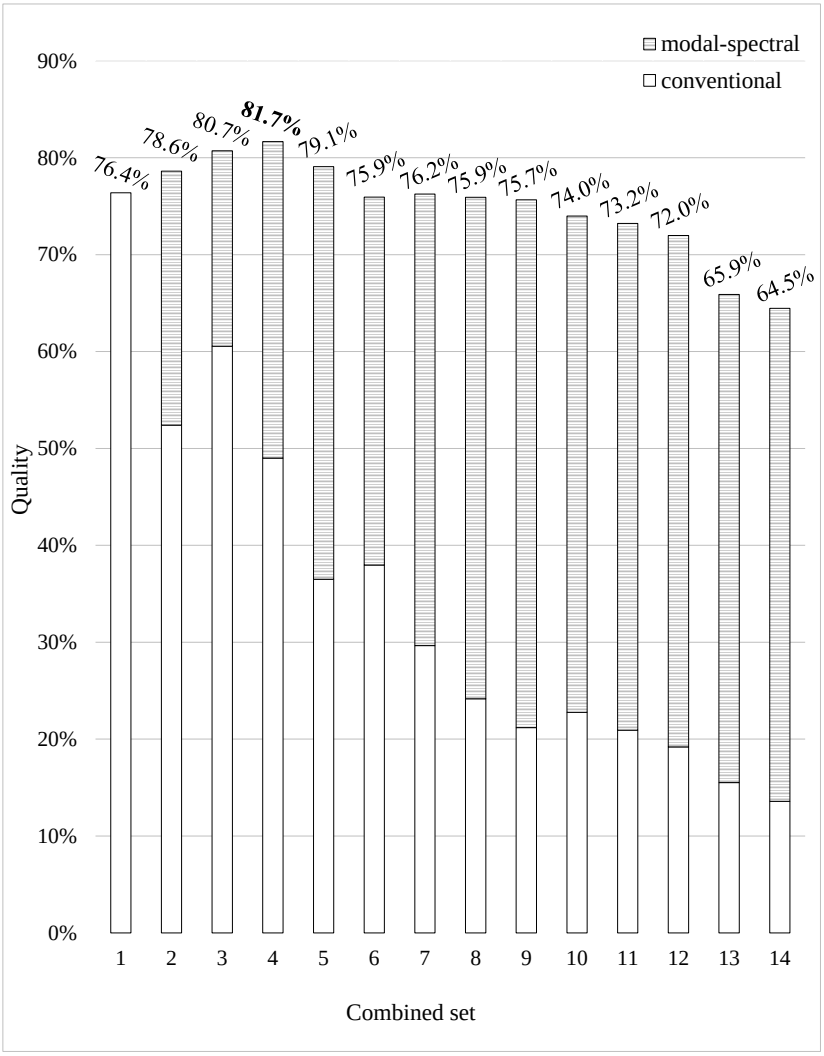


Figure 7.2: Performance of combined sets including both modal-spectral and conventional features based on the strategy presented in Fig. 6.4 and using the Support Vector Machine (SVM) classifier.

Table 7.3: Features automatically selected by using the Linear Discriminant Analysis (LDA) classifier, corresponding to the first four combined sets presented in Fig. 7.3.

Combined set k	Features			QI
1	raw FHR		σ	76.7%
	raw FHR		mad	
	raw FHR		ApEn	
	raw FHR		SampEn	
2	IMF_4	E^c	ApEn	79.1%
	IMF_4	E^c	SampEn	
3	PBL		σ	80.8%
	PBL		mad	
4	IMF_{10}	E^c	ApEn	81.5%
	IMF_{10}	E^c	SampEn	

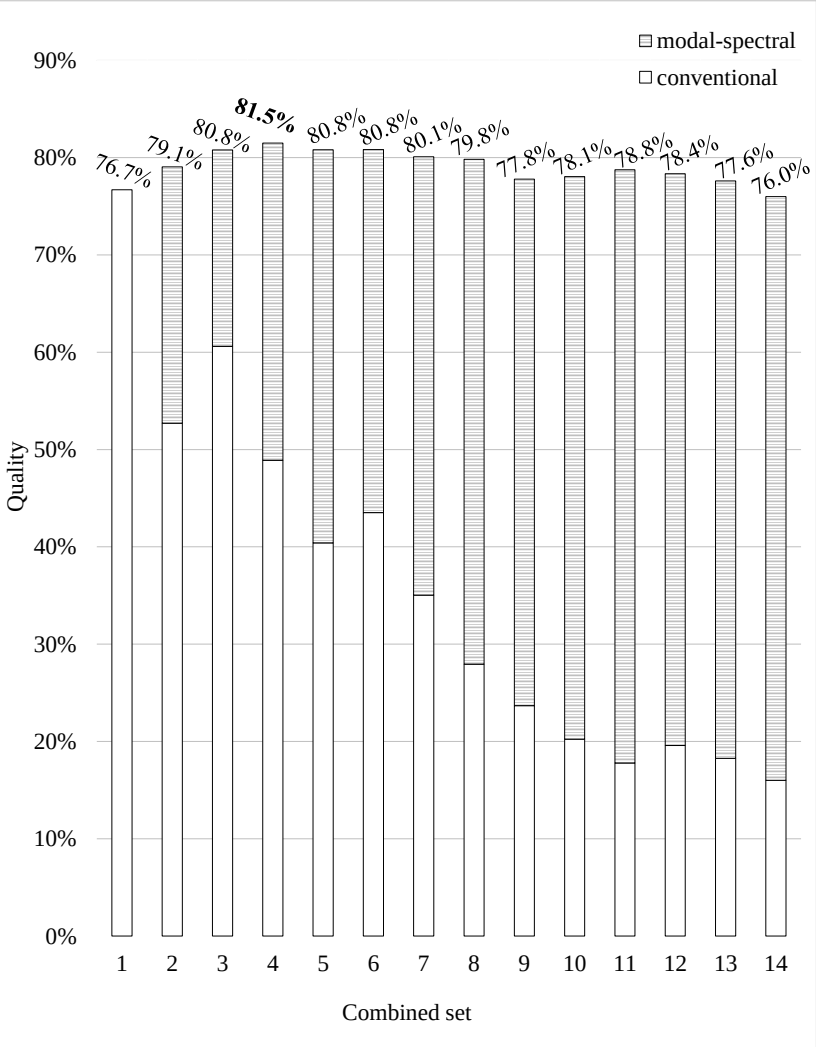


Figure 7.3: Performance of combined sets including both modal-spectral and conventional features based on the strategy presented in Fig. 6.4 by using the Linear Discriminant Analysis (LDA) classifier.

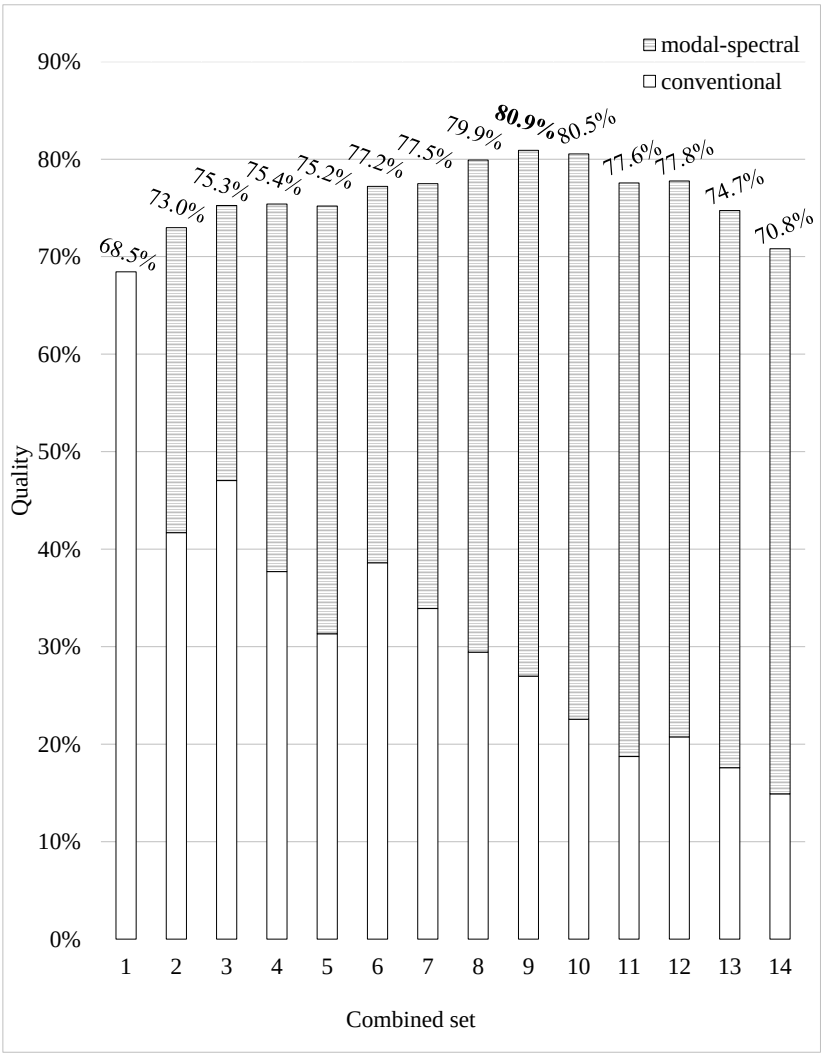


Figure 7.4: Performance of combined sets including both modal-spectral and conventional features based on the strategy presented in Fig. 6.4 by using the *k*-Nearest Neighbors (*k*-NN) classifier.

Table 7.4: Features automatically selected by using the k-Nearest Neighbors (k-NN) classifier, corresponding to the first nine combined sets presented in Fig. 7.4.

Combined set k	Features			QI
1	raw FHR		σ	68.5%
2	raw FHR		mad	
3	raw FHR		ApEn	
4	raw FHR		SampEn	
5	IMF_1	E^c	μ	73.0%
6	IMF_1	E^c	M	
7	IMF_1	E^c	RMS	
8	IMF_6		ApEn	75.3%
9	IMF_{10}	E^c	ApEn	75.4%
	IMF_{10}	E^c	SampEn	
	IMF_4	E^c	ApEn	75.2%
	IMF_4	E^c	SampEn	
	PBL		σ	77.2%
	PBL		mad	
	IMF_2	E^c	ApEn	77.5%
	IMF_2	E^c	SampEn	
	IMF_1	E	μ	79.9%
	IMF_1	E	M	
	IMF_1	E	RMS	
	IMF_{10}	E	ApEn	80.9%
	IMF_{10}	E	SampEn	

Table 7.5: Performance comparison between conventional features and then including the modal-spectral features, as well as for each employed classifier.

Classifier	Conventional	Including modal-spectral
SVM	78.0%	81.7%
LDA	77.9%	81.5%
k-NN	71.4%	80.9%

As shown in Fig. 7.2, 7.3, and 7.4, the highest classification performance achieved by the SVM, LDA, and k-NN classifiers were 81.7%, 81.5% and 80.9% of QI, respectively. Likewise, in order to achieve such performance, the three classifiers require the contribution of both modal-spectral and conventional features. As presented in Tables 7.2, 7.3, and 7.4, for the SVM and LDA classifiers, four feature sets are required to obtain that the highest performance, whereas for the k-NN classifier, nine feature sets are needed.

For studying the significance of the previous results, we calculate the classification performance obtained by the conventional features. The idea is to compare the classification performance obtained by including the modal-spectral with the highest performance achieved by only the conventional features. For this operation, we employ the same strategy described in Fig. 6.4, but now considering the feature sets that include only the conventional features. As performed above, it is also performed for the three classifiers: SVM, LDA, and k-NN. The results of this study are presented in Table 7.5, which shows the classification performance by using only the conventional features and then by including the modal-spectral features for each column, respectively. Here, we can observe that the best performance achieved by using only conventional features is 78.0%, and by including the modal-spectral, it has improved to 81.7%, obtained by using the SVM classifier.

Since the highest performance is achieved by using the SVM classifier, the subsequent analysis concerning the feature contribution performed in this section is focused on the results obtained by this classifier. Particularly, Table 7.2 shows that the highest performance is achieved by the contribution of four feature sets (two modal-spectral and two conventional feature sets). Concerning the feature sets presented in this table, we can observe that by considering one set of conventional features and then including one set of modal-spectral features, the quality improves from 76.4% to 78.6%. Then, the combined set 3 achieved 80.7% of quality by including one more set of conventional features. Finally, the best classification performance ($QI = 81.7\%$) is obtained by the combined set 4, including two sets of modal-spectral and two sets of conventional features.

Note that, as presented in Fig. 7.2, the combined sets 5 to 14 show that by including a larger set of features as input to the classifier, the classification performance may decrease. This phenomenon can be explained by the fact that although those features are considered as significant (p-value $< .05$; criterion explained in Section 6.2.3), they can involve less informative attributes about the studied biological process compared to the features included in the combined set 4.

In summary, as shown in Table 7.2, the selected modal-spectral features are extracted from IMF_4 and IMF_{10}, and the selected conventional features are extracted from the raw FHR signal and PBL. Besides, from the complete set of features studied in this work, the optimal set consists of four modal-spectral and six conventional features. As a result, considering only conventional features (see Table 7.5), the performance achieved 78.0%. Then, by including the modal-spectral features, it improved to 81.7%.

7.2 Contribution of FHR decelerations and resting periods

As explained n Section 1.1, FHR decelerations are one of the most evident patterns represented in the FHR signal resulting from the fetal compensatory mechanism activity. Certainly, these patterns correspond to FHR variations generated by the ANS modulation after a perceived oxygen insufficiency episode such as a UC event (Freeman et al., 2012). Despite they are considered as one of the most significant patterns to interpret the CTG signal, they are one of the most challenging patterns to assess by the clinical staff (Hruban et al., 2015; Spilka et al., 2014b). For this reason, we proposed to study their characteristics in order to investigate whether these patterns, analyzed independently, can involve modulated dynamics that can provide significant information about fetal distress.

The feature evaluation performed in the previous section (Section 7.1) focused on the contribution of the modal-spectral features extracted from the complete signal segment (S-features). In this section, we consider the analysis of S-features as well as features extracted during decelerations (D-features) and during resting periods (R-features) in order to evaluate their contribution in CTG classification. The main idea is to determine whether by considering D-features and R-features, the classification performance can improve compared with the analysis based only on S-features.

As explained in Section 6.2.2, S-features and R-features consist of 210 modal-spectral and 69 conventional features, each one individually. Besides, D-features includes 150 modal-spectral and 63 conventional features.

The studied features are extracted from a FHR epoch corresponding to the last 35 minutes of the recording, which is considered as an informative FHR segment according to the criteria explained in Section 7.1.1. Then, in order to identify the significant features, we apply a feature elimination step (see Section 6.2.3). This operation was already performed above for S-features (see Section 7.1.2), whose resulting significant features (38 features in total) were presented in Table 7.1. In this section, we present that information in Table 7.6. Likewise, we apply the elimination step for D-features and R-features. As a result, we obtained 23 and 27 significant D-features and R-features, which are shown in Table 7.7 and Table 7.8, respectively. These tables show the significant features grouped in feature sets, formed as explained in Section 6.2.4.

For the feature sets evaluation, we make use of the strategy explained in Section 6.2.4, which uses the geometric mean as a metric of classification performance. Considering the

Table 7.6: Feature sets and corresponding significant S-features.

Feature set	Significant features		Type
S_1	IMF_1-E	μ, M, RMS	modal-spectral
S_2	IMF_1-E^c	μ, M, RMS	
S_3	IMF_2-E^c	ApEn, SampEn	
S_4	IMF_4-E^c	ApEn, SampEn	
S_5	IMF_6-E	μ, σ, mad, RMS	
S_6	IMF_6-E^c	μ, σ, mad, RMS	
S_7	IMF_8-E	μ, σ, mad, RMS	
S_8	IMF_8-E^c	μ, σ, mad, RMS	
S_9	IMF_{10}-E	ApEn, SampEn	
S_{10}	IMF_{10}-E^c	ApEn, SampEn	
S_{11}	FHR signal	σ, mad, ApEn, SampEn	conventional
S_{12}	PBL	σ, mad	
S_{13}	IMF_6	ApEn	
S_{14}	IMF_7	ApEn	

Table 7.7: Feature sets and corresponding significant D-features.

Feature set	Significant features		Type
D_1	IMF_1-E	μ, σ, mad, RMS	modal-spectral
D_2	IMF_4-E	σ	
D_3	IMF_4-E^c	σ, RMS	
D_4	IMF_5-E^c	M	
D_5	IMF_6-E	μ, M, σ, mad, RMS	
D_6	IMF_6-E^c	μ, M, σ, mad, RMS	
D_7	IMF_6-f^c	M	
D_8	FHR signal	μ, M, RMS	conventional
D_9	PBL	μ	

Table 7.8: Feature sets and corresponding significant R-features.

Feature set	Significant features		Type
R_1	IMF_1-E	μ, M, RMS	modal-spectral
R_2	IMF_1-E^c	μ, M, RMS	
R_3	IMF_2-E	μ, mad	
R_4	IMF_2-E^c	ApEn	
R_5	IMF_6-E	σ, ApEn, SampEn, RMS	
R_6	IMF_8-E	ApEn, SampEn	
R_7	IMF_{10}-E^c	SampEn	
R_8	IMF_7-f^c	μ, σ, mad	
R_9	FHR signal	σ, mad, ApEn, SampEn	conventional
R_{10}	PBL	σ, mad	
R_{11}	detrended FHR	ApEn, SampEn	

Table 7.9: Classification performance considering S-features, D-features and R-features, based on the SVM classifier.

