

Review

Computerised Cardiotocography Analysis for the Automated Detection of Fetal Compromise during Labour: A Review

Lochana Mendis ^{1,*}, Marimuthu Palaniswami ¹, Fiona Brownfoot ^{2,†} and Emerson Keenan ^{1,2,†}

- ¹ Department of Electrical and Electronic Engineering, The University of Melbourne, Parkville, VIC 3010, Australia; palani@unimelb.edu.au (M.P.); e.keenan@ieee.org (E.K.)
- ² Obstetric Diagnostics and Therapeutics Group, Department of Obstetrics and Gynaecology, The University of Melbourne, Heidelberg, VIC 3084, Australia; fiona.brownfoot@unimelb.edu.au
- * Correspondence: lochana.mendis@ieee.org
- † These authors contributed equally to this work.

Abstract: The measurement and analysis of fetal heart rate (FHR) and uterine contraction (UC) patterns, known as cardiotocography (CTG), is a key technology for detecting fetal compromise during labour. This technology is commonly used by clinicians to make decisions on the mode of delivery to minimise adverse outcomes. A range of computerised CTG analysis techniques have been proposed to overcome the limitations of manual clinician interpretation. While these automated techniques can potentially improve patient outcomes, their adoption into clinical practice remains limited. This review provides an overview of current FHR and UC monitoring technologies, public and private CTG datasets, pre-processing steps, and classification algorithms used in automated approaches for fetal compromise detection. It aims to highlight challenges inhibiting the translation of automated CTG analysis methods from research to clinical application and provide recommendations to overcome them.



Citation: Mendis, L.; Palaniswami, M.; Brownfoot, F.; Keenan, E. Computerised Cardiotocography Analysis for the Automated Detection of Fetal Compromise during Labour: A Review. *Bioengineering* **2023**, *10*, 1007. <https://doi.org/10.3390/bioengineering10091007>

Academic Editors: Aimée Lovers, Giulia Baldazzi, Nicolò Pini, Danilo Pani, Antoniya Georgieva, Patrice Abry and Martin Gerbert Frasch

Received: 12 July 2023

Revised: 17 August 2023

Accepted: 22 August 2023

Published: 25 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: fetal compromise; intrapartum fetal monitoring; cardiotocography; fetal heart rate; artificial intelligence

1. Introduction

Despite ongoing advances in maternal and fetal healthcare, over 2 million babies are stillborn worldwide every year [1–3]. Depending on global region, between 6% and 49% of these stillbirths occur during labour compared to the antepartum period [1]. Many of these stillbirths could be prevented with monitoring during pregnancy [4] and timely obstetric intervention for complications during childbirth [5]. The most common cause of stillbirth globally is intrapartum asphyxia [6], which typically occurs when mechanical pressure during labour limits blood flow to the baby, resulting in a reduction in oxygen delivery to the fetus. When oxygen levels are critically low, the fetus changes its metabolism from aerobic to anaerobic, resulting in hydrogen ion formation in tissues, known as acidosis [7]. In addition to total oxygen deprivation, which may result in stillbirth, this build-up of acid is extremely damaging, especially to neuronal cells in the fetal brain [8]. This can result in short- and long-term brain injury, leading to complications, including cerebral palsy and intellectual impairment for surviving infants [9]. For this reason, the early detection of fetal compromise is extremely important so that clinicians can intervene before irreversible harm occurs.

Due to these hypoxic events triggering the fetal sympathetic and parasympathetic nervous systems, variations in the fetal heart rate (FHR) can be monitored to detect signs of fetal compromise [8,10,11]. The primary source of FHR surveillance in the early half of the twentieth century was via the use of intermittent auscultation, and in the late 1960s, with the development of ultrasound technology, continuous cardiotocography (CTG) was

introduced for fetal monitoring [12]. CTG extracts both the FHR and uterine contractions (UC) simultaneously, allowing for the continuous assessment of both signals.