Features	Optimal feature sets	QI (%)
S	S_{11}, S_4, S_{12}, S_{10}	81.7
D	D_8, D_3, D_1	66.8
R	R_{10}, R_8, R_9	76.3
D_R	R_{10}, R_8, R_9	76.3
S_D	S_{11}, S_4, S_{12}, D_4, D_8	82.5
S_R	S_{11}, S_4, S_{12}, R_5	83.2
S_D_R	S_{11}, S_4, S_{12}, R_5	83.2

Table 7.10: Classification performance considering S-features, D-features and R-features, based on the LDA classifier.

Features	Optimal feature sets	QI (%)
S	S_{11}, S_4, S_{12}, S_{10}	81.5
D	D_7, D_9, D_2, D_1	68.4
R	R_{10}, R_8, R_9	75.1
D_R	R_{10}, R_8, R_9, D_9, D_6	76.4
S_D	S_{11}, S_4, S_{12}, S_{10}, D_6	84.1
S_R	S_{11}, S_4, S_{12}, R_5, S_{13}	84.1
S_D_R	S_{11}, S_4, S_{12}, R_5, D_7, D_4, S_{13}	84.8

results obtained in Section 7.1.2, for this evaluation, we employ the SVM and LDA classifiers, because they have achieved the highest performance by using the conventional and modal-spectral features (see Table 7.5). It is important to note that in Fuentealba et al. (2019e), we have published the results obtained by the SVM classifier. Nevertheless, in order to evaluate the contribution of our proposed features further, in this work, we employ both the SVM and the LDA classifier. The obtained results are presented in Tables 7.9 and 7.10 for the SVM and LDA classifiers, respectively. In these tables, the first, second, and third columns show the studied features, optimal feature sets, and the obtained classification performance, respectively. The optimal feature sets are ordered in the same way that they were selected by the feature selection strategy presented in Fig. 6.4. The first three rows show the classification performance achieved by using S-features, D-features, and R-features, independently. Then, the last four rows present the results considering a combination of the studied features in order to examine whether the contribution of D-features and R-features, compared with the traditional analysis based only on S-features, can improve the CTG classification performance.

On the one hand, concerning the results obtained by the SVM classifier (see Table 7.9), S-features achieve 81.7% of quality, whose value is higher than the performance achieved by D-features, R-features, and the combination of them (D_R). By considering S-features in

combination with D-features (S_D), the classification performance increases from 81.7% to 82.5% of quality. Then, by using S-features in combination with R-features (S_R), the classification performance improves to 83.2%, whose value corresponds to the best classification performance achieved by using the SVM classifier. Particularly, as shown in the last row of Table 7.9 (S_D_R), by including D-features, the classification quality did not improve compared to the results obtained by S_R. Besides, the optimal feature sets were the same for both S_R and S_D_R. Therefore, by using the SVM classifier, D-features do not contribute to improving the CTG classification compared to S_R.

On the other hand, according to the LDA classifier, as presented in Table 7.10, S-features achieve 81.5% of quality, whose value is higher than the performance achieved by D, R, and D_R features. Continuing, for both S_D features and S_R features, the performance increases to 84.2% of quality. Finally, as can be observed in the last row of Table 7.10, by considering S_D_R features, the classification quality improves to 84.8%, which corresponds to the higher classification performance achieved by the LDA classifier. Differently to the previous case based on SVM, by using the LDA classifier, D-features contribute to CTG classification compared with S_R. Therefore, both D-features and R-features allow improving the CTG classification compared with the analysis based on only S-features.

In summary, for the two employed classifiers, the combination of S-features, D-features, and R-features allows a higher classification performance compared with the performance obtained by the contribution of only S-features.

It is important to note that for both classifiers, the classification performance obtained by D-features in combination with R-features (D_R), is still lower compared to the performance achieved by S-features. Therefore, the contribution of S-features is still required to achieve the highest classification performance. In other words, the analysis based on decelerations and resting periods should not replace the analysis based on the complete signal, but rather include them for the assessment.

The obtained results coincide with a recent approach (Warmerdam et al., 2018), where authors employ a different CTG database, but similar class formation criteria and evaluation strategy. That approach shows that a combination of features extracted from the entire FHR segment and contraction-dependent features might improve the detection of fetal distress, whose classification performance improved from 70% to 79% of quality (based on the geometric mean metric). Our approach, based on ICEEMDAN and TV-AR modeling, shows that considering only the complete segment, and then including features extracted from decelerations and resting periods, the performance improved from 81.5% to 84.8% of quality.

The LDA classifier achieved the best classification performance by including the features sets S_{11}, S_4, S_{12}, R_5, D_7, D_4, and S_{13}. As can be observed in Tables 7.6, 7.7, and 7.8, four of these sets correspond to modal-spectral features, whereas the other three sets correspond to conventional features. The selected modal-spectral features are extracted from

the IMF_4-E^c, IMF_5-E^c, IMF_6-E, and IMF_6-f^c. The selected conventional features are extracted from the FHR signal, PBL, and IMF_6 in the time-domain.

In order to analyze the obtained results from a physiological perspective, we focus on examining the potential physiological information involved in the selected features. In this context, features extracted from the FHR signal and PBL describe information in the time-domain that is usually assessed by the clinical staff during labor. Therefore, the potential physiological information associated with those indicators is well defined in CTG guidelines (Ayres-de Campos et al., 2015). Concerning the selected features extracted from the IMFs, we propose to examine the spectral information involved in the IMF_4, IMF_5, and IMF_6 (selected IMFs). The main idea is to identify the frequency information involved in each IMF in relation to the FHR frequency bands established in the literature. For this purpose, making use of the indicator of the frequency of the main component (f^c), we perform a frequency analysis for each IMF as performed in Section 6.2.2. The results of this operation are presented in Fig. 7.5, where each point corresponds to the average of f^c for each IMF and each FHR recording. Note that we employ a semi-log chart, where the ordinate axis is in a logarithmic scale of base ten.

As can be observed in Fig. 7.5, the IMF_4, IMF_5, and IMF_6 can involve spectral information in the range of the LF band ($0.03-0.15$ Hz, demarcated by the dotted lines). From a physiological point of view, these results are highly interesting, because according to the literature (see Section 3.1.1), this LF band is associated with the sympathetic ANS activity of the fetus. Therefore, these results reveal that S-features, D-features, and R-features based on the ICEEMDAN and TV-AR spectral analysis involve significant information that can be directly related to the fetal sympathetic ANS activity. Besides, an independent analysis of S-features, D-features, and R-features can help to assess the fetal condition, and thereby improve the CTG classification performance. Note that the band mentioned above (LF band) is used only as a referential frequency range in order to associate the automatically selected features with a potential fetal physiological phenomenon. It is an important characteristic of our approach because the analysis is not performed only on a limited frequency range established in the literature; rather, it is based on tracking FHR dynamics in the complete band and then examine the significant dynamics that can involve information in the established FHR frequency bands.

7.3 Discussion of results

In the previous sections, we have presented the results obtained by an automatic selection of features extracted from the CTG signal. This feature selection was based on the classification performance estimated by using three machine learning classifiers such as SVM, LDA, and k-NN. As a result, the best performance was 84.8% of quality, including the proposed features computed from the complete signal segment, during decelerations, and during resting periods, using the LDA classifier.

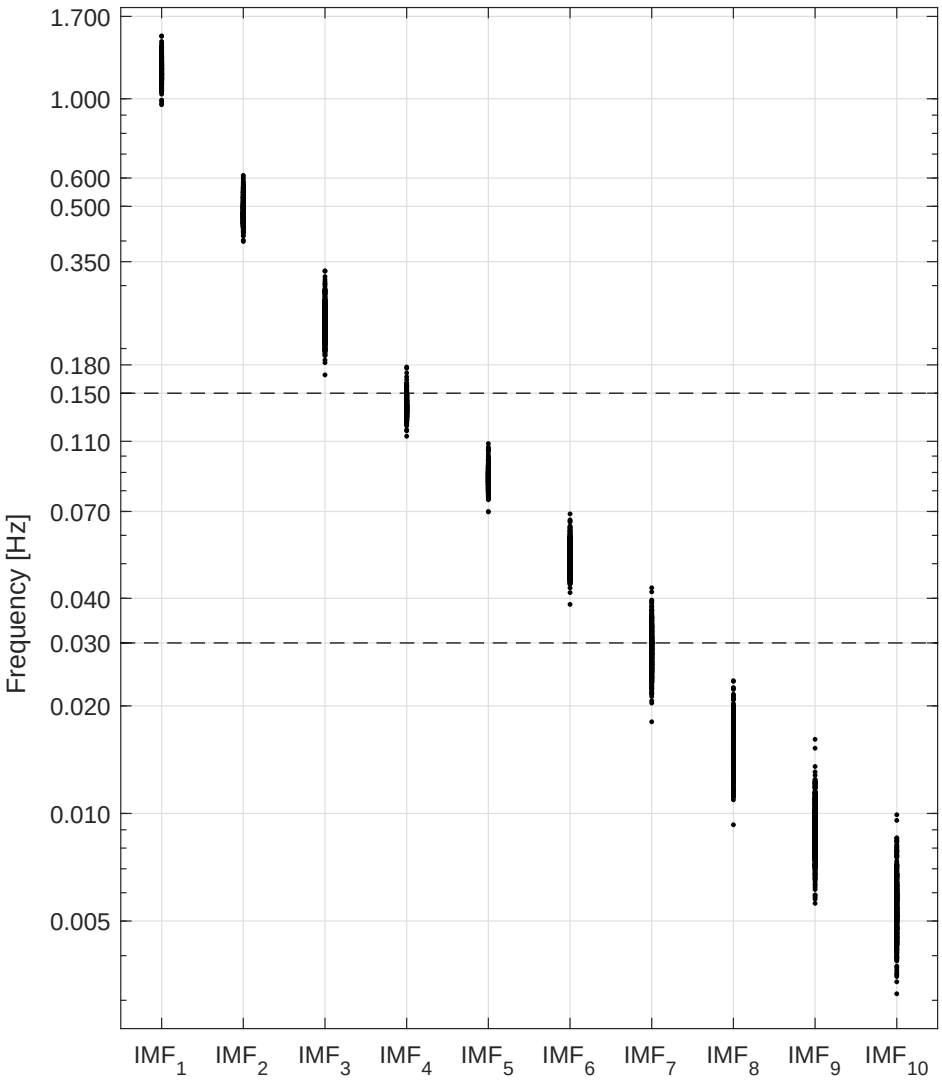


Figure 7.5: Average of frequencies involved in each IMF computed on the entire CTG database.

Table 7.11: Display of recent related works in CTG classification and using pH for class formation.

Reference	Database	# of recordings (abnormal/normal)	CTG selection criteria	*Se* (%)	*Sp* (%)	*QI* (%)
Spilka et al. (2012)	Private	217 (94/123)	pH< 7.15	73.4	76.3	74.8
Georgieva et al. (2013)	Private	252 (126/126)	pH< 7.10 7.27 <pH< 7.33	60.3	67.5	63.8
Xu et al. (2014)	Private	510 (255/255)	pH< 7.05 7.27 <pH< 7.33	83	66	74
Spilka et al. (2016a)	Private	1288 (37/1251)	pH≤ 7.05	73	75	74
Warmerdam et al. (2018)	Private	100 (20/80)	pH< 7.05	81	77	79
Rotariu et al. (2014b)	CTU-UHB	289 (25/264)	pH< 7.20; BDecf> 8	96.0	87.6	91.7
Rotariu et al. (2014a)	CTU-UHB	552 (168/384)	pH< 7.20; apgar< 6	73.2	88.2	80.4
Georgoulas et al. (2017)	CTU-UHB	552 (44/508)	pH≤ 7.05	68.5	77.7	72.9
Yu et al. (2017)	CTU-UHB	88 (44/44) 122 (61/61)	pH≤ 7.05; pH> 7.20 pH≤ 7.10; pH> 7.20	75.3 65.4	84.4 75.4	79.7 70.2
Cömert et al. (2018)	CTU-UHB	552 (113/439)	pH≤ 7.15	63.5	65.9	64.7
Zarmehri et al. (2019)	CTU-UHB	552 (44/508)	pH≤ 7.05	63.6	80.1	71.4
This approach	**CTU-UHB**	**372 (18/354)**	**pH<7.05; BDecf≥12** **pH>7.20; BDecf<12**	**84.9**	**86.0**	**84.8**

In order to determine the significance of the obtained classification performance, we compare it to related works published in the literature, as presented in Table 7.11. This table exhibits a list of recent related approaches that involve comparable class formation criteria based on the pH outcome parameter for CTG classification.

In Table 7.11, the columns from left to right, display the related work (first author and year), employed database, number of recordings (total, abnormal, and normal) considered for the analysis, class formation criteria for the recording selection, sensitivity (*Se*), specificity (*Sp*), and quality (*QI*), respectively. Note that some of these works do not provide the *QI* explicitly; thus, considering that the *QI* metric is employed as a reference for the result comparison, it was computed based on the provided metric of *Se* and *Sp* by using Eq. (6.2). Regarding the results of *Se*, *Sp*, and *QI* obtained in this work (see the last row of the table), they do not satisfy the relation presented in Eq. (6.2). It is because *QI* was not computed directly from the average of *Se* and *Sp*, but rather, as explained in Section 6.2.4, it was calculated for each iteration and then averaged.

As can be observed in Table 7.11, the pH outcome value used for the class formation criteria can significantly differ from one approach to another. This phenomenon can be explained by the fact that whilst the pH outcome is considered as a gold standard measure for outcome assessment, the connection of the neonatal outcome with a precise value of pH at birth is still inconclusive (Perveen et al., 2015). It is important to note that pH is a logarithmic measure; hence slight differences in pH values used for class formation can lead to significant variations in the selected CTG dataset. In consequence, the recordings selected and labeled as abnormal and normal conditions can also differ from one work to another, which can also affect the results of classification performance. Therefore, the comparison of results is complex, or in some cases, even impossible.

Table 7.12: Selected conventional and modal-spectral features, as explained in Section 7.2.

Selected features			Type
S_{11}	FHR signal	σ, mad, ApEn, SampEn	conventional
S_{12}	PBL	σ, mad	
S_{13}	IMF_6	ApEn	
S_4	IMF_4-E^c	ApEn, SampEn	modal-spectral
D_4	IMF_5-E^c	M	
D_7	IMF_6-f^c	M	
R_5	IMF_6-E	σ, ApEn, SampEn, RMS	

Nevertheless, according to the information presented in Table 7.11, the results obtained in our approach overcome most of the results presented in related works, with a classification performance of 84.8%. Particularly, compared with the results achieved by Rotariu et al. (2014b), based on data extracted from the same database (CTU-UHB), our classification performance might seem considerably lower. The higher classification performance achieved by Rotariu et al. (2014b) (91.7%) can be attributed to the criteria used for the dataset selection. Despite they consider a threshold based on pH value (pH< 7.20), after the class formation, they excluded 263 recordings from the CTG database, considering only the recordings with a SNR of over 20%. As a result, they used only 289 recordings for the analysis, whose smaller proportion of the original database and better quality of the signals involved for the analysis make these results not comparable to our approach.

Although our obtained results show an improvement in CTG classification, we consider that the contribution of this thesis work can not be directly evaluated by comparing the percentage of classification performance with related works. In order to evaluate the contribution of our work, we address the central hypothesis (see Section 1.2), which postulates that the fetal response modulated by the ANS can describe different dynamics, whose changes and their progression over time can be related to the fetal condition.

In this context, results obtained from the feature selection procedure show that the highest classification performance is obtained by including the features sets S_4, S_{11}, S_{12}, S_{13}, D_7, D_4, and R_5, and using the LDA classifier (see Section 7.2). As shown in Table 7.12, these selected feature sets include both conventional and modal-spectral features. The selected conventional features are extracted from the FHR signal, PBL, and IMF_6 in the time-domain, whereas, the selected modal-spectral features are extracted from TV-AR spectral information computed from the IMF_4, IMF_5, and IMF_6.

A pairwise correlation matrix of the selected features is presented in Fig. 7.6. Here, we can observe that most of the features have low correlations with each other (close to 0). However, some features show high correlations, which correspond to statistical values of high proximity in nature, and computed in the same feature set. For example, for S_{11} and S_{12}, features computed by σ have high correlation with mad. Likewise, for S_{11}, S_4 and R_5, features computed by SampEn and have high correlation with ApEn. Besides, for R_5,

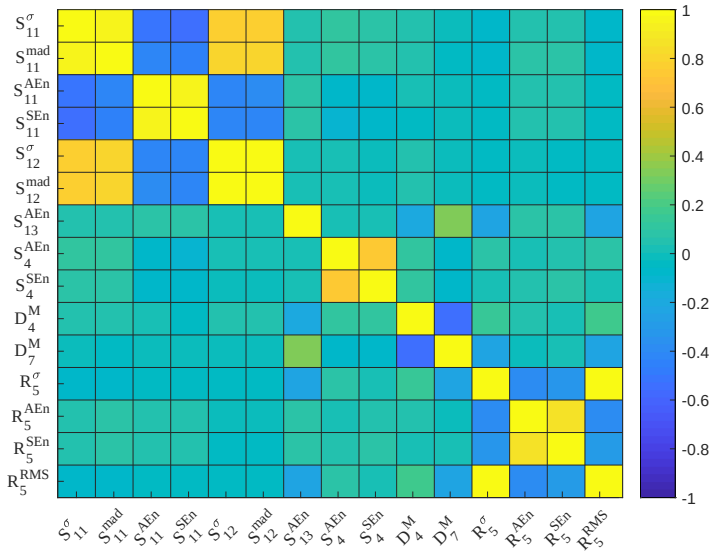


Figure 7.6: Pairwise correlation matrix of the selected features presented in Table 7.12.

σ and RMS show high correlation. It is important to note that, in order to deal with potential redundancy that the studies features may convey, as explained in Section 6.2.4, the feature selection strategy involved a feature preprocessing step based on principal component analysis. Therefore, redundant information involved in the studied features does not affect the feature evaluation and selection. Nevertheless, from features of high proximity in nature and considered as redundant, one can be selected and the others can be removed for further analysis.

In order to determine whether the selected features have a potential physiological connection with the fetal ANS activity, we analyze them based on biomedical research reported in the literature. In this context, the first two traces, corresponding to the FHR signal and PBL, describe information that is usually assessed by the clinical staff during labor. On the one hand, as can be observed in Table 7.12, all the selected features extracted from the FHR signal correspond to statistics of dispersion such as standard deviation, mean absolute deviation, approximate entropy, and sample entropy. These last two coefficients allow quantifying the signal data complexity, which, according to Pincus and Viscarello (1992), might be related to the interaction of the sympathetic and parasympathetic ANS activity of the fetus. On the other hand, as described in CTG guidelines (Ayres-de Campos et al., 2015), a normal FHR baseline involves values between 110 and 160 bpm, whereas lower and higher values lasting more than 10 min are related to fetal bradycardia and tachycardia, respectively. Therefore, the selected features computed from the FHR baseline, described by statistics of dispersion, might represent a certain interaction related to the fetal condition.

Concerning the selected features extracted from the IMF_4, IMF_5, and IMF_6, we proposed to examine their spectral information. The main idea is to identify the frequency information involved in each IMF in relation to the standard FHR frequency bands established in

the literature. As explained in Section 7.2, the IMF_4, IMF_5, and IMF_6 can involve spectral information in the range of the low frequency (LF) band ($0.03-0.15$ Hz). Physiologically, these results are highly interesting, because as explained in Section 3.1.1, this LF band is associated with the ANS activity of the fetus. Therefore, these results represent one of the main contributions of this thesis work. Note that the frequency band mentioned above (LF band) is used only as a reference in order to associate the automatically selected features with a potential fetal physiological phenomenon. Indeed, it is an important characteristic of our approach, because, compared with most of related works, the analysis is not performed only on a limited frequency range. On the contrary, it is based on tracking FHR dynamics in the complete FHR band ($0-2$ Hz), and then the significant dynamics are automatically selected. Besides, considering that the proposed method based on ICEEMDAN and TV-AR modeling can provide significant information extracted from the LF band, it can be an important approach for future research that involves the study of the fetal ANS activity.

Regarding the results obtained in feature selection and classification performance, the proposed features have shown a significant (p-value $< .05$) discriminant capability for CTG analysis, which can contribute to estimate the fetal condition based on automatic classification. From a physiological point of view, significant features and their examination in classification can help to guide the direction of the research in CTG analysis. However, CTG classification is not the object of intrapartum FHR monitoring (Ugwumadu, 2017). Therefore, as mentioned above, a certain percentage of performance in CTG classification, can not be directly applicable nor explain the studied physiological phenomenon in real scenarios. In other words, only metrics related to classification performance are not of clinical interest and not meaningful when assessing the fetal status during labor. In this context, it is essential to explore for a transition towards more significant and representative information, i.e., a connection between the obtained results in classification with more significant information that can be clinically applicable and better understood by the clinical staff.

Considering the central hypothesis of this thesis work (explained above), a FHR segment closer to the delivery can involve information more related to the fetal outcome parameters such as the pH value. On the contrary, FHR segments that are farther to the delivery time could be less connected to the fetal condition at the delivery point.

In order to explore this phenomenon, we propose to examine not only the last 35 minutes but also FHR information farther from the delivery time. For this purpose, we first employ the last 35 minutes to train a model based on the LDA classifier and features selected in Section 7.2. Then, we observe the model prediction over FHR epochs of 35 minutes length by windowing in steps of 2 minutes. For each studied epoch, we compute the selected features, which are the input to the classifier model. Then, as the model output, for each epoch, we obtain a predicted status, which can be normal or acidotic (defined according to the class formation criteria explained in Section 6.2.1).

Although the proposed model has been designed to classify CTG recordings between normal and acidotic classes, we are in an early step to ensure that the proposed model is able to characterize the real fetal status. Indeed, the amount of CTG data studied in this work (from the employed database, only 18 cases labeled as acidotic) is not enough to validate our results. For the moment, we can only associate the predicted status by the model with a particular fetal response that can potentially characterize such predicted fetal status. Nevertheless, the main idea of this discussion is only to observe how the fetal response, represented by the model prediction, changes over time, as well as compare its behavior between the two classes. Therefore, for this proposed analysis, we will consider that a prediction of the trained model classified as acidotic can be associated with a high probability of fetal distress, whereas, a prediction of normal status means a high probability of a normal condition.

For a graphic observation of the proposed analysis, we employ the 18 cases labeled as acidotic (see Section 6.2.1), and we select a similar number of normal cases. For the selection of the normal cases to be used for this analysis, we employ a selection criterion based on values of pH $>$ 7.37 and BDecf $<$ 12. As a result, 20 normal cases were selected. The results of this operation are shown in Fig. 7.7 and 7.8 for the studied normal and acidotic cases, respectively.

In this display of results, the information is presented in such a way that we are able to observe the model prediction for each epoch over time. Considering that the length of the CTG signals available in the database is not uniform, we decided to locate the data on the abscissa axis from 90 to 0 minutes, where 0 corresponds to the delivery time. Each box corresponds to one epoch of 35 minutes, and its color indicates the status predicted by the model. Green and red colors represent normal and acidotic predictions, respectively, whereas an empty box represents epochs involving time instants outside of the range of the available FHR signal data, i.e., outside of the range between the first and the last FHR signal sample available in the recording.

As explained in Section 6.1, the CTU-UHB database also provides the evaluation performed by nine experienced obstetricians, whose annotations are divided into four steps. The first three annotations correspond to the evaluation of the CTG recording divided into three segments (two segments corresponding to the first stage of labor and one segment to the second stage of labor). The fourth annotation step is performed with the objective of predicting the labor outcome (probable range of pH value after delivery), taking into account the information of previous steps. As presented by Hruban et al. (2015), the described evaluation was performed by the nine experts to the complete CTU-UHB database, where they provided a decision about the fetal status classified as normal, suspicious, pathological, or uninterpretable.

In this proposed analysis, we make use of the first three annotations in order to have an overview concerning the evaluation of expert clinicians throughout the CTG recording. From these annotations, we consider the decision when more than 50% of the experts

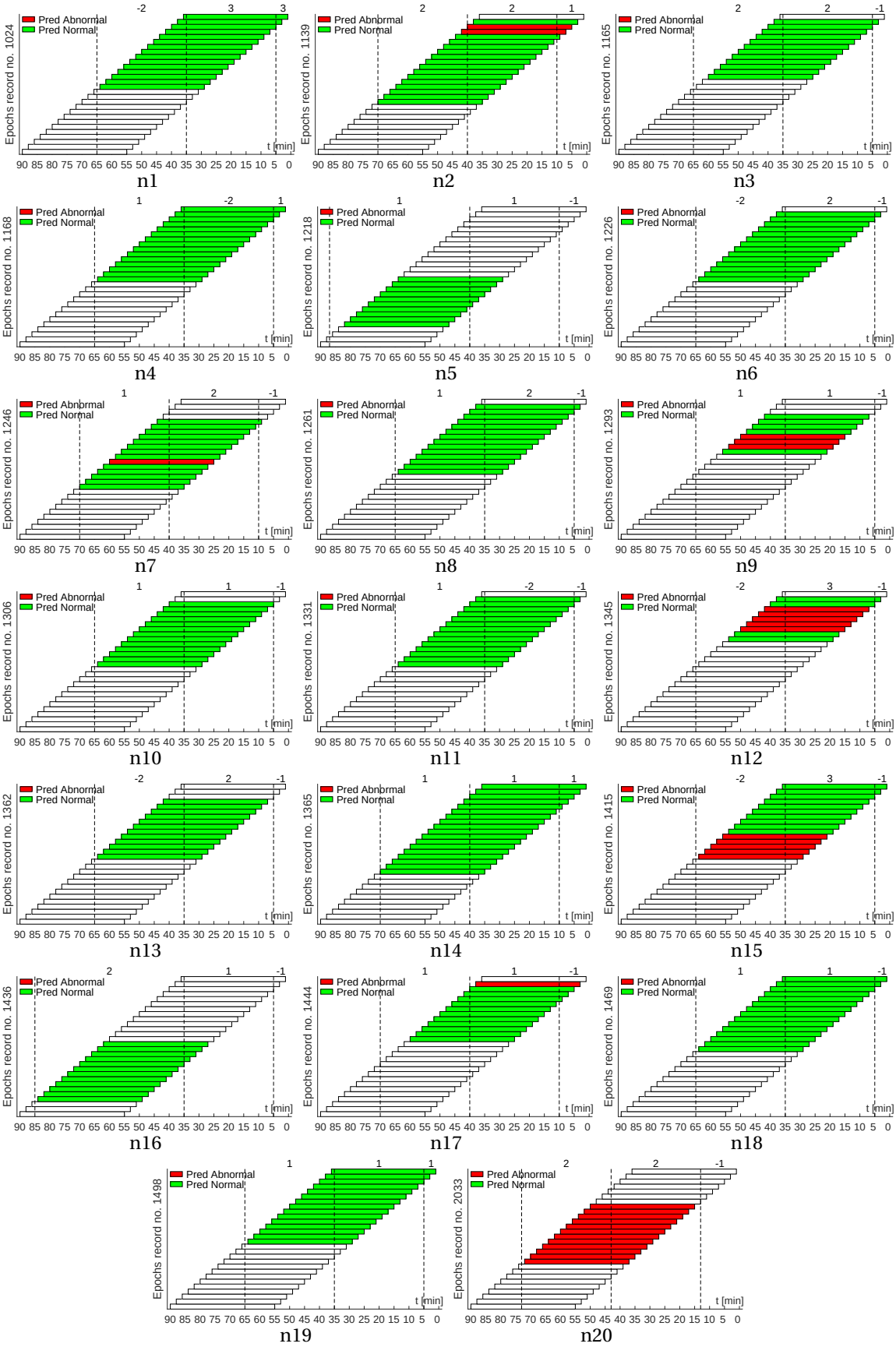


Figure 7.7: Model prediction represented by epochs over time for selected normal cases.

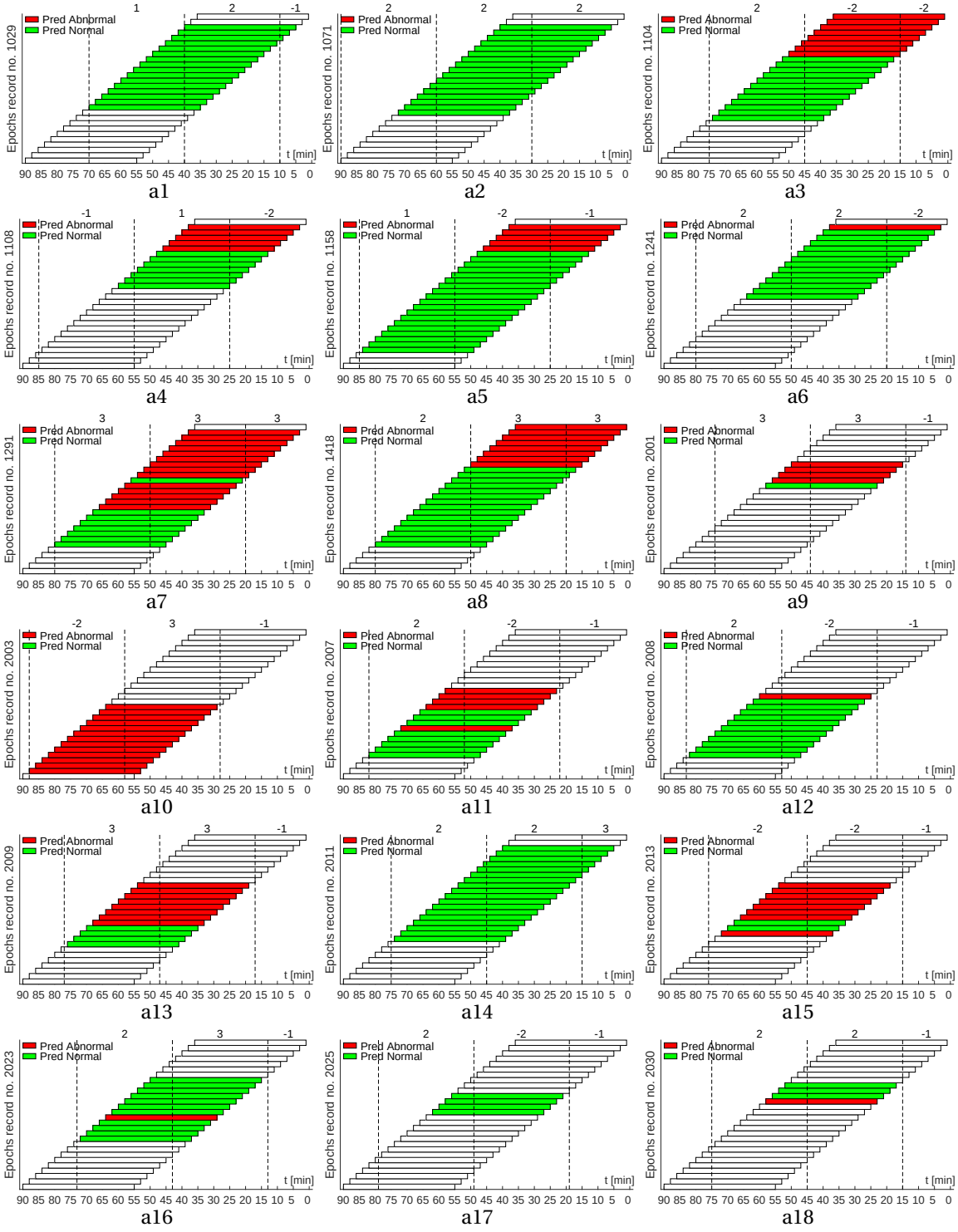


Figure 7.8: Model prediction represented by epochs over time for selected abnormal cases.

are in agreement about the fetal status; otherwise, we label the step as in disagreement. This information is represented in the result display by number labels such as 1 (normal), 2 (suspicious), 3 (pathological), -1 (uninterpretable), and -2 (in disagreement). As can be observed in Fig. 7.7 and 7.8, the evaluated steps are demarcated by three vertical dotted lines, where each line indicates the starting point of each step, and the delivery point is the ending point of the third step. The information is displayed for each step, represented by the number located in the upper part of the figures. It is important to note that this display and analysis does not aim to determine the performance in the expert evaluation nor compare them with our model prediction. The idea is to examine whether, in the studied CTG recordings, the evaluation of the fetal status can change from one step to another.

As can be observed in the figures, both classes show different behavior with each other. For the cases of the normal class, in general, the last epochs indicate a normal status. Likewise, for most of the cases of the acidotic class, the last epochs indicate a high probability of fetal distress. Unfortunately, several of the studies CTG recordings available in the database (such as no. 1218, 1436, 2033, 2001, 2003, 2007, 2008, 2009, 2013, 2023, 2025, and 2030) have limited information of the FHR by the end of labor, whose signal loss is represented by empty boxes by the end of the process. Nevertheless, as shown in most of the cases, we can still observe that the characteristics involved in the fetal response, represented by the model prediction, might change over time. This phenomenon also has been reported in previous research (Coletta et al., 2012; Devoe, 2016; Parer and Ikeda, 2007), which indicates that the risk of the fetal condition, to evolve from one health state to another, frequently change in the course of labor. .

In Fig. 7.7, corresponding to normal condition, the cases n1, n3, n4, n6, n8, n10, n11, n13, n14, n18, and n19, show that, according to the model prediction, the fetal response is associated with a normal status, whose response remains constant during the complete process. Cases n2, n7, n9, n12, and n15 show that several minutes before delivery, there are episodes (represented by red blocks) when the fetal response could be associated with distress but after a specific time by the end of the process, it presents a response that indicates a normal condition. For the case n20, it indicates distress in the first and second steps of labor; nevertheless, we do not have available FHR information in the last 14 minutes to continue tracking the process. Considering that for the case n17, the model predicts distress by the end of the process, it can be considered as an outlier for the model.

In Fig. 7.8, corresponding to the acidotic group, we can observe that the cases a3, a4, a5, a6, a7, and a8 show a prediction of fetal distress by the end of the process. Although some of these cases do not have FHR information in the last epoch (signal loss in the last 2 minutes), the distress status is predicted by the model 16, 12, 12, 4, 20, and 16 minutes before delivery, respectively. Unfortunately, for cases a9, a10, a11, a12, a13, a15, a16, a17 and, a18, we can not track the model prediction during the complete process because of the considerable signal loss in the last step of labor. Nevertheless, in most of the cases, it is possible to observe that the model prediction can change from normal to distress in the course of labor. The cases a1, a2, and a14 show a prediction of normal status for all the studied

epochs, considering that the signal loss before delivery involves only 4 minutes, it might be considered as an outlier for the model.