FHR and UC are typically extracted using either invasive or non-invasive methods. The most widely used method is the non-invasive Doppler ultrasound and tocodynamometer, which uses two external transducers placed on the mother's abdomen. These signals, however, are subject to signal dropout due to fetal and maternal movements, and their signal quality deteriorates with an increase in maternal body mass index (BMI) [13]. Therefore, more accurate readings are sometimes obtained using invasive monitoring methods such as the direct fetal electrocardiogram (fECG) acquired via a fetal scalp electrode for FHR and intrauterine pressure catheter for UC [14,15]. Despite their benefits, these invasive modes of measurement have an increased risk of infection and can only be used by rupturing the membranes, making them suitable for use during labour only [14,16]. As a result, new methods such as the non-invasive fetal electrocardiogram and electrohysterogram are being introduced to overcome the shortcomings of their predecessors [17]. Throughout this paper, we use the term CTG to refer to any method that measures both FHR and UC simultaneously.

Since its introduction, CTG has become a routinely used technique in clinical practice for fetal monitoring in labour. However, despite the wide use of CTG, it has not been able to show a significant reduction in fetal mortality [18–20]. In practice, CTG recordings are visually evaluated and assessed based on guidelines, leading to substantial inter- and intra-observer disagreement among clinicians [21–23]. Furthermore, the high false positive rate of visual CTG evaluation is cited as one of the reasons behind the increasing rate of caesarean sections (CS) and unnecessary operative deliveries [24]. Therefore, there is an unmet clinical need for a more reliable, reproducible and objective CTG evaluation system.

As CTG evaluation provides an indirect measure of fetal asphyxia, a more direct and quantitative measure of fetal asphyxia is pH. To evaluate pH, a blood sample is taken from either the fetal scalp during labour or the umbilical cord immediately after birth [25]. The pH value measures the level of fetal acidemia arising from the dispersion of accumulated lactic acid in the fetal circulation [26]. The base deficit (BDecf) is another parameter determined from the umbilical cord blood, which is closely linked to fetal asphyxia. pH values below 7.05 and BDecf values above 10 mmol/L are strongly related to adverse neonatal outcomes [26,27]. Another metric used to assess fetal compromise is the Apgar score, which evaluates the physical condition of the baby shortly after delivery, where the 5-min Apgar score has been shown to be a predictor of the risk of neonatal death [28]. Therefore, umbilical artery pH, BDecf and Apgar are common clinical endpoints used to define fetal compromise.

With the availability of digital CTG records and clinical endpoints defining the presence of fetal asphyxia, alternative computerised data-driven approaches have been proposed to reduce misinterpretation and improve the accuracy of CTG evaluation. Initially, these methods tried to approximate expert interpretations based on guidelines provided by the International Federation of Gynaecology and Obstetrics (FIGO) by analysing baseline variability, accelerations and decelerations of the FHR [29]. The Dawes–Redman System and the Porto System are two systems that first adopted these features for automated CTG analysis in clinical practice [30,31], followed by the INFANT system [32,33]. Thereafter, additional FHR features from time, frequency and non-linear domains were introduced in addition to the morphological features. A CTG analysis prototype system called the OxSys [34] based on one of the non-linear FHR features, decelerative capacity, was recently introduced to analyse and trigger real-time alerts to clinicians.

Concurrent to these clinical methods, FHR features were also utilised in data-driven approaches to train machine learning (ML) algorithms to detect fetal compromise [35,36]. Subsequently, modern data-driven methods like deep learning (DL) algorithms were proposed, which independently learn information from the raw CTG waveforms minimizing the need for human-guided feature extraction and selection [37]. However, comparisons between the performance of these novel data-driven methods are difficult as different

outcome measures and different case selection criteria have been used across studies. The most commonly used biomarker in differentiating CTGs with fetal compromise in these computer-based methods is umbilical artery pH with a threshold set at 7.05 [35,37–40]. Other alternative cut-off pH values such as 7.10 [41,42], 7.15 [43–46] and 7.20 [47] have also been employed by previous studies. Nevertheless, these algorithms have shown the potential to improve the sensitivity of detecting fetal compromise but have not yet been translated into widespread clinical use.

This narrative review aims to synthesise existing literature on computerised cardiotocography analysis for fetal compromise detection, with a particular focus on automated data-driven techniques and a comprehensive overview of their associated challenges. This complements recent reviews associated with the physiology of FHR variability during labour [48], signal processing techniques for FHR analysis [49–51], challenges for developing AI techniques for FHR monitoring [52,53], and computerised CTG analysis in clinical practice [54]. One of the primary focuses of this paper is to identify a standardised approach for CTG segment length, pre-processing, and definition of outcomes for fetal compromise to enable the effective comparison of data-driven analysis methods.