The analysis presented above, based on normal and acidotic cases, coincide with some approaches proposed in the medical research area (Garabedian et al., 2017; Ugwumadu, 2014), which emphasize that each fetus has a particular behavior and control, and during labor, its condition depends on how it is compensating and defending itself over time.

Concerning the evaluation performed by the expert obstetricians, if we observe the annotations of normal (1), suspicious (2) and pathological (3) conditions (excluding the annotations -1 and -2), in cases n1, n3, n4, n5, n9, n10, n14, n17, n18, n19, n20, a2, a6, a7, a9, a13, and a18, the evaluation remains stable from one step to the next one. Nevertheless, in cases n2, n7, n8, n16, a1, a8, a14, and a16, we can observe that the evaluation can change from one step to another, either from normal (1) to suspicious (2), or from suspicious (2) to pathological (3). Unfortunately, for the remaining cases (n6, n11, n12, n13, n15, a3, a4, a5, a10, a11, a12, a15, and a17), in more than one step the annotations are uninterpretable or in disagreement. Therefore, in these cases, it is not possible to determine if there are changes in the evaluation from one step to the next one.

In summary, considering the previous analysis, we can observe that the prediction of fetal status might change in the course of labor, which is shown not only by our model prediction but also evidenced by recent approaches in medical research, and the expert evaluation explained above.

It is important to note that this performed analysis showed only preliminary results that can be further investigated. Nevertheless, as a future perspective, we consider that the proposed way of visualization presented in Fig. 7.7 and 7.8, can be significant to assist clinicians in real scenarios.

On the one hand, considering that the fetal response can change over time, we can not estimate the condition based only on a snapshot of the process. On the other hand, given that the CTG analysis requires a certain FHR epoch length to be assessed, the prediction can not be displayed by each signal sample, because the analysis does not provide such information. For example, we can not identify whether the predicted status corresponds at the beginning, at the middle, or at the end of the studied epoch. In this context, our proposed display of results deals with these characteristics, offering a visualization that can help to track the fetal status over time based on binary classification. Besides, it can provide information concerning the evaluation of past epochs as a memory of the process, and thereby clinicians can have access to revise past episodes to perform a long-term assessment.

Additional future work and perspectives concerning this analysis are addressed in Section 8.2.

8

Conclusions and Future Work

8.1 Conclusions

This section aims to address the research questions proposed in Section 1.2, which are listed in the following: (1) Is it possible to identify and track modulated time-variant dynamics involved in the FHR signal that are related to the fetal health condition? (2) Could the dynamics involved in FHR decelerations, analyzed independently from the complete signal, improve the automatic CTG signal analysis? (3) Are these studied FHR signal dynamics linked to the fetal compensatory mechanisms modulated by the ANS? (4) Could this approach based on a time-variant analysis of the fetal response improve the automatic computer-based CTG classification compared with the conventional CTG analysis?

Regarding the first research question, as explained in Chapter 3, the FHR signal involves highly complex time-variant characteristics, which are mainly associated with nonlinearities resulting from the physiological regulation mechanisms modulated by the ANS contributing to the FHR activity. In this context, we have postulated that if such characteristics are not previously decomposed or demodulated, the FHR signal can still involve modulated dynamics of high complexity that are difficult to analyze. As explained in Section 5.5, in order to investigate these FHR dynamics, we have proposed a new approach by combining two signal processing methods: ICEEMDAN and TV-AR spectral estimation. On the one hand, ICEEMDAN allows decomposing signals involving nonlinear and non-stationary characteristics into IMFs. Considering that this decomposition operation has similar principles of signal demodulation, this method can be appropriate for studying the FHR dynamics modulated by the ANS activity. On the other hand, in order to study their time-variant characteristics in the spectral-domain, we employed TV-AR modeling.

In the spectral-domain, the information of interest is represented by the spectral density peaks, whose dynamical behavior describes the contribution of the frequency dynamics over time. As explained in detail in Section 4.1, TV-AR modeling can effectively describe the peaks of a narrow-band power spectrum, is more robust to noise, and requires less sampling frequency than standard methods to obtain the same spectral resolution. Therefore, it results appropriate for the spectral analysis of the extracted IMFs. Consequently, ICEEMDAN allows obtaining a less complex signal by decomposing its dynamics into the IMFs, which are better suited for the spectral analysis performed by TV-AR modeling. The

time-variant frequency characteristics are easier to track because the dynamics are already decomposed, and the spectral analysis focused on tracking only one main frequency component of interest for each IMF over time.

Based on the proposed methodology explained above, we track and investigate whether certain FHR signal dynamics are potentially related to fetal distress. For this purpose, we have analyzed CTG recordings corresponding to normal and acidotic fetuses extracted from the CTU-UHB database. As presented in Chapter 5, the performed time-variant spectral analysis revealed that it is possible to decompose and track different spectral dynamics involved in the FHR signal, which are strongly variant in time. In order to examine the connection of these dynamics with the fetal health condition, we compute a set of features based on their spectral characteristics. Then, we studied their discriminant capability by using a Wilcoxon rank-sum test, applied under the hypothesis that the median value of the normal class data significantly (p-value $< .05$) differs from the median value of the acidotic class data. As a result, the behavior of the spectral dynamics significantly differs between normal and acidotic cases. It is important to note that the described information also coincides with results presented in Section 7.1, where these FHR signal characteristics are evaluated further based on their performance in CTG classification. In conclusion, according to the obtained results, it is possible to identify and track modulated time-variant FHR signal dynamics, whose behavior over time can depend significantly (p-value $< .05$) on the fetal health condition.

Regarding the second research question, as described in Chapter 1, FHR decelerations are one of the most evident patterns represented in the FHR signal that result from the fetal compensatory mechanism activity. Certainly, these patterns correspond to FHR variations generated by the ANS modulation after a perceived oxygen insufficiency episode. Although they are considered as one of the patterns that involve more significant about the fetal health condition, they are one of the most challenging patterns to interpret by the clinical staff. For this reason, we proposed to study their characteristics in both the time-domain and the spectral-domain in order to investigate whether these patterns can involve modulated dynamics that can provide significant information about the fetal condition. The main idea is to study the FHR decelerations, and resting periods, together and independently from the complete FHR signal in order to investigate their contribution in CTG assessment.

As presented in Section 5.6, we found that the FHR deceleration episodes and resting periods can involve significant (p-value $< .05$) information about the fetal health condition. In order to examine their contribution to CTG assessment, in Section 7.2, we have evaluated their performance in CTG classification. For this purpose, we first computed the proposed signal features from the complete segment (S-features), during decelerations (D-features), and during resting periods (R-features), independently. Then, we evaluated the extracted features in order to examine whether the contribution of D-features and R-features, can improve the CTG classification performance compared with the analysis based only on S-features. It is important to note that the features extracted from the complete FHR signal

segment are still required to achieve the highest classification performance. In other words, the analysis performed during decelerations and resting periods should not replace the analysis based only on the complete signal, but rather include them separately for the analysis. As a result, the CTG classification performance by including S-features, D-features, and R-features has improved compared with the performance achieved by only S-features. Therefore, according to the obtained results, both the FHR decelerations and resting periods can contribute to improving the CTG signal analysis compared with the analysis based only on the complete FHR signal segment.

These obtained results coincide with a recent approach (Warmerdam et al., 2018) based on the analysis of contraction-dependent features. In this approach, the authors employed a different CTG database, but a similar strategy for evaluation and class formation criteria. As a result, they revealed that features extracted from the entire FHR segment in combination with contraction-dependent features might improve the detection of fetal distress, whose classification performance, improved from 70% to 79% of quality. Our approach, based on ICEEMDAN and TV-AR modeling, has shown that considering the complete FHR segment. Then including features extracted from decelerations and resting periods, the performance improved from 81.5% to 84.8% of quality.

Regarding the third research question, in Section 7.2, we have presented the evaluation of the proposed features, which involved two main steps, feature elimination and feature selection based on their classification performance. As a result of this operation, we provided a detailed explanation concerning the feature selection, where the LDA classifier provided the highest classification performance. This classification performance was achieved by the contribution of features sets from S-features, D-features, and R-features. These selected features sets include both conventional and modal-spectral features, extracted from the FHR signal, PBL, IMF_4, IMF_5, and IMF_6.

As explained in Section 7.3, in order to determine a potential physiological connection between the selected features and the fetal ANS activity, we analyzed them based on results reported in the literature in the field of biomedical research. On the one hand, the first two traces (FHR and PBL) describe information that is usually assessed by the clinical staff during labor, which according to the CTG guidelines, involves patterns that can be related to the fetal health condition. Besides, according to the literature (Pincus and Viscarello, 1992), the FHR signal complexity might be related to the interaction of the sympathetic and parasympathetic ANS activity of the fetus. On the other hand, the last three traces (IMF_4, IMF_5, and IMF_6) can involve spectral information in the range of the FHR LF band ($0.03 - 0.15$ Hz). From a physiological perspective, these results are highly interesting, because as explained in Section 3.1.1, this LF band is mainly associated with the ANS activity of the fetus. Therefore, these results represent one of the main contributions of this thesis work. It is important to note that the LF frequency band mentioned above was used only as a reference to associate the automatically selected features with a potential fetal physiological phenomenon. In other words, the analysis was not performed only on a limited frequency range; on the contrary, the proposed approach was based on tracking FHR

dynamics in the entire FHR band ($0-2$ Hz), and then the significant dynamics were automatically selected. For a detailed explanation about the features selection strategy and a discussion on their significance, please refer to Section 6.2 and 7.3, respectively.

In summary, those results indicate that the features extracted and automatically selected can involve significant information related to the fetal health condition. The studied FHR dynamics, described by these selected features, can be linked to the fetal compensatory mechanisms modulated by the ANS. However, in order to validate these results, more investigation is required.

Regarding the fourth research question, in Section 7.1, we have studied the contribution of the proposed modal-spectral features in comparison with conventional features. For this purpose, we have selected an optimal FHR epoch corresponding to the last 35 minutes before delivery, where the features are computed. The complete set of features consisted of 210 modal-spectral and 69 conventional features, which after a feature elimination step, resulted in 30 and 8 features, respectively. Their classification performance was evaluated based on three machine learning classifiers (SVM, LDA, and k-NN). First, only the conventional features were considered, and then the modal-spectral were included in order to evaluate their contribution. On the one hand, the classification performance achieved by the conventional features was 78.0%, 77.9%, and 71.4% of quality, for SVM, LDA, and k-NN, respectively. On the other hand, the classification performance achieved by the modal-spectral features was 81.7%, 81.5%, and 80.9% of quality, for SVM, LDA, and k-NN, respectively. As a result, based on the contribution of the proposed modal-spectral features, the proposed approach involving a time-variant analysis of the fetal response has improved the CTG classification performance compared with the conventional features. Besides, as presented in Section 7.2, this classification performance has even increased to 84.8% of quality by including the independent analysis of deceleration and resting periods based on the LDA classifier.

In order to analyze the significance of our proposed approach and obtained classification results, we have compared them with related works reported in the literature. As presented in Section 7.3, Table 7.11, recent related approaches, which involve the pH value as class formation criteria, have been considered for their comparison with our approach. Considering the diversity of class formation criteria used in the literature, which can significantly differ from one approach to another, a direct comparison of results is difficult. Nevertheless, the results obtained in our approach overcome most of the results presented in related works, with a classification performance of 84.8%. Particularly, only one compared approach (Rotariu et al., 2014b) might show better results than our approach. The higher classification performance reported by Rotariu et al. (2014b) (91.7%) can be attributed to the criteria used for the dataset selection. Despite they consider a threshold based on pH value (pH$<$ 7.20), after the class formation, they cleared 263 recordings from the CTG database, considering only recordings with a higher signal-to-noise ratio. Therefore, they consider a smaller proportion of the original database (289 recordings) with better signal quality for the analysis. In consequence, these results are not comparable to our approach.

In summary, our proposed methodology based on ICEEMDAN and TV-AR modeling has shown to be a promising approach for fetal distress estimation during labor. It has overcome most of the results presented in related works. However, considering the diversity of class formation criteria and the number of recordings used in recent related works, further investigation is needed. In order to validate and prove our classification results, together with the proposed analysis and evaluation, it would be necessary the study different class formation criteria, other techniques for automatic classification, and a higher number of CTG recordings extracted from different databases.

8.2 Future Work

As explained in Section 7.3, from a physiological point of view, significant features, and a high performance in classification can help to guide the direction of the research in CTG analysis. However, CTG classification, applied only over a particular FHR segment length, is not the object of intrapartum FHR monitoring (Ugwumadu, 2017). Therefore, the numerical results explained above, based on a certain percentage of performance in CTG classification, can not be directly applicable in real scenarios nor describe completely the studied fetal physiological phenomenon. In other words, only classification performance metrics are not of clinical interest and not meaningful when assessing the fetal status during labor.

As future work, it would be essential to explore for a transition towards more significant and representative information, i.e., a connection between the obtained results in feature classification and more significant information that can be clinically applicable and better understood by the clinical staff.

In this context, as explained in Section 1.2, we defined the important concept of *progressive fetal response*. For example, under this concept, FHR segments closer to the delivery can involve more representative information connected to the fetal outcome than segments that are farther to the delivery time.

In order to explore this phenomenon, we propose to examine not only the last 35 minutes but also FHR information farther from the delivery time. In Section 7.3, we have presented preliminary results and discussion concerning this analysis, based on 18 and 20 acidotic and normal cases, respectively. For this purpose, we made use of the prediction of a model trained in the last 35 minutes, and then we observe the model prediction in FHR segments farther from the delivery time. Results showed that normal and acidotic cases could exhibit different behavior to each other.

On the one hand, most of the cases corresponding to a normal condition showed that the model prediction does not change over time. However, some of them showed that several minutes before delivery, the prediction could be associated with distress, but after a certain time instant, by the end of the process, it can change to a normal condition. On the other hand, most of the acidotic cases showed that the model prediction could change over time. First, several minutes before delivery, these cases can involve episodes where

the model can predict a normal condition. Then, after a particular time instant, the prediction changes from normal to distress. This phenomenon is also shown by the evaluation performed by expert clinicians, which revealed that the fetal condition could change in the course of labor.

Considering these preliminary results, we firmly recommend further investigation of this phenomenon. For a future step, we propose the following main research questions:

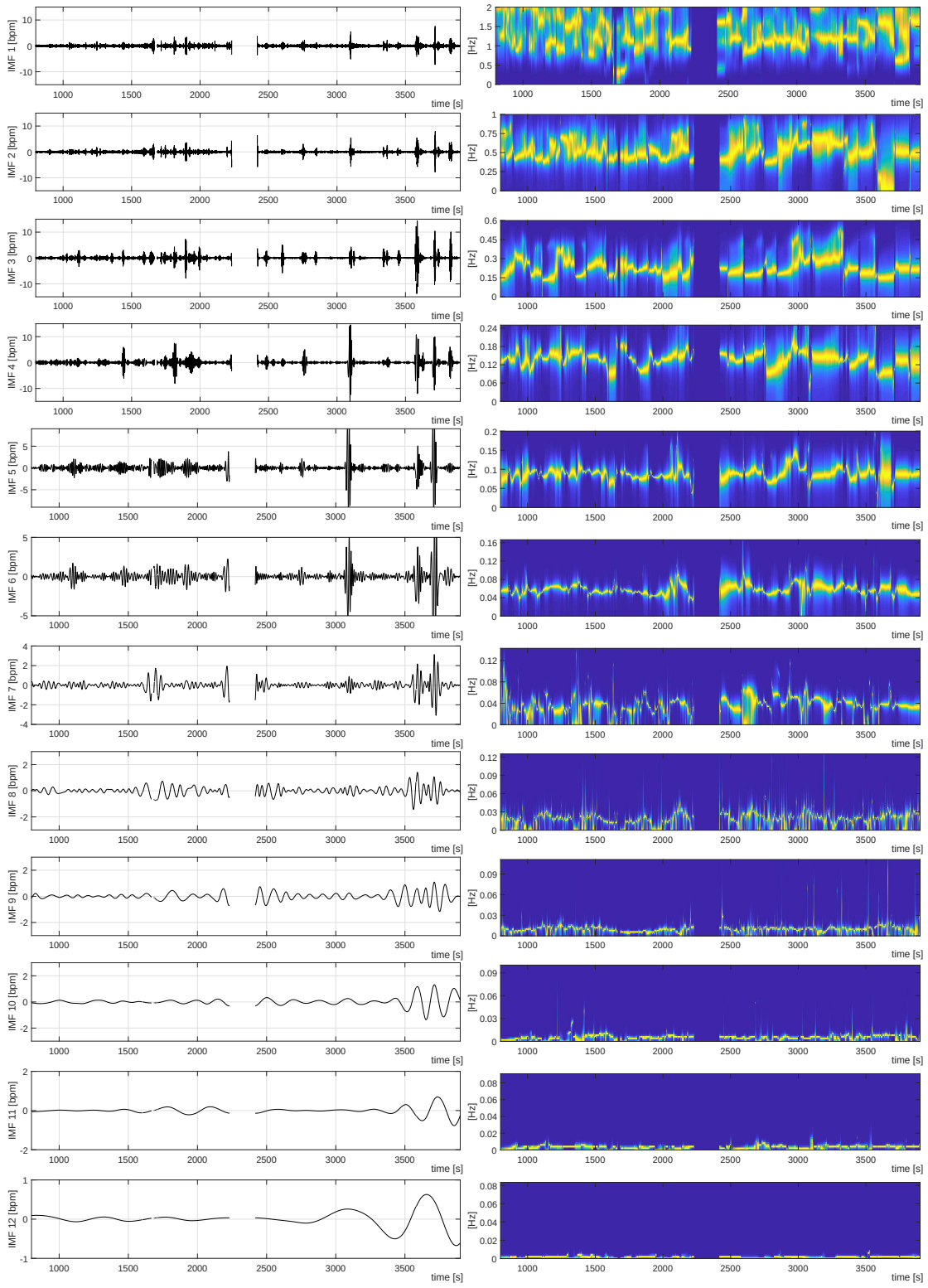
- **Is it possible to characterize and track episodes of fetal distress over time?**
- **How can these episodes of distress affect the fetal outcome or lead to an acidotic condition?**
- **Is there a relationship between the duration and frequency of these episodes with the fetal health condition?**
- **Under which conditions a fetus can tolerate a distressed status, defense itself, and recover to a normal condition?**

All of the previous inquiries open perspectives for an automatic analysis, which considers the evolution of fetal distress and during labor. Therefore, it can help to integrate the evolution concept in the analysis in order to improve the reliability of the CTG assessment. Besides, considering that the proposed method based on ICEEMDAN and TV-AR modeling could provide significant information that can be related to the fetal ANS activity, it can be aplied in future research that involves the study of the interaction of the fetal compensatory mechanisms during labor.

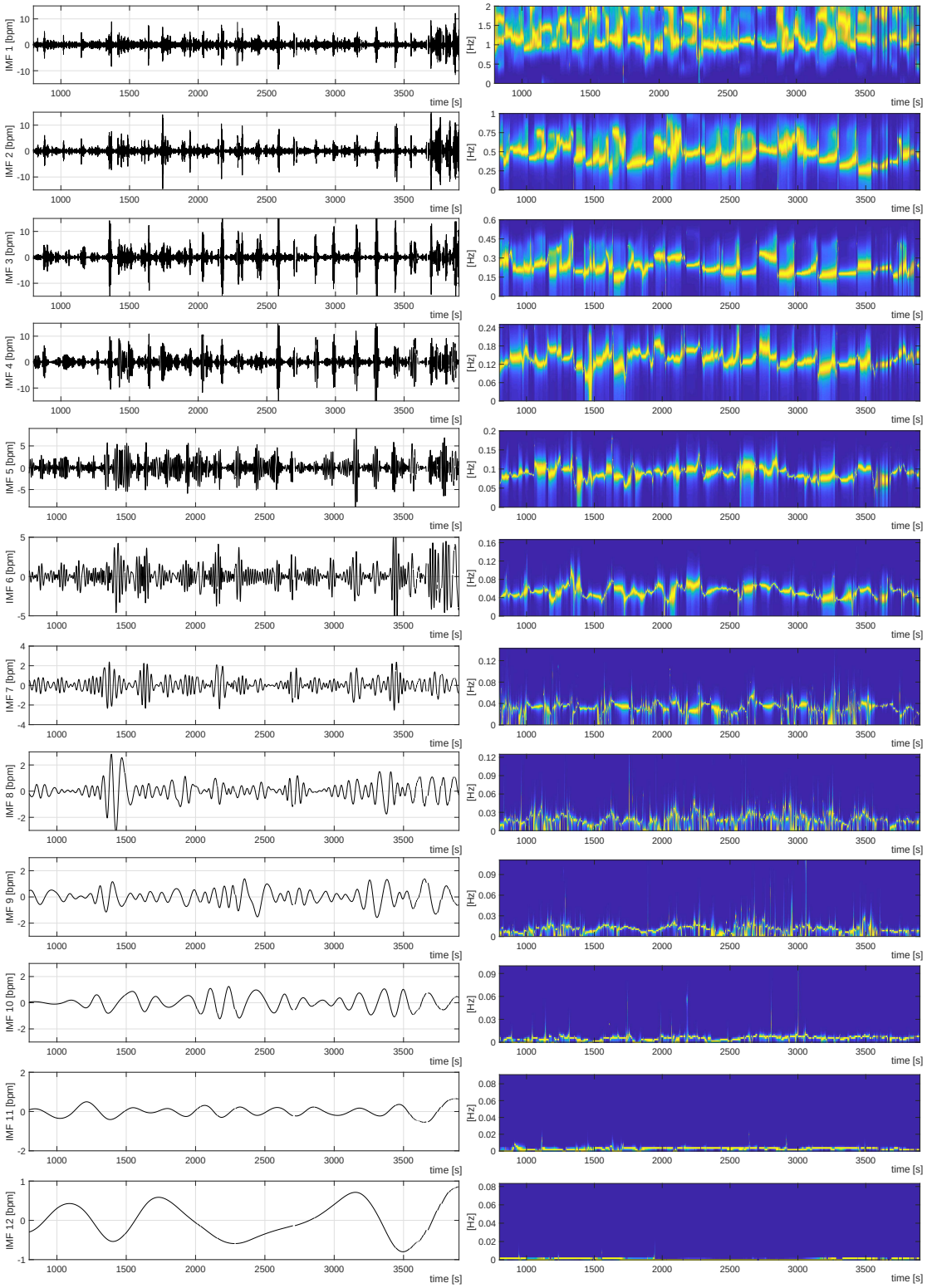
A

Appendix

A.1 Examples of ICEEMDAN and TV-AR spectrum computation

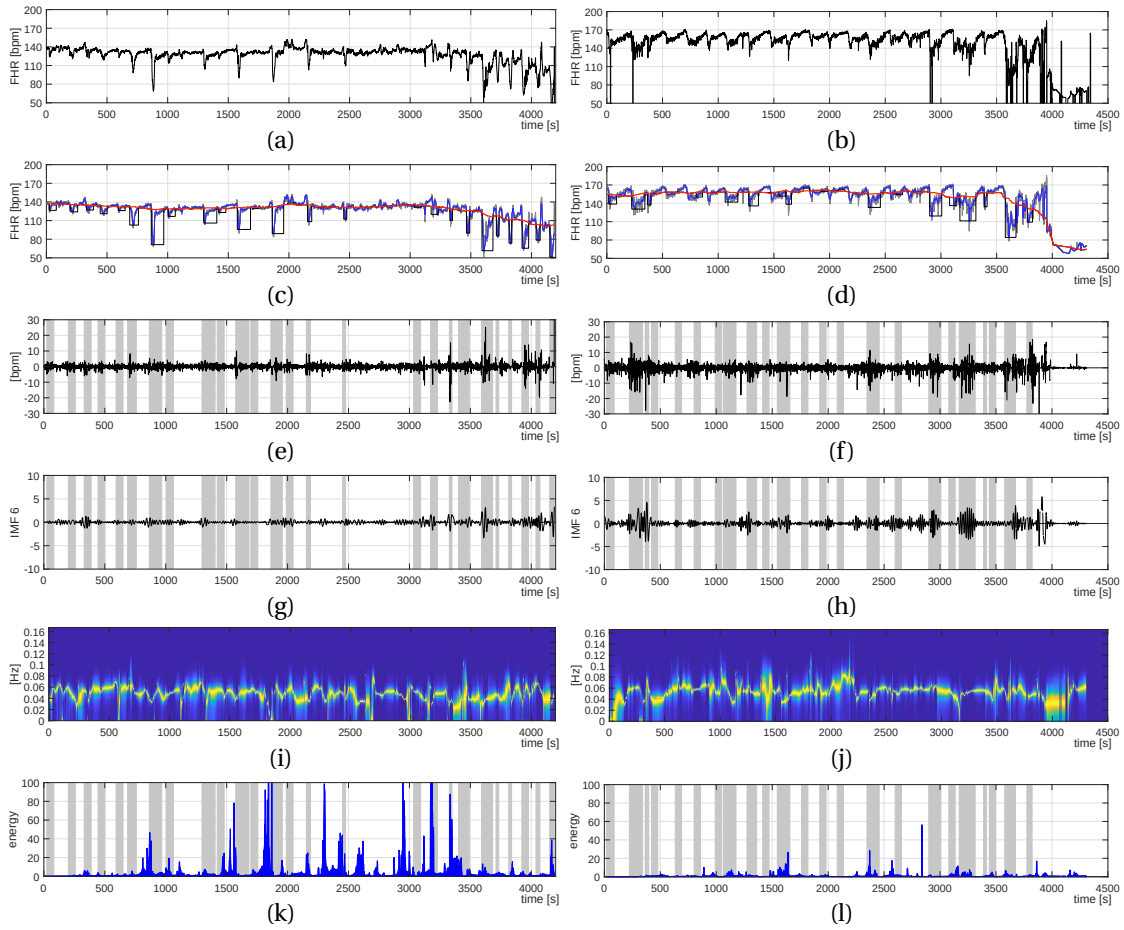


ICEEMDAN decomposition and TV-AR spectrum; (left column) IMFs computed from the detrended FHR signal of recording no. 1024; (right column) TV-AR spectrum of each IMF.

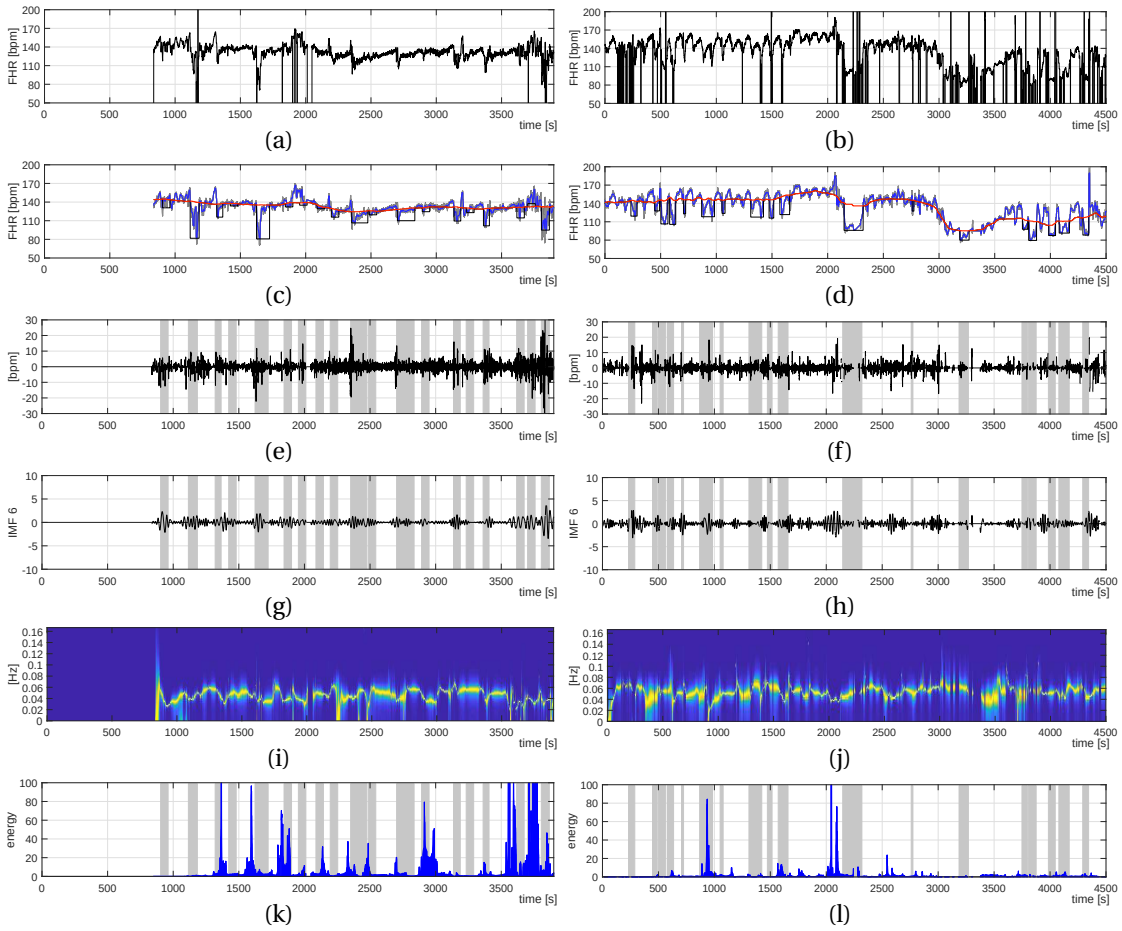


ICEEMDAN decomposition and TV-AR spectrum; (left column) IMFs computed from the detrended FHR signal of recording no. 1176; (right column) TV-AR spectrum of each IMF.

A.2 Representative cases of proposed features



Representative cases; (a) raw FHR signal of recording no. 1167, pH = 7.36, BDecf = 5.03; (b) raw FHR signal of recording no. 2011, pH = 7.01, BDecf = 12.1; (c-d) preprocessed FHR signal (gray), PBL (red), floating-line (blue), and deceleration episodes (black); (e-f) detrended FHR signal; (g-h) IMF_6; (i-j) TV-AR spectrum computed from IMF_6; (k-l) E from (i-j).



Representative cases; (a) raw FHR signal of recording no. 1189, pH = 7.36, BDecf = 0.43; (b) raw FHR signal of recording no. 1104, pH = 6.92, BDecf = 23.75; (c-d) preprocessed FHR signal (gray), PBL (red), floating-line (blue), and deceleration episodes (black); (e-f) detrended FHR signal; (g-h) IMF_6; (i-j) TV-AR spectrum computed from IMF_6; (k-l) E from (i-j).

B

Abbreviations and Notations

Dataset and clustering acronyms

Acronym	Meaning
ADASYN	Adaptive synthetic sampling
ANS	Autonomic nervous system
ApEn	Approximate entropy
AR	Autoregressive
BE	Base excess
BDecf	Base deficit of extracellular fluid
CEEMDAN	Complete ensemble empirical mode decomposition with adaptive noise
CO_2	Carbon dioxide
CTG	Cardiotocograph
CS	Cesarean section
DC	Zero-frequency component
DFT	Discrete Fourier transform
ECG	Electrocardiogram
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
FBS	Fetal blood sampling
FECG	Fetal electrocardiogram
FIGO	International Federation of Gynecology and Obstetrics
FFT	Fast Fourier transform
FHR	Fetal heart rate
FMCG	Fetal magnetocardiography
FN	False negative
FP	False positive
HF	High frequency
HMM	Hidden Markov Model
ICEEMDAN	Improved complete ensemble empirical mode decomposition with adaptive noise
IMF	Intrinsic mode function

I – FECG	Invasive fetal electrocardiogram
***k* – NN**	*k*-Nearest Neighbors
LDA	Linear discriminant analysis
LF	Low frequency
MF	Medium frequency
NI	Number of sifting iterations
NI – FECG	Non-invasive fetal electrocardiogram
NR	Number of realizations
Nstd	Noise standard deviation
PBL	Progressive baseline
PCA	Principal component analysis
PCG	Phonocardiography
pCO_2	Partial pressure of carbon dioxide
pH	Artery pH
pO_2	Partial pressure of oxygen
PO	Pulse oximetry
PSD	Power spectral density
QI	Quality metric
RMS	Root mean square
SampEn	Sample entropy
Se	Sensitivity
Sp	Specificity
SQUID	Superconducting quantum interference device
STAN	ST analysis
STFT	Short-time Fourier transform
SVM	Support vector machine
TN	True negative
TP	True positive
TV – AR	Time-varying autoregressive
UC	Uterine contraction
US	Ultrasound
VBL	Virtual baseline
VLF	Very low frequency

C

List of Figures

D

List of Tables

E

Bibliography

3B Scientific® (2020). Ctgi - simulierter fetalmonitor. `https://www.3bscientific.de/geburtshilfe,pg_895.html`. Accessed: 2020-03-12.

Abdulhay, E. W., Oweis, R. J., Alhaddad, A. M., Sublaban, F. N., Radwan, M. A., and Almasaeed, H. M. (2014). Monitoring techniques. *Biomedical Science*, 2(3):53–67.

Abramovich, Y. I., Spencer, N. K., and Turley, M. D. (2007). Time-varying autoregressive (tvar) models for multiple radar observations. *IEEE Transactions on Signal Processing*, 55(4):1298–1311.

Agostinelli, A., Braccili, E., Marchegiani, E., Rosati, R., Sbrollini, A., Burattini, L., Morettini, M., Di Nardo, F., Fioretti, S., and Burattini, L. (2017). Statistical baseline assessment in cardiotocography. In *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, pages 3166–3169. IEEE.

Al-Angari, H. M., Kimura, Y., Hadjileontiadis, L. J., and Khandoker, A. H. (2017). A hybrid emd-kurtosis method for estimating fetal heart rate from continuous doppler signals. *Frontiers in physiology*, 8:641.

Alfirevic, Z., Devane, D., Gyte, G. M., and Cuthbert, A. (2017). Continuous cardiotocography (ctg) as a form of electronic fetal monitoring (efm) for fetal assessment during labour. *The Cochrane Library*.

American College of Obstetricians and Gynecologists and others (2010). Practice bulletin no. 116: Management of intrapartum fetal heart rate tracings. *Obstetrics and gynecology*, 116(5):1232.

Artal, R. and O'toole, M. (2003). Guidelines of the american college of obstetricians and gynecologists for exercise during pregnancy and the postpartum period. *British journal of sports medicine*, 37(1):6–12.