To achieve this, we first provide a brief overview of the physiological basis for CTG monitoring and then summarise FHR and UC monitoring technologies and associated challenges with data extraction. Then, we provide an overview of the current state of the existing computerised CTG systems in clinical practice. Next, we provide a summary of public and private CTG datasets, followed by a review of pre-processing techniques reported in the literature. Following this, we comprehensively review the data-driven computerised approaches for classifying fetal compromise, from proposed methods based on feature-based machine learning to the latest advancements in deep learning. Finally, the current obstacles that limit progress within this field are identified, and recommendations for future research are presented. A summary of the process for the computerised CTG analysis of fetal compromise detection and the core topics addressed in this review are shown in Figure 1.

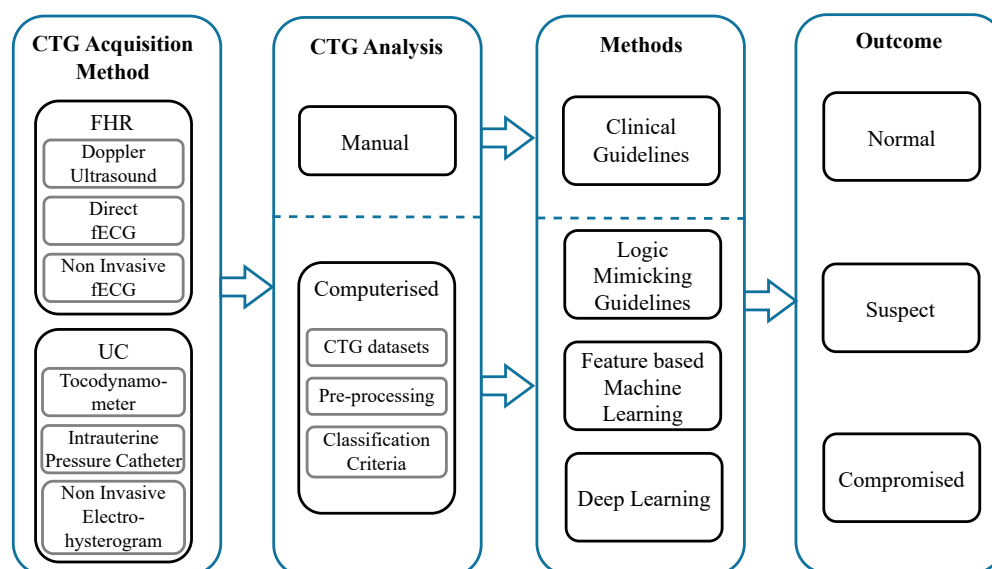


Figure 1. Summary of the computerised CTG analysis process for fetal compromise detection. The process begins with the CTG acquisition methods for fetal heart rate and uterine contractions. The output of these signals is then analysed by two methods: visual inspection or a computerised approach. The prior is based on clinical guidelines, and the latter uses artificial intelligence to analyse the FHR. The outcome of the analysis is to predict the status of the baby as normal, suspect or compromised.

2. Physiological Basis for CTG Monitoring during Labour

The fetal heart rate is regulated by the autonomic nervous system, primarily through sympathetic and parasympathetic influences [8,10,55]. During labour, the FHR is influenced by various factors such as changes in oxygenation as a result of insufficient placental perfusion, fetal movement and umbilical cord compression. These factors give rise to distinct FHR patterns, which are typically categorised as changes in baseline, variability, accelerations and decelerations. The baseline is the average FHR, excluding accelerations and decelerations. The variability is the fluctuation of the FHR around the baseline, reflecting the balance between the sympathetic and parasympathetic activity. Accelerations are transient increases in FHR above the baseline, indicating a healthy fetal response to stimuli. Decelerations are transient decreases in FHR below the baseline, which can be either physiological or pathological depending on their shape, timing, duration, and association with uterine contractions [11,56].