Ayres-de Campos, D. and Bernardes, J. (1999). Early, variable and late decelerations: can a consensus be reached in their identification? *International Journal of Gynecology & Obstetrics*, 65(3):305–306.

Ayres-de Campos, D. and Bernardes, J. (2010). Twenty-five years after the figo guidelines for the use of fetal monitoring: Time for a simplified approach? *International Journal of Gynecology & Obstetrics*, 110(1):1–6.

Ayres-de Campos, D., Bernardes, J., Marsal, K., Nickelsen, C., Makarainen, L., Banfield, P., Xavier, P., and Campos, I. (2004). Can the reproducibility of fetal heart rate baseline estimation be improved? *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 112(1):49–54.

Ayres-de Campos, D., Sousa, P., Costa, A., and Bernardes, J. (2008). Omniview-sisporto® 3.5–a central fetal monitoring station with online alerts based on computerized cardiotocogram+ st event analysis. *Journal of Perinatal Medicine*, 36(3):260–264.

Ayres-de Campos, D., Spong, C. Y., Chandraharan, E., et al. (2015). Figo consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *International Journal of Gynecology & Obstetrics*, 131(1):13–24.

Bastos, L. F., Van Meurs, W., and Ayres-de Campos, D. (2012). A model for educational simulation of the evolution of uterine contractions during labor. *Computer methods and programs in biomedicine*, 107(2):242–247.

Beard, R., Filshie, G., Knight, C., and Roberts, G. (1971). The significance of the changes in the continuous fetal heart rate in the first stage of labour. *BJOG: An International Journal of Obstetrics & Gynaecology*, 78(10):865–881.

Behar, J., Andreotti, F., Zaunseder, S., Oster, J., and Clifford, G. D. (2016). A practical guide to non-invasive foetal electrocardiogram extraction and analysis. *Physiological measurement*, 37(5):R1.

Bernardes, J., Costa-Pereira, A., Ayres-de Campos, D., Van Geijn, H., and Pereira-Leite, L. (1997). Evaluation of interobserver agreement of cardiotocograms. *International Journal of Gynecology & Obstetrics*, 57(1):33–37.

Bhatnagar, D. and Maheshwari, P. (2016). Classification of cardiotocography data with weka. *International Journal of Computer Science and Network-IJCSN*, 5(2).

Bhide, A., Chandraharan, E., and Acharya, G. (2016). Fetal monitoring in labor: Implications of evidence generated by new systematic review. *Acta obstetricia et gynecologica Scandinavica*, 95(1):5–8.

Böttrich, M., Ley, S., and Husar, P. (2015). Simulation based investigation of source-detector configurations for non-invasive fetal pulse oximetry. *Current Directions in Biomedical Engineering*, 1(1):450–453.

Bursa, M. and Lhotska, L. (2017). The use of convolutional neural networks in biomedical data processing. In *International Conference on Information Technology in Bio-and Medical Informatics*, pages 100–119. Springer.

Cahill, A. G., Roehl, K. A., Odibo, A. O., and Macones, G. A. (2012). Association and prediction of neonatal acidemia. *American journal of obstetrics and gynecology*, 207(3):206–e1.

Campbell, J. Q., Best, T. H., Eswaran, H., and Lowery, C. L. (2006). Fetal and maternal magnetocardiography during flecainide therapy for supraventricular tachycardia. *Obstetrics & Gynecology*, 108(3, Part 2):767–771.

Cao, H., Lake, D. E., Ferguson, J. E., Chisholm, C. A., Griffin, M. P., and Moorman, J. R. (2006). Toward quantitative fetal heart rate monitoring. *IEEE Transactions on Biomedical Engineering*, 53(1):111–118.

Cattani, C., Doubrovina, O., Rogosin, S., Voskresensky, S. L., and Zelianko, E. (2006). On the creation of a new diagnostic model for fetal well-being on the base of wavelet analysis of cardiotocograms. *Journal of medical systems*, 30(6):489–494.

Cazares, S., Moulden, M., Redman, C. W., and Tarassenko, L. (2001). Tracking poles with an autoregressive model: a confidence index for the analysis of the intrapartum cardiotocogram. *Medical engineering & physics*, 23(9):603–614.

Cesarelli, M., Romano, M., and Bifulco, P. (2009). Comparison of short term variability indexes in cardiotocographic foetal monitoring. *Computers in biology and medicine*, 39(2):106–118.

Cesarelli, M., Romano, M., Bifulco, P., Fedele, F., and Bracale, M. (2007). An algorithm for the recovery of fetal heart rate series from ctg data. *Computers in biology and medicine*, 37(5):663–669.

Cesarelli, M., Romano, M., D'Addio, G., Ruffo, M., Bifulco, P., Pasquariello, G., and Fratini, A. (2011). Floatingline estimation in fhr signal analysis. In *5th European Conference of the International Federation for Medical and Biological Engineering*, pages 179–182. Springer.

Cesarelli, M., Romano, M., Ruffo, M., Bifulco, P., and Pasquariello, G. (2010). Foetal heart rate variability frequency characteristics with respect to uterine contractions. *Journal of Biomedical Science and Engineering*, 3(10):1014.

Chudáček, V., Spilka, J., Burša, M., Janku, P., Hruban, L., Huptych, M., and Lhotská, L. (2014). Open access intrapartum CTG database. *BMC pregnancy and childbirth*, 14(1):16.

Clark, S. L., Hamilton, E. F., Garite, T. J., Timmins, A., Warrick, P. A., and Smith, S. (2017). The limits of electronic fetal heart rate monitoring in the prevention of neonatal metabolic acidemia. *American journal of obstetrics and gynecology*, 216(2):163–e1.

Clark, S. L. and Hankins, G. D. (2003). Temporal and demographic trends in cerebral palsy-fact and fiction. *American journal of obstetrics and gynecology*, 188(3):628–633.

Clifford, G. D., Silva, I., Behar, J., and Moody, G. B. (2014). Non-invasive fetal ecg analysis. *Physiological measurement*, 35(8):1521.

Coletta, J., Murphy, E., Rubeo, Z., and Gyamfi-Bannerman, C. (2012). The 5-tier system of assessing fetal heart rate tracings is superior to the 3-tier system in identifying fetal acidemia. *American journal of obstetrics and gynecology*, 206(3):226–e1.

Colominas, M. A., Schlotthauer, G., and Torres, M. E. (2014). Improved complete ensemble emd: A suitable tool for biomedical signal processing. *Biomedical Signal Processing and Control*, 14:19–29.

Cömert, Z., Kocamaz, A. F., and Güngör, S. (2016). Cardiotocography signals with artificial neural network and extreme learning machine. In *Signal Processing and Communication Application Conference (SIU), 2016 24th*, pages 1493–1496. IEEE.

Cömert, Z., Kocamaz, A. F., and Subha, V. (2018). Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Computers in biology and medicine*, 99:85–97.

Costa, M. D., Schnettler, W. T., Amorim-Costa, C., Bernardes, J., Costa, A., Goldberger, A. L., and Ayres-de Campos, D. (2014). Complexity-loss in fetal heart rate dynamics during labor as a potential biomarker of acidemia. *Early human development*, 90(1):67–71.

Cover, T. M., Hart, P., et al. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.

Cristianini, N., Shawe-Taylor, J., et al. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Cuneo, B. F., Strasburger, J. F., Yu, S., Horigome, H., Hosono, T., Kandori, A., and Wakai, R. T. (2013). In utero diagnosis of long qt syndrome by magnetocardiography. *Circulation*, 128(20):2183–2191.

Czabanski, R., Jezewski, M., Wrobel, J., Horoba, K., and Jezewski, J. (2008). A neuro-fuzzy approach to the classification of fetal cardiotocograms. In *14th Nordic-Baltic Conference on Biomedical Engineering and Medical Physics*, pages 446–449. Springer.

Das, S., Roy, K., and Saha, C. (2011). A novel approach for extraction and analysis of variability of baseline. In *Recent Trends in Information Systems (ReTIS), 2011 International Conference on*, pages 336–339. IEEE.

Devoe, L., Golde, S., Kilman, Y., Morton, D., Shea, K., and Waller, J. (2000). A comparison of visual analyses of intrapartum fetal heart rate tracings according to the new national institute of child health and human development guidelines with computer analyses by an automated fetal heart rate monitoring system. *American journal of obstetrics and gynecology*, 183(2):361–366.

Devoe, L. D. (2016). Future perspectives in intrapartum fetal surveillance. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 30:98–106.

Di Tommaso, M., Seravalli, V., Cordisco, A., Consorti, G., Mecacci, F., and Rizzello, F. (2013). Comparison of five classification systems for interpreting electronic fetal monitoring in predicting neonatal status at birth. *The Journal of Maternal-Fetal & Neonatal Medicine*, 26(5):487–490.

Dong, S., Azemi, G., and Boashash, B. (2014). Improved characterization of hrv signals based on instantaneous frequency features estimated from quadratic time–frequency distributions with data-adapted kernels. *Biomedical Signal Processing and Control*, 10:153–165.

Doria, V., Papageorghiou, A., Gustafsson, A., Ugwumadu, A., Farrer, K., and Arulkumaran, S. (2007). Review of the first 1502 cases of ecg-st waveform analysis during labour in a teaching hospital. *BJOG: An International Journal of Obstetrics & Gynaecology*, 114(10):1202–1207.

East, C. E., Begg, L., Colditz, P. B., and Lau, R. (2014). Fetal pulse oximetry for fetal assessment in labour. *Cochrane Database Syst Rev*, 10.

Elliott, C., Warrick, P. A., Graham, E., and Hamilton, E. F. (2010). Graded classification of fetal heart rate tracings: association with neonatal metabolic acidosis and neurologic morbidity. *American journal of obstetrics and gynecology*, 202(3):258–e1.

Eun, S., Lee, J. M., Yi, D. Y., Lee, N. M., Kim, H., Yun, S. W., Lim, I., Choi, E. S., and Chae, S. A. (2016). Assessment of the association between apgar scores and seizures in infants less than 1 year old. *Seizure*, 37:48–54.

Evans, M. I., Eden, R. D., Britt, D. W., Evans, S. M., and Schifrin, B. S. (2018). Re-engineering the interpretation of electronic fetal monitoring to identify reversible risk for cerebral palsy: a case control series. *The Journal of Maternal-Fetal & Neonatal Medicine*, pages 1–9.

Feng, G., Quirk, J. G., and Djurić, P. M. (2019). Inference about causality from cardiotocography signals using gaussian processes. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2852–2856. IEEE.

Fergus, P., Selvaraj, M., and Chalmers, C. (2018). Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using cardiotocography traces. *Computers in biology and medicine*, 93:7–16.

Ferrario, M., Signorini, M. G., and Magenes, G. (2009). Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses. *Medical & biological engineering & computing*, 47(9):911–919.

Ferrario, M., Signorini, M. G., Magenes, G., and Cerutti, S. (2006). Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress. *IEEE Transactions on Biomedical Engineering*, 53(1):119–125.

Ferreira Jr, M. and Zanesco, A. (2016). Heart rate variability as important approach for assessment autonomic modulation. *Motriz: Revista de Educação Física*, 22(2):3–8.

Freeman, R. K., Garite, T. J., Nageotte, M. P., and Miller, L. A. (2012). *Fetal heart rate monitoring*. Lippincott Williams & Wilkins.

Fritz, J., Walker, D. M., Cohen, S., Angeles, G., and Lamadrid-Figueroa, H. (2017). Can a simulation-based training program impact the use of evidence based routine practices at birth? results of a hospital-based cluster randomized trial in mexico. *PloS one*, 12(3):e0172623.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2017a). Analysis of the foetal heart rate in cardiotocographic recordings through a progressive characterization of decelerations. *Current Directions in Biomedical Engineering*, 3(2):423–427.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2017b). Progressive fetal distress estimation by characterization of fetal heart rate decelerations response based on signal variability in cardiotocographic recordings. In *2017 Computing in Cardiology (CinC)*, pages 1–4. IEEE.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2018a). Foetal heart rate signal spectral analysis by using time-varying autoregressive modelling. *Current Directions in Biomedical Engineering*, 4(1):579–582.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2018b). Spectral-based analysis of progressive dynamical changes in the fetal heart rate signal during labor by using empirical mode decomposition. In *2018 Computing in Cardiology (CinC)*, pages 1–4. IEEE.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2019a). Automatic fetal distress assessment during labor based on a progressive analysis of the cardiotocographic recording. In *1.Doktorandentagung am 28.& 29. Januar, Magdeburg, Germany, 2019*. Faculty of Computer Science, Otto-von-Guericke University.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2019b). Cardiotocograph data classification improvement by using empirical mode decomposition. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2019c). Cardiotocographic signal feature extraction through ceemdan and time-varying autoregressive spectral-based analysis for fetal welfare assessment. *IEEE Access*.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2019d). Foetal heart rate assessment by empirical mode decomposition and spectral analysis. *Current Directions in Biomedical Engineering*, 5(1):381–383.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2019e). Independent analysis of decelerations and resting periods through ceemdan and spectral-based feature extraction improves cardiotocographic assessment. *Applied Sciences*, 9(24):5421.

Fuentealba, P., Illanes, A., and Ortmeier, F. (2019f). A study on the classification performance of cardiotocographic data vs. class formation criteria. In *2019 International Student Conference (ISC) IEEE EMBS, Magdeburg, Germany*. IEEE.

Fukushima, A., Nakai, K., Kanasugi, T., Terata, M., and Sugiyama, T. (2011). Assessment of fetal autonomic nervous system activity by fetal magnetocardiography: comparison of normal pregnancy and intrauterine growth restriction. *Journal of pregnancy*, 2011.

Garabedian, C., De Jonckheere, J., Butruille, L., Deruelle, P., Storme, L., and Houfflin-Debarge, V. (2017). Understanding fetal physiology and second line monitoring during labor. *Journal of gynecology obstetrics and human reproduction*, 46(2):113–117.

Garibaldi, J. M., Westgate, J. A., and Ifeachor, E. C. (1999). The evaluation of an expert system for the analysis of umbilical cord blood. *Artificial intelligence in medicine*, 17(2):109–130.

Georgieva, A., Payne, S. J., Moulden, M., and Redman, C. W. (2013). Artificial neural networks applied to fetal monitoring in labour. *Neural Computing and Applications*, 22(1):85–93.

Georgoulas, G., Karvelis, P., Spilka, J., Chudáček, V., Stylios, C. D., and Lhotská, L. (2017). Investigating ph based evaluation of fetal heart rate (fhr) recordings. *Health and Technology*, 7(2-3):241–254.

Georgoulas, G., Stylios, D., and Groumpos, P. (2006). Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines. *IEEE Transactions on biomedical engineering*, 53(5):875–884.

Georgoulas, G. G., Stylios, C., Nokas, G., and Groumpos, P. P. (2004). Classification of fetal heart rate during labour using hidden markov models. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 3, pages 2471–2475. IEEE.

Ghi, T., Morganelli, G., Bellussi, F., Rucci, P., Giorgetta, F., Rizzo, N., Frusca, T., and Pilu, G. (2016). Cardiotocographic findings in the second stage of labor among fetuses delivered with acidemia: a comparison of two classification systems. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 203:297–302.

Gibbons, J. D. and Chakraborti, S. (2011). Nonparametric statistical inference. In *International encyclopedia of statistical science*, pages 977–979. Springer.

Gibbons, L., Belizán, J. M., Lauer, J. A., Betrán, A. P., Merialdi, M., Althabe, F., et al. (2010). The global numbers and costs of additionally needed and unnecessary caesarean sections performed per year: overuse as a barrier to universal coverage. *World health report*, 30:1–31.

GOLDABER, K. G., GILSTRAP III, L. C., LEVENO, K. J., DAX, J. S., and McINTIRE, D. D. (1991). Pathologic fetal acidemia. *Obstetrics & Gynecology*, 78(6):1103–1107.

Gonçalves, H., Amorim-Costa, C., Ayres-de Campos, D., and Bernardes, J. (2018). Evolution of linear and nonlinear fetal heart rate indices throughout pregnancy in appropriate, small for gestational age and preterm fetuses: A cohort study. *Computer methods and programs in biomedicine*, 153:191–199.

Gonçalves, H., Rocha, A. P., Ayres-de Campos, D., and Bernardes, J. (2006). Internal versus external intrapartum foetal heart rate monitoring: the effect on linear and nonlinear parameters. *Physiological measurement*, 27(3):307.

Granero-Belinchon, C., Roux, S. G., Abry, P., Doret, M., and Garnier, N. B. (2017). Information theory to probe intrapartum fetal heart rate dynamics. *Entropy*, 19(12):640.

Gupta, S., Natarajan, G., Gupta, D., Karnati, S., Dwaihy, M., Wang, B., and Chawla, S. (2016). Variability in apgar score assignment among clinicians: Role of a simple clarification. *American journal of perinatology*.

Hamilton, E. and Kimanani, E. K. (1994). 3 intrapartum prediction of fetal status and assessment of labour progress. *Baillière's clinical obstetrics and gynaecology*, 8(3):567–581.

Haran, S. S. and Everett, T. R. (2017). Antenatal fetal wellbeing. *Obstetrics, Gynaecology and Reproductive Medicine*, 27(2):44–49.

Haritopoulos, M., Illanes, A., and Nandi, A. K. (2016). Survey on cardiotocography feature extraction algorithms for foetal welfare assessment. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, pages 1193–1198. Springer.

He, H., Bai, Y., Garcia, E. A., and Li, S. (2008). Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pages 1322–1328. IEEE.

Hruban, L., Spilka, J., Chudáček, et al. (2015). Agreement on intrapartum cardiotocogram recordings between expert obstetricians. *Journal of evaluation in clinical practice*.

Huang, J. (2014a). Study of autoregressive) ar (spectrum estimation algorithm for vibration signals of industrial stream turbines. *International Journal of Control and Automation*, 7(8).

Huang, M.-L. and Hsu, Y.-Y. (2012). Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. *Journal of Biomedical Science and Engineering*, 5(09):526.

Huang, N. E. (2014b). Introduction to the hilbert–huang transform and its related mathematical problems. In *Hilbert–Huang transform and its applications*, pages 1–26. World Scientific.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, volume 454, pages 903–995. The Royal Society.

HUNTLEIGH - Arjo (2020). Sonicaid FetalCare3 Doctor's Office & Clinics. `http://www.huntleigh-diagnostics.com/store/obstetric-monitoring/sonicaid-fetal-monitoring/software-fetal`. Accessed: 2020-03-12.

Illanes, A. and Haritopoulos, M. (2015). Fetal heart rate feature extraction from cardiotocographic recordings through autoregressive model's power spectral-and pole-based analysis. In *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, pages 5842–5845. IEEE.

Illanes, A., Haritopoulos, M., Robles, F., and Guerra, F. (2015). A qualitative dynamical model for cardiotocography simulation. In *Computing in Cardiology Conference (CinC), 2015*, pages 129–132. IEEE.

Improta, G., Romano, M., Ponsiglione, A., Bifulco, P., Faiella, G., and Cesarelli, M. (2014). Computerized cardiotocography: a software to generate synthetic signals. *Journal of Health & Medical Informatics*, 5(4):162.

iSimulate CTGi (2020). Real-time fetal heart rate simulator. `https://www.isimulate.com/ctgi/`. Accessed: 2020-03-12.

Jezewski, M., Wrobel, J., Labaj, P., Leski, J., Henzel, N., Horoba, K., and Jezewski, J. (2007). Some practical remarks on neural networks approach to fetal cardiotocograms classification. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pages 5170–5173. IEEE.

Jiménez-González, A. and James, C. J. (2009). Extracting sources from noisy abdominal phonograms: a single-channel blind source separation method. *Medical & biological engineering & computing*, 47(6):655–664.

Jimenez-Gonzalez, A. and James, C. J. (2010). Time-structure based reconstruction of physiological independent sources extracted from noisy abdominal phonograms. *Biomedical Engineering, IEEE Transactions on*, 57(9):2322–2330.

Jongen, G. J., van der Hout-van, M. B., Oei, S. G., van de Vosse, F. N., Bovendeerd, P. H., et al. (2017). Simulation of fetal heart rate variability with a mathematical model. *Medical Engineering & Physics*, 42:55–64.

Jongen, G. J., van der Hout-van, M. B., van de Vosse, F. N., Oei, S. G., Bovendeerd, P. H., et al. (2016a). A mathematical model to simulate the cardiotocogram during labor. part a: Model setup and simulation of late decelerations. *Journal of biomechanics*, 49(12):2466–2473.

Jongen, G. J., van der Hout-van, M. B., van de Vosse, F. N., Oei, S. G., Bovendeerd, P. H., et al. (2016b). A mathematical model to simulate the cardiotocogram during labor. part b: Parameter estimation and simulation of variable decelerations. *Journal of biomechanics*, 49(12):2474–2480.

K2 Medical Systems^{TM} (2020). K2 infant-guardian®. `https://www.k2ms.com/infant-guardian/`. Accessed: 2020-03-12.

K2 Medical Systems^{TM} PTP (2020). Perinatal training programme. `https://www.k2ms.com/ptp/`. Accessed: 2020-03-12.

Kahankova, R., Martinek, R., Jaros, R., Behbehani, K., Matonia, A., Jezewski, M., and Behar, J. A. (2019). A review of signal processing techniques for non-invasive fetal electrocardiography. *IEEE reviews in biomedical engineering*.

Kay, S. M. and Marple, S. L. (1981). Spectrum analysis—a modern perspective. *Proceedings of the IEEE*, 69(11):1380–1419.

Keith, R. D. and Greene, K. R. (1994). 4 development, evaluation and validation of an intelligent system for the management of labour. *Bailliere's clinical obstetrics and gynaecology*, 8(3):583–605.

Khokhlova, L., Seleznev, A., Zhdanov, D., Zemlyakov, I. Y., and Kiseleva, E. Y. (2016). A device for fetal monitoring by means of control over cardiovascular parameters based on acoustic data. In *Journal of Physics: Conference Series*, volume 671, page 012064. IOP Publishing.

Kolip, P., Nolting, H.-D., and Zich, K. (2012). Faktencheck gesundheit. In *Kaiserschnittgeburten–Entwicklung und regionale Verteilung*. Bertelsmann-Stiftung.

Kovács, F., Horváth, C., Balogh, Á. T., and Hosszú, G. (2011). Fetal phonocardiography—past and future possibilities. *Computer methods and programs in biomedicine*, 104(1):19–25.

Krupa, B., Ali, M. M., and Zahedi, E. (2009). The application of empirical mode decomposition for the enhancement of cardiotocograph signals. *Physiological measurement*, 30(8):729.

Krupa, N., Ali, M., Zahedi, E., Ahmed, S., and Hassan, F. M. (2011). Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine. *Biomedical engineering online*, 10:6–6.

Kuehnle, E., Herms, S., Kohls, F., Kundu, S., Hillemanns, P., and Staboulidou, I. (2016). Correlation of fetal scalp blood sampling ph with neonatal outcome umbilical artery ph value. *Archives of gynecology and obstetrics*, pages 1–8.

Kumar, N., Suman, A., and Sawant, K. (2016). Relationship between immediate postpartum umbilical cord blood ph and fetal distress. *International Journal of Contemporary Pediatrics*, 3(1):113–119.

Kupka, T., Wrobel, J., Jezewski, J., Gacek, A., and Jezewski, M. (2006). Evaluation of fetal heart rate baseline estimation method using testing signals based on a statistical model. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 3728–3731. IEEE.

Kwak, S. G. and Park, S.-H. (2019). Normality test in clinical research. *Journal of Rheumatic Diseases*, 26(1):5–11.

Kwon, J. Y., Park, I. Y., Shin, J. C., Song, J., Tafreshi, R., and Lim, J. (2012). Specific change in spectral power of fetal heart rate variability related to fetal acidemia during labor: comparison between preterm and term fetuses. *Early human development*, 88(4):203–207.

Lewis, D., Downe, S., et al. (2015). Figo consensus guidelines on intrapartum fetal monitoring: Intermittent auscultation. *International Journal of Gynecology & Obstetrics*, 131(1):9–12.

Li, J. and Farajidavar, A. (2015). A portable magnetocardiography system: Using a magnetoimpedance sensor. In *Biomedical Circuits and Systems Conference (BioCAS), 2015 IEEE*, pages 1–4. IEEE.

Li, Y., Nie, W., Ye, F., and Li, A. (2016). A fetal electrocardiogram signal extraction algorithm based on the temporal structure and the non-gaussianity. *Computational and mathematical methods in medicine*, 2016.

Lu, Y., Li, X., Wei, S., and Liu, X. (2014). Fetal heart rate baseline estimation with analysis of fetal movement signal. *Bio-medical materials and engineering*, 24(6):3763–3769.

Lutomski, J. E., Meaney, S., Greene, R. A., Ryan, A. C., and Devane, D. (2013). Expert systems for fetal assessment in labour. *Cochrane Database Syst Rev*.

Lutomski, J. E., Meaney, S., Greene, R. A., Ryan, A. C., and Devane, D. (2015). Expert systems for fetal assessment in labour. *The Cochrane Library*.

MacLennan, A. (1999). A template for defining a causal relation between acute intrapartum events and cerebral palsy: international consensus statement. *BMJ: British Medical Journal*, 319(7216):1054.

Macones, G. A., Hankins, G. D., Spong, C. Y., Hauth, J., and Moore, T. (2008). The 2008 national institute of child health and human development workshop report on electronic fetal monitoring: update on definitions, interpretation, and research guidelines. *Journal of Obstetric, Gynecologic, & Neonatal Nursing*, 37(5):510–515.

Maeda, K., Utsu, M., Noguchi, Y., Matsumoto, F., and Nagasawa, T. (2012). Central computerized automatic fetal heart rate diagnosis with a rapid and direct alarm system. *The Open Medical Devices Journal*, 4(1).

Magenes, G., Bellazzi, R., Fanelli, A., and Signorini, M. (2014). Multivariate analysis based on linear and non-linear fhr parameters for the identification of iugr fetuses. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 1868–1871. IEEE.

Manganaro, R., Mamì, C., and Gemelli, M. (1994). The validity of the apgar scores in the assessment of asphyxia at birth. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 54(2):99–102.

Marques, J. A. L., Cortez, P. C., Madeiro, J. P., de Albuquerque, V. H. C., Fong, S. J., and Schlindwein, F. S. (2020). Nonlinear characterization and complexity analysis of cardiotocographic examinations using entropy measures. *The Journal of Supercomputing*, 76(2):1305–1320.

Marques, J. A. L., Cortez, P. C., Madeiro, J. P. D. V., Fong, S. J., Schlindwein, F. S., and De Albuquerque, V. H. C. (2019). Automatic cardiotocography diagnostic system based on hilbert transform and adaptive threshold technique. *IEEE Access*, 7:73085–73094.

Martinek, R., Sincl, A., Vanus, J., Kelnar, M., Bilik, P., Machacek, Z., and Zidek, J. (2016). Modelling of fetal hypoxic conditions based on virtual instrumentation. In *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015*, pages 249–259. Springer.

Martinek, R. and Žídek, J. (2012). A system for improving the diagnostic quality of fetal electrocardiogram.

Maso, G., Piccoli, M., De Seta, F., Parolin, S., Banco, R., Mattos, L. C., Bogatti, P., and Alberico, S. (2015). Intrapartum fetal heart rate monitoring interpretation in labour: a critical appraisal. *Minerva Ginecol*, 67(1):65–79.

Mika, S., Ratsch, G., Weston, J., Scholkopf, B., and Mullers, K.-R. (1999). Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468)*, pages 41–48. Ieee.

Monteiro-Santos, J., Henriques, T., Nunes, I., Amorim-Costa, C., Bernardes, J., and Costa-Santos, C. (2020). Complexity of cardiotocographic signals as a predictor of labor. *Entropy*, 22(1):104.

Nagendra, V., Gude, H., Sampath, D., Corns, S., and Long, S. (2017). Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data. In *Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2017 IEEE Conference on*, pages 1–6. IEEE.

Nageotte, M. P. (2015). Fetal heart rate monitoring. In *Seminars in Fetal and Neonatal Medicine*, volume 20, pages 144–148. Elsevier.

Nunes, I. and Ayres-de Campos, D. (2016). Computer analysis of foetal monitoring signals. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 30:68–78.

Ocak, H. (2013). A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being. *Journal of medical systems*, 37(2):9913.

Ortiz, M., Bojorges, E., Aguilar, S., Echeverria, J., Gonzalez-Camarena, R., Carrasco, S., Gaitan, M., and Martinez, A. (2005). Analysis of high frequency fetal heart rate variability using the empirical mode decomposition. *Computers in Cardiology*, 32(1):675–678.

Pardey, J., Moulden, M., and Redman, C. W. (2002). A computer system for the numerical analysis of nonstress tests. *American journal of obstetrics and gynecology*, 186(5):1095–1103.

Pardey, J., Roberts, S., and Tarassenko, L. (1996). A review of parametric modelling techniques for eeg analysis. *Medical engineering & physics*, 18(1):2–11.

Parer, J. (1980). The effect of acute maternal hypoxia on fetal oxygenation and the umbilical circulation in the sheep. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 10(2):125–136.

Parer, J., King, T., Flanders, S., Fox, M., and Kilpatrick, S. (2006). Fetal acidemia and electronic fetal heart rate patterns: Is there evidence of an association? *The Journal of Maternal-Fetal & Neonatal Medicine*, 19(5):289–294.

Parer, J. T. (2014). Standardization of fetal heart rate pattern management: is international consensus possible? *Hypertension Research in Pregnancy*, 2(2):51–58.

Parer, J. T. and Hamilton, E. F. (2010). Comparison of 5 experts and computer analysis in rule-based fetal heart rate interpretation. *American journal of obstetrics and gynecology*, 203(5):451–e1.

Parer, J. T. and Ikeda, T. (2007). A framework for standardized management of intrapartum fetal heart rate patterns. *American journal of obstetrics and gynecology*, 197(1):26–e1.

Pehrson, C., Sorensen, J., and Amer-Wåhlin, I. (2011). Evaluation and impact of cardiotocography training programmes: a systematic review. *BJOG: An International Journal of Obstetrics & Gynaecology*, 118(8):926–935.

Perveen, F., Khan, A., Ali, T., and Rabia, S. (2015). Umbilical cord blood ph in intrapartum hypoxia. *J Coll Physicians Surg Pak*, 25(9):667–670.

Peters, M., Crowe, J., Piéri, J.-F., Quartero, H., Hayes-Gill, B., James, D., Stinstra, J., and Shakespeare, S. (2001). Monitoring the fetal heart non-invasively: a review of methods. *Journal of perinatal medicine*, 29(5):408–416.

Petrozziello, A., Jordanov, I., Papageorghiou, T. A., Redman, W. C., and Georgieva, A. (2018). Deep learning for continuous electronic fetal monitoring in labor. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5866–5869. IEEE.

PHC NORDIC (2020). Ctgi real-time obstetric simulator. `https://phcnordic.com/isimulate/ctgi/`. Accessed: 2020-03-12.

Pinas, A. and Chandraharan, E. (2016). Continuous cardiotocography during labour: Analysis, classification and management. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 30:33–47.

Pincus, S. (1995). Approximate entropy (apen) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 5(1):110–117.

Pincus, S. M. and Viscarello, R. R. (1992). Approximate entropy: a regularity measure for fetal heart rate analysis. *Obstet Gynecol*, 79(2):249–255.

Richman, J. S. and Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049.

Rivolta, M., Stampalija, T., Frash, M., and Sassi, R. (2019). Theoretical value of deceleration capacity points to deceleration reserve of fetal heart rate. *IEEE transactions on bio-medical engineering*.

Romano, M., Bifulco, P., Cesarelli, M., Sansone, M., and Bracale, M. (2006a). Foetal heart rate power spectrum response to uterine contraction. *Medical and Biological Engineering and Computing*, 44(3):188–201.

Romano, M., Bracale, M., Cesarelli, M., Campanile, M., Bifulco, P., De Falco, M., Sansone, M., and Di Lieto, A. (2006b). Antepartum cardiotocography: A study of fetal reactivity in frequency domain. *Computers in biology and medicine*, 36(6):619–633.

Romano, M., Faiella, G., Clemente, F., Iuppariello, L., Bifulco, P., and Cesarelli, M. (2016a). Analysis of foetal heart rate variability components by means of empirical mode decomposition. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, pages 71–74. Springer.

Romano, M., Iuppariello, L., Ponsiglione, A. M., Improta, G., Bifulco, P., and Cesarelli, M. (2016b). Frequency and time domain analysis of foetal heart rate variability with traditional indexes: A critical survey. *Computational and Mathematical Methods in Medicine*, 2016.

Rooth, G., Huch, A., and Huch, R. (1987). Figo news: guidelines for the use of fetal monitoring. *Int J Gynecol Obstet*, 25:159–67.

Rosén, K. G., Blad, S., Larsson, D., Norén, H., and Outram, N. (2007). Assessment of the fetal bioprofile during labor by fetal ecg analysis. *Expert Review of Obstetrics & Gynecology*, 2(5):609–620.

Rotariu, C., Pasarica, A., Andruseac, G., Costin, H., and Nemescu, D. (2014a). Automatic analysis of the fetal heart rate variability and uterine contractions. In *2014 International Conference and Exposition on Electrical and Power Engineering (EPE)*, pages 553–556. IEEE.

Rotariu, C., Pasarica, A., Costin, H., and Nemescu, D. (2014b). Spectral analysis of fetal heart rate variability associated with fetal acidosis and base deficit values. In *2014 International Conference on Development and Application Systems (DAS)*, pages 210–213. IEEE.

Royal College of Obstetricians and Gynaecologists (Great Britain). Clinical Effectiveness Support Unit (2001). *The use of electronic fetal monitoring: The use and interpretation of cardiotocography in intrapartum fetal surveillance*. Royal College of Obstetricians and Gynaecologists.

Saccone, G., Schuit, E., Amer-Wåhlin, I., Xodo, S., and Berghella, V. (2016). Electrocardiogram st analysis during labor: A systematic review and meta-analysis of randomized controlled trials. *Obstetrics & Gynecology*, 127(1):127–135.

Sahin, H. and Subasi, A. (2015). Classification of the cardiotocogram data for anticipation of fetal risks using machine learning techniques. *Applied Soft Computing*, 33:231–238.

Salamalekis, E., Hintipas, E., Salloum, I., Vasios, G., Loghis, C., Vitoratos, N., Chrelias, C., and Creatsas, G. (2006). Computerized analysis of fetal heart rate variability using the matching pursuit technique as an indicator of fetal hypoxia during labor. *The Journal of Maternal-Fetal & Neonatal Medicine*, 19(3):165–169.