Uterine contractions also play a central role during labour, resulting from the coordinated activity of uterine smooth muscles. Oxytocin stimulates contractions' frequency and intensity, while prostaglandins contribute to cervical ripening [57]. These contractions can impact FHR through changes in uteroplacental perfusion or umbilical cord compression. Intense or frequent UCs may compromise fetal oxygenation by limiting blood flow, potentially leading to FHR decelerations [55]. Alternatively, UC-induced stress can stimulate fetal sympathetic activity, resulting in FHR accelerations [26,57]. Therefore, the simultaneous monitoring of UC and FHR patterns using CTG can aid clinicians in assessing both labour progression and fetal status.

3. Fetal Heart Rate and Uterine Contraction Monitoring Technologies

Although the FHR and UC signals produced across CTG monitoring systems are displayed in a consistent format, they may be acquired using different modalities that have different accuracy profiles, benefits and limitations. These signals are critical for computerised CTG analysis, and problems with input signal quality can lead to incorrect predictions in subsequent steps. To highlight the importance of this area, this section provides a review of various technologies utilised to extract FHR and UC signals. An overview of fetal monitoring technologies for capturing FHR and UC is shown in Figure 2.

3.1. Fetal Heart Rate Monitoring Technologies

3.1.1. Doppler Ultrasound

Doppler ultrasound is a non-invasive form of monitoring the fetal heart rate. It comprises an ultrasound transducer strapped to the maternal abdomen, which emits an ultrasound wave and measures the reflected waveforms resonated with the fetal cardiac structures. Fetal heartbeats are then approximated using signal modulation and autocorrelation techniques [58]. Due to the nature of ultrasound monitoring, the extracted FHR recordings are prone to signal losses caused by the movements of the baby, maternal movements and transducer displacement. Signal artefacts also occur particularly due to accidental monitoring of the maternal heart rate [11] and half or double counting fetal heart rate values when using autocorrelation techniques for FHR approximation [59]. These noises are typically seen more often during the second stage of labour [14]. The quality of the FHR measured using the Doppler ultrasound method is also hampered by maternal factors like increased BMI [60]. In terms of reliability, the mean positive percent agreement of the Doppler ultrasound compared to fetal scalp electrode during labour is between 62 and 73% [61,62]. Despite these limitations, Doppler ultrasound has been widely used in monitoring FHR during labour since its introduction in the 1960s. According to a United States (U.S.) national survey, 98% of mothers who gave birth in a U.S. hospital from 2011–2012 underwent Doppler ultrasound CTG monitoring at some point during labour [63].

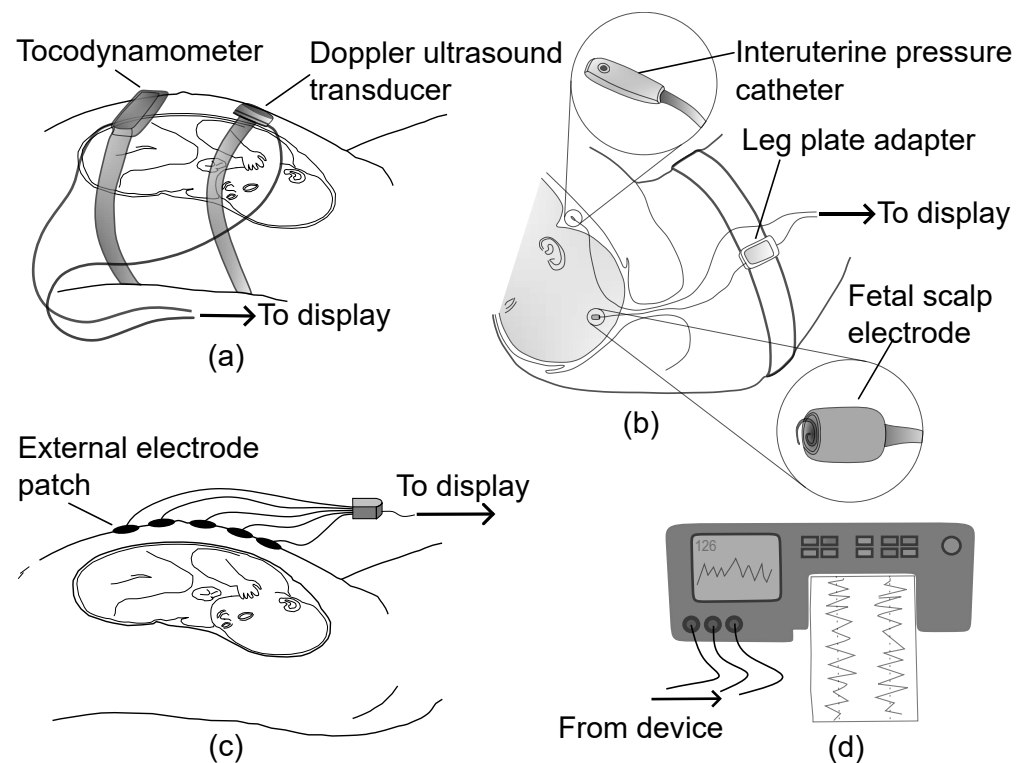


Figure 2. An overview of fetal monitoring technologies for capturing FHR and UC showing (a) Doppler ultrasound and tocodynamometer, (b) direct fetal electrocardiogram using scalp electrode and intrauterine pressure catheter, (c) non-invasive fetal electrocardiogram and electrohysterogram using external electrodes, and (d) the display unit for monitoring the extracted data in digital or paper format.

3.1.2. Direct Fetal Electrocardiogram

The direct fetal electrocardiogram is an invasive method where a fetal scalp electrode is inserted through the mother's cervix and attached to the presenting part of the fetal scalp. It enables the measurement of the time difference between two consecutive heartbeats by identifying the R waves of the direct fECG [11,64], allowing for a more accurate beat-to-beat FHR to be obtained. However, the electrode can only be placed when the amniotic sac around the baby has ruptured and the cervix is dilated [64]. This limits its usage to monitoring FHR only during labour. Further, it is typically avoided in fetuses under the gestational age of 32 weeks and in mothers who are infected by blood-borne viruses [11]. Additionally, by using this method in clinical practice there is an increased risk of infections, injury, and bruising of the fetal scalp. This method is also more expensive to use as the electrodes are disposable [11]. However, the direct fECG monitoring of FHR through a scalp electrode is the most accurate method for fetal surveillance during labour [14,65].

3.1.3. Non-Invasive Fetal Electrocardiogram

A desire to overcome the limitations of Doppler ultrasound and the invasiveness of the direct fECG has led to the introduction of an alternate method of monitoring called the non-invasive fetal electrocardiogram (NI-fECG). This method uses several electrodes placed on the maternal abdomen to record the electrical activity of the fetal heart which is used to extract the FHR rhythm. The main challenge faced by this method is the low signal-to-noise ratio of the fetal ECG compared to that of maternal ECG and the presence of different dielectric media between the sensors and the electrodes [66–69]. However, the NI-fECG method provides more reliable FHR readings compared to Doppler ultrasound in labour with a mean positive percent agreement between 81–83% than the fetal scalp electrode [61,62], with no significant decrease in performance for high BMI patients [13]. However, there are limitations to these studies, where agreement analysis

was only performed in a subset of recording segments where fetal heart rate extraction was successful. The percentage of successful recording segments varied between Doppler and NI-fECG methods, as well as there were variations in the number of participants included for analysis during the second stage of labour. As such, further studies are required to demonstrate that NI-FECG offers comparable performance to direct fetal ECG scalp monitoring.

3.2. Uterine Contraction Technologies

3.2.1. External Tocodynamometer

An external tocodynamometer (TOCO) is a non-invasive strain gauge placed over the maternal abdomen to monitor uterine activity. This device provides information on the frequency and approximate duration of uterine contractions but not their absolute intensity [16,70]. This method can be used both before and during labour but suffers from periods of signal loss arising from misalignment following maternal movements. Furthermore, performance reduces with increasing BMI [71]. Nevertheless, TOCO is the most commonly used modality to monitor uterine contractions as it is non-invasive and simple to use [70].

3.2.2. Intrauterine Pressure Catheter

An intrauterine pressure catheter (IUPC) is an invasive device placed inside the amniotic space during labour to monitor uterine contractions. This method is unaffected by maternal position and obesity and provides the most accurate measurement of the frequency of contractions as well as their strength and duration [15]. Therefore, it is considered the current gold standard for measuring UC. However, it requires the rupture of the membranes and carries the risk of infection and other complications, which limit its widespread use [72].