Saleem, S., Naqvi, S. S., Manzoor, T., Saeed, A., Mirza, J., et al. (2019). A strategy for classification ofvaginal vs. cesarean section'delivery: bivariate empirical mode decomposition of cardiotocographic recordings. *Frontiers in Physiology*, 10:246.

Sameni, R., Shamsollahi, M., and Jutten, C. (2008). Model-based bayesian filtering of cardiac contaminants from biomedical recordings. *Physiological Measurement*, 29(5):595.

Sankhe, M. S. and Desai, K. D. (2016). Fetal heart rate variability: multiple regression models using autoregressive analysis and fast fourier transform. In *Innovations in Bio-Inspired Computing and Applications*, pages 447–462. Springer.

Santo, S. and Ayres-de Campos, D. (2012). Human factors affecting the interpretation of fetal heart rate tracings: an update. *Current Opinion in Obstetrics and Gynecology*, 24(2):84–88.

Santo, S., Ayres-de Campos, D., Costa-Santos, C., Schnettler, W., Ugwumadu, A., and Da Graça, L. M. (2017). Agreement and accuracy using the figo, acog and nice cardiotocography interpretation guidelines. *Acta obstetricia et gynecologica Scandinavica*, 96(2):166–175.

Sartwelle, T. P., Johnston, J. C., Arda, B., and Zebenigus, M. (2019). Cerebral palsy, cesarean sections, and electronic fetal monitoring: All the light we cannot see. *Clinical Ethics*, 14(3):107–114.

Schiermeier, S., Pildner von Steinburg, S., Thieme, A., Reinhard, J., Daumer, M., Scholz, M., Hatzmann, W., and Schneider, K. (2008). Sensitivity and specificity of intrapartum computerised figo criteria for cardiotocography and fetal scalp ph during labour: multicentre, observational study. *BJOG: An International Journal of Obstetrics & Gynaecology*, 115(12):1557–1563.

Schneider, U., Schleussner, E., Fiedler, A., Jaekel, S., Liehr, M., Haueisen, J., and Hoyer, D. (2009). Fetal heart rate variability reveals differential dynamics in the intrauterine development of the sympathetic and parasympathetic branches of the autonomic nervous system. *Physiological measurement*, 30(2):215.

Schönauer, M., Thomas, A., Morbach, S., Niebauer, J., Schönauer, U., and Thiele, H. (2008). Cardiac autonomic diabetic neuropathy. *Diabetes and Vascular Disease Research*, 5(4):336–344.

Signorini, M. G., De Angelis, A., Magenes, G., Sassi, R., Arduini, D., and Cerutti, S. (2000). Classification of fetal pathologies through fuzzy inference systems based on a multiparametric analysis of fetal heart rate. In *Computers in Cardiology 2000*, pages 435–438. IEEE.

Signorini, M. G. and Magenes, G. (2016). Advanced signal processing techniques for ctg analysis. In *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, pages 1205–1210. Springer.

Signorini, M. G., Magenes, G., Cerutti, S., and Arduini, D. (2003). Linear and nonlinear parameters for the analysisof fetal heart rate signal from cardiotocographic recordings. *IEEE Transactions on Biomedical Engineering*, 50(3):365–374.

Signorini, M. G., Pini, N., Malovini, A., Bellazzi, R., and Magenes, G. (2020). Dataset on linear and non-linear indices for discriminating healthy and iugr fetuses. *Data in brief*, 29:105164.

Siira, S. M., Ojala, T. H., Vahlberg, T. J., Jalonen, J. O., Välimäki, I. A., Rosén, K. G., and Ekholm, E. M. (2005). Marked fetal acidosis and specific changes in power spectrum analysis of fetal heart rate variability recorded during the last hour of labour. *BJOG: An International Journal of Obstetrics & Gynaecology*, 112(4):418–423.

Siira, S. M., Ojala, T. H., Vahlberg, T. J., Rosén, K. G., and Ekholm, E. M. (2013). Do spectral bands of fetal heart rate variability associate with concomitant fetal scalp ph? *Early human development*, 89(9):739–742.

Sletten, J., Kiserud, T., and Kessler, J. (2016). Effect of uterine contractions on fetal heart rate in pregnancy: a prospective observational study. *Acta obstetricia et gynecologica Scandinavica*, 95(10):1129–1135.

Spilka, J., Abry, P., Goncalves, P., and Doret, M. (2014a). Impacts of first and second labour stages on hurst parameter based intrapartum fetal heart rate analysis. In *Computing in Cardiology 2014*, pages 777–780. IEEE.

Spilka, J., Chudáček, V., Janku, P., Hruban, L., Burša, M., Huptych, M., Zach, L., and Lhotská, L. (2014b). Analysis of obstetricians decision making on ctg recordings. *Journal of biomedical informatics*, 51:72–79.

Spilka, J., Chudácek, V., Koucký, M., and Lhotská, L. (2009). Assessment of non-linear features for intrapartal fetal heart rate classification. In *2009 9th International Conference on Information Technology and Applications in Biomedicine*, pages 1–4. IEEE.

Spilka, J., Chudáček, V., Koucký, M., Lhotská, L., Huptych, M., Janku, P., Georgoulas, G., and Stylios, C. (2012). Using nonlinear features for fetal heart rate classification. *Biomedical Signal Processing and Control*, 7(4):350–357.

Spilka, J., Frecon, J., Leonarduzzi, R., Pustelnik, N., Abry, P., and Doret, M. (2016a). Sparse support vector machine for intrapartum fetal heart rate classification. *IEEE journal of biomedical and health informatics*, 21(3):664–671.

Spilka, J., Frecon, J., Leonarduzzi, R., Pustelnik, N., Abry, P., and Doret, M. (2017). Sparse support vector machine for intrapartum fetal heart rate classification. *IEEE journal of biomedical and health informatics*, 21(3):664–671.

Spilka, J., Georgoulas, G., Karvelis, P., Chudáček, V., Stylios, C. D., and Lhotská, L. (2014c). Discriminating normal from "abnormal" pregnancy cases using an automated fhr evaluation method. In *Hellenic Conference on Artificial Intelligence*, pages 521–531. Springer.

Spilka, J., Georgoulas, G., Karvelis, P., Oikonomou, V. P., Chudáček, V., Stylios, C., Lhotská, L., and Janku, P. (2013). Automatic evaluation of fhr recordings from ctu-uhb ctg database. In *Information Technology in Bio-and Medical Informatics*, pages 47–61. Springer.

Spilka, J., Leonarduzzi, R., Chudácek, V., Abry, P., and Doret, M. (2016b). Fetal heart rate classification: First vs. second stage of labor. In *Proceedings of the 8th International Workshop on Biosignal Interpretation, Osaka, Japan*, pages 1–3.

Spyridou, K., Chouvarda, I., Hadjileontiadis, L., and Maglaveras, N. (2018). Linear and nonlinear features of fetal heart rate on the assessment of fetal development in the course of pregnancy and the impact of fetal gender. *Physiological measurement*, 39(1):015007.

Steer, P. J. (2008). Has electronic fetal heart rate monitoring made a difference? In *Seminars in Fetal and Neonatal Medicine*, volume 13, pages 2–7. Elsevier.

Stingl, K., Paulsen, H., Weiss, M., Preissl, H., Abele, H., Goelz, R., and Wacker-Gussmann, A. (2013). Development and application of an automated extraction algorithm for fetal magnetocardiography–normal data and arrhythmia detection. *Journal of perinatal medicine*, 41(6):725–734.

Sundar, C., Chitradevi, M., and Geetharamani, G. (2012). Classification of cardiotocogram data using neural network based machine learning technique. *International Journal of Computer Applications*, 47(14).

Svenvik, M., Brudin, L., and Blomberg, M. (2015). Preterm birth: A prominent risk factor for low apgar scores. *BioMed research international*.

Sweha, A., Hacker, T. W., and Nuovo, J. (1999). Interpretation of the electronic fetal heart rate during labor. *American family physician*, 59:2487–2506.

Takatani, T., Takahashi, Y., Yoshida, R., Imai, R., Uchiike, T., Yamazaki, M., Shima, M., Nishikubo, T., Ikada, Y., and Fujimoto, S. (2018). Relationship between frequency spectrum of heart rate variability and autonomic nervous activities during sleep in newborns. *Brain and development*, 40(3):165–171.

Thellesen, L., Hedegaard, M., Bergholt, T., Colov, N. P., Hoegh, S., and Sorensen, J. L. (2015). Curriculum development for a national cardiotocography education program: a delphi survey to obtain consensus on learning objectives. *Acta obstetricia et gynecologica Scandinavica*, 94(8):869–877.

Torres, M. E., Colominas, M. A., Schlotthauer, G., and Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. In *Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on*, pages 4144–4147. IEEE.

Ugwumadu, A. (2014). Are we (mis) guided by current guidelines on intrapartum fetal heart rate monitoring? case for a more physiological approach to interpretation. *BJOG: An International Journal of Obstetrics & Gynaecology*, 121(9):1063–1070.

Ugwumadu, A. (2017). Author's reply re: Time to optimise and enforce training in interpretation of intrapartum cardiotocograph. *BJOG: An International Journal of Obstetrics & Gynaecology*, 124(1):169–170.

Ugwumadu, A., Steer, P., Parer, B., Carbone, B., Vayssiere, C., Maso, G., and Arulkumaran, S. (2016). Time to optimise and enforce training in interpretation of intrapartum cardiotocograph. *BJOG: An International Journal of Obstetrics & Gynaecology*, 123(6):866–869.

Vairavan, S., Ulusar, U., Eswaran, H., Preissl, H., Wilson, J., Mckelvey, S., Lowery, C., and Govindan, R. (2016). A computer-aided approach to detect the fetal behavioral states using multi-sensor magnetocardiographic recordings. *Computers in Biology and Medicine*, 69:44–51.

van der Hout-van, M. B., Jongen, G. J., Bovendeerd, P. H., Oei, S. G., et al. (2013a). Insight into variable fetal heart rate decelerations from a mathematical model. *Early Human Development*, 89(6):361–369.

van der Hout-van, M. B., Oei, S. G., Bovendeerd, P. H., et al. (2012). A mathematical model for simulation of early decelerations in the cardiotocogram during labor. *Medical engineering & physics*, 34(5):579–589.

van der Hout-van, M. B., Oei, S. G., Bovendeerd, P. H., et al. (2013b). Simulation of reflex late decelerations in labor with a mathematical model. *Early human development*, 89(1):7–19.

Van Laar, J., Peters, C., Houterman, S., Wijn, P. F., Kwee, A., and Oei, S. (2011). Normalized spectral power of fetal heart rate variability is associated with fetal scalp blood ph. *Early human development*, 87(4):259–263.

Van Laar, J., Peters, C., Vullings, R., Houterman, S., Bergmans, J., and Oei, S. (2010). Fetal autonomic response to severe acidaemia during labour. *BJOG: An International Journal of Obstetrics & Gynaecology*, 117(4):429–437.

Van Laar, J., Peters, C., Vullings, R., Houterman, S., and Oei, S. (2009). Power spectrum analysis of fetal heart rate variability at near term and post term gestation during active sleep and quiet sleep. *Early human development*, 85(12):795–798.

Van Laar, J., Porath, M., Peters, C., and Oei, S. (2008). Spectral analysis of fetal heart rate variability for fetal surveillance: review of the literature. *Acta obstetricia et gynecologica Scandinavica*, 87(3):300–306.

Victory, R., Penava, D., da Silva, O., Natale, R., and Richardson, B. (2004). Umbilical cord ph and base excess values in relation to adverse outcome events for infants delivering at term. *American journal of obstetrics and gynecology*, 191(6):2021–2028.

Visser, G. H., Ayres-de Campos, D., Barnea, E. R., Bernis, L. D., Carlo, G., Renzo, D., Fernanda, M., Vidarte, E., Lloyd, I., Nassar, A. H., et al. (2018). Figo position paper: how to stop the caesarean section epidemic.

Wakai, R. T. (2004). Assessment of fetal neurodevelopment via fetal magnetocardiography. *Experimental neurology*, 190:65–71.

Warmerdam, G., Vullings, R., Van Laar, J., Bergmans, J., Schmitt, L., Oei, S., et al. (2016a). Using uterine activity to improve fetal heart rate variability analysis for detection of asphyxia during labor. *Physiological measurement*, 37(3):387.

Warmerdam, G. J., Vullings, R., Van Laar, J. O., Bergmans, J., Schmitt, L., Oei, S., et al. (2016b). Selective heart rate variability analysis to account for uterine activity during labor and improve classification of fetal distress. In *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, pages 2950–2953. IEEE.

Warmerdam, G. J., Vullings, R., Van Laar, J. O., Van der Hout-Van, M. B., Bergmans, J. W., Schmitt, L., Oei, S. G., et al. (2018). Detection rate of fetal distress using contraction-dependent fetal heart rate variability analysis. *Physiological measurement*.

Warrick, P. A. and Hamilton, E. F. (2012). Fetal heart-rate variability response to uterine contractions during labour and delivery. In *Computing in Cardiology (CinC), 2012*, pages 417–420. IEEE.

Warrick, P. A. and Hamilton, E. F. (2015). Mutual information estimates of ctg synchronization. In *Computing in Cardiology Conference (CinC), 2015*, pages 137–139. IEEE.

Warrick, P. A., Hamilton, E. F., Precup, D., and Kearney, R. E. (2009). Identification of the dynamic relationship between intrapartum uterine pressure and fetal heart rate for normal and hypoxic fetuses. *IEEE Transactions on Biomedical Engineering*, 56(6):1587–1597.

Warrick, P. A., Hamilton, E. F., Precup, D., and Kearney, R. E. (2010). Classification of normal and hypoxic fetuses from systems modeling of intrapartum cardiotocography. *IEEE Transactions on Biomedical Engineering*, 57(4):771–779.

Wehrwein, E. A. and Joyner, M. J. (2013). Regulation of blood pressure by the arterial baroreflex and autonomic nervous system. In *Handbook of clinical neurology*, volume 117, pages 89–102. Elsevier.

Wei, S.-y., Lu, Y.-S., and Liu, X.-l. (2012). Fetal heart rate analysis using a non-linear baseline and variability estimation method. In *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, pages 532–536. IEEE.

Westerhuis, M. E., Moons, K. G., van Beek, E., Bijvoet, S. M., Drogtrop, A. P., van Geijn, H. P., van Lith, J. M., Mol, B. W., Nijhuis, J. G., Oei, G. S., et al. (2007). A randomised clinical trial on cardiotocography plus fetal blood sampling versus cardiotocography plus st-analysis of the fetal electrocardiogram (stan®) for intrapartum monitoring. *BMC pregnancy and childbirth*, 7(1):1.

Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, volume 460, pages 1597–1611. The Royal Society.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in adaptive data analysis*, 1(01):1–41.

Xu, L., Redman, C. W., Payne, S. J., and Georgieva, A. (2014). Feature selection using genetic algorithms for fetal heart rate analysis. *Physiological measurement*, 35(7):1357.

Yeh, P., Emary, K., and Impey, L. (2012). The relationship between umbilical cord arterial ph and serious adverse neonatal outcome: analysis of 51 519 consecutive validated samples. *BJOG: An International Journal of Obstetrics & Gynaecology*, 119(7):824–831.

Yentes, J. M., Hunt, N., Schmid, K. K., Kaipust, J. P., McGrath, D., and Stergiou, N. (2013). The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of biomedical engineering*, 41(2):349–365.

Yılmaz, E. and Kılıkçıer, Ç. (2013). Determination of fetal state from cardiotocogram using ls-svm with particle swarm optimization and binary decision tree. *Computational and mathematical methods in medicine*, 2013.

Yu, K., Quirk, J. G., and Djurić, P. M. (2017). Dynamic classification of fetal heart rates by hierarchical dirichlet process mixture models. *PloS one*, 12(9):e0185417.

Zarmehri, M. N., Castro, L., Santos, J., Bernardes, J., Costa, A., and Santos, C. C. (2019). On the prediction of foetal acidaemia: A spectral analysis-based approach. *Computers in biology and medicine*, 109:235–241.

Zhang, P., Ye, S., Huang, Z., Jiaerken, D., Zhao, S., Zhang, L., and Wu, J. (2019). A noninvasive continuous fetal heart rate monitoring system for mobile healthcare based on fetal phonocardiography. In *Advances in Body Area Networks I*, pages 191–204. Springer.

Zhao, Z., Zhang, Y., and Deng, Y. (2018). A comprehensive feature analysis of the fetal heart rate signal for the intelligent assessment of fetal state. *Journal of clinical medicine*, 7(8):223.

Zhdanov, D., Bureev, A., Kutsov, M., Kiseleva, E., and Kistenev, Y. (2015). Algorithm for extraction of fetal heart tones during fetal phonocardiography. *Biol Med (Aligarh)*, 7(3):2.

Zourabian, A., Siegel, A. M., Chance, B., Ramanujam, N., Rode, M. E., and Boas, D. A. (2000). Trans-abdominal monitoring of fetal arterial blood oxygenation using pulse oximetry. *Journal of biomedical optics*, 5(4):391–406.

Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht. Insbesondere habe ich nicht die Hilfe eines kommerziellen Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert,
- fremde Forschungsergebnisse verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadensersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, den 16.11.2020

Patricio Fuentealba Ortiz

Anlage 6

Text Wortlaut der Erklärung zur strafrechtlichen Verurteilung

„Ich erkläre hiermit, nicht wegen einer Straftat verurteilt worden zu sein, die Wissenschaftsbezug hat.“

Magdeburg, den 16.11.2020

Patricio Fuentealba Ortiz