3.2.3. Non-Invasive Electrohysterogram

The electrohysterogram is a promising non-invasive technique that uses abdominal electrodes to measure uterine electrical activity and derive uterine contractions [73]. The main benefits of this method are that it gives a reliable measurement of the contractions even with obese women compared to TOCO [16,71] and unlike IUPC can be used for continuous UC monitoring in both pregnancy and during labour.

4. Computerised CTG Analysis Systems in Clinical Practice

The first computerised CTG analysis system was introduced by Dawes and Redman [30] by analysing the morphology of the FHR to alert clinicians of the risk of pathological outcomes during the antenatal period. In the same year, the Porto system was introduced, which quantitatively adapts the FIGO guidelines for automated CTG analysis during labour [31]. Later, this system was commercialised as SisPorto [74] and Omniview-SisPorto (Speculum, Lisbon, Portugal) [75]. The latest version of this system, SisPorto 4.0 [76], incorporates the FIGO's 2015 guidelines and provides real-time alerts to features that require attention or clinical intervention. INFANT (K2 Medical Systems, Plymouth, UK) is another decision support system that analyses the FHR and UC patterns to provide alerts to help clinicians during labour. Recently, a new prototype system primarily based on the decelerative capacity of FHR called the OxSys was introduced to analyse and trigger real-time alerts to clinicians [34].

Most of these systems have undergone years of work, rigorous testing, and clinical validation through randomised control trials (RCTs). The largest RCT on 46,042 women evaluated the INFANT system and concluded that the incidence of poor neonatal outcomes was the same regardless of using the decision support system, 0.7% in both groups (adjusted risk ratio 1.01, 95% CI:0.82–1.25) [33]. Another RCT on 7730 patients evaluated continuous central fetal monitoring by computer analysis and real-time alerts of the Omniview-SisPorto 3.5 system with visual analysis and concluded that while both study arms reported lower

than expected rate of newborn metabolic acidosis, no significant reduction in the metabolic rate or obstetric intervention was achieved using computerised analysis (relative risk 0.69, 95% CI:0.36–1.31) [77]. Two recent meta-analyses also concluded similar findings that computerised CTG analysis did not improve neonatal outcomes compared to conventional evaluation [78,79]. In another study, a retrospective database of 22,790 women in labour was evaluated using the prototype OxSys 1.5 system. They found that the OxSys system potentially increased the sensitivity of fetal compromise detection (43.3% vs. 38.0% for severe, $p = 0.3$ and 36.1% vs. 31.0% for moderate, $p = 0.06$) while reducing the false positive rate (14.4% vs. 16.3%, $p < 0.001$) compared to conventional clinical diagnosis [34]; however, the sensitivity increase was not statistically significant.

5. CTG Datasets

In order to develop new data-driven methods for CTG analysis during labour, it can be seen that large datasets are necessary to capture uncommon clinical outcomes. CTG recordings are typically stored by hospitals in paper format or as digital records in electronic health systems. Research groups with access to digitally accessible records have proposed a range of computer-based CTG analyses with varying dataset sizes, data sampling frequencies, and outcome measures. Nevertheless, there are only two CTG datasets currently available for public access: (1) The University of California, Irvine (UCI) Machine Learning repository CTG dataset, consisting of 2126 records, each with 23 features of FHR and UC signals classified into three classes of normal, suspect and pathological by three expert obstetricians [80]; and (2) The Czech Technical University and University Hospital in Brno (CTU-UHB) dataset, consisting of 552 CTG records with raw FHR and UC signals [81]. It is important to note that the UCI dataset does not contain the raw FHR and UC signals. Additionally, the Lyon dataset [38] and the Oxford dataset are used in the literature [37], but these are not publically available. An overview of these four frequently used datasets is shown in Table 1.

Table 1. Overview of public and private CTG datasets.

Dataset	Signal Format	Outcome Data	Number of Recordings	Availability
UCI ML Repository [80]	FHR and UC features only	Three classes: Normal (N) Suspect (S) Pathological (P)	2126 N = 1655 S = 295 P = 176	Public
CTU-UHB Dataset [81]	Raw FHR and UC at 4 Hz	pH	552	Public
Lyon Dataset [38]	Raw FHR and UC at 10 Hz	pH	1288	Private
Oxford Dataset [37]	Raw FHR and UC at 4 Hz	pH	35,429	Private

6. FHR Pre-Processing Techniques

In routine practice, the FHR is monitored through Doppler ultrasound CTG or direct fetal scalp electrode, or sometimes a combination of both. For example, FHR records of the CTU-UHB dataset were acquired by a mixture of Doppler ultrasound CTG and direct fECG [81]. These signals are contaminated by noises composed of artefacts and periods of missing FHR values (denoted by zeros) as shown in Figure 3. Fetal/maternal movements and the displacement of the transducers add noise to Doppler ultrasound CTG monitoring, while vaginal examinations and maternal pushing contribute to noise in CTG monitoring by direct fECG [14]. Generally, for Doppler ultrasound-based CTG, median signal loss is reported to be between 5 and 8% for the first stage of labour and between 9 and 20% for the second stage of labour. The direct fetal scalp electrode demonstrates lower median

signal loss between 0.8 and 1% and 3 and 4% in the first and second stages of labour, respectively [14,82,83]. For the above reasons, raw FHR data require pre-processing before analyzing with computer-based methods like machine learning because low-quality data prevent stability and convergence in the learning process [84]. Hence, pre-processing is a crucial step in the automatic evaluation of FHR signals and generally involves the following five steps, which may be performed in differing order:

1. Segment Selection

In clinical practice, labour is divided into three stages: Stage I is where the cervix starts dilating (<10 cm) and frequent contractions occur; Stage II is the period from when the cervix is fully dilated (10 cm) to when the baby is born; Stage III starts after the baby is born and continues until delivery of placenta and membranes. In the segment selection step, a signal segment with sufficient quality that is closer to the delivery of the baby is typically selected as this is representative of the level of asphyxia and correlates best with the cord pH at birth. Unfortunately, the FHR signal is typically most affected by noise and artefacts at this point of labour [85]. FIGO guidelines require the signal loss to be less than 20% for a signal to be acceptable for evaluation [14]. Different prior studies have selected signal lengths varying from 10–60 min before birth to analyse, whilst others simply select the segment of trace during a certain stage of the labour [44].

2. Artefact Removal

A typical baseline heart rate of a normal fetus varies between 110 bpm and 160 bpm, and accelerations or decelerations occur when amplitude varies 15 bpm above or below the baseline lasting for more than 15 s respectively [11]. In the artefact removal step, values below 50 bpm and above 200 bpm are typically considered outliers and removed [36]. In some studies, consecutive missing values of more than 15 s (long gaps) are removed from the analysis [36,40,86]. For others, when the difference between two adjacent FHR values exceeds 25 bpm, the corresponding signal segment from the previous FHR value to the next stable segment is considered unstable and removed [45,85]. A stable segment is a signal segment with five consecutive FHR values having a difference of less than 10 bpm between them [31].

3. Signal Interpolation

The signal interpolation step employs techniques like linear [87] and spline [35] interpolation to fill the missing FHR values created from the previous steps. Generally, the interpolation is performed for gaps < 15 s, and the gaps > 15 s in the FHR are either skipped or removed in subsequent feature extraction and deep learning training processes [35,39,86]. In linear interpolation, these missing gaps are approximated using the slope of the data points on either side of the gap. Spline interpolation uses a set of low-degree polynomials called a spline to estimate the missing gaps to make the signal smoother and continuous. When polynomials of degree 3 are used in the spline, the resulting interpolation is called cubic spline interpolation. The Hermite spline interpolation uses polynomials defined by the values and the derivatives at the endpoints of the corresponding interval to estimate the missing values.

4. Downsampling

A typical fetal heart beats less than 3 times per second (<180 bpm), making some data of the original FHR signals sampled at 4 Hz redundant [46,87]. Therefore, in this step, the FHR signal is sometimes downsampled to reduce the computational complexity and memory needed to process the input signals. For example, only 900 values are required to represent a 60 min FHR signal at 0.25 Hz, compared to the 14,400 values required for the same signal at 4 Hz.

5. Smoothing and Detrending

In some works, the final step in pre-processing was to smooth the FHR signal using a median filter [44,88] or detrend the signal before using non-linear signal processing techniques [44,89].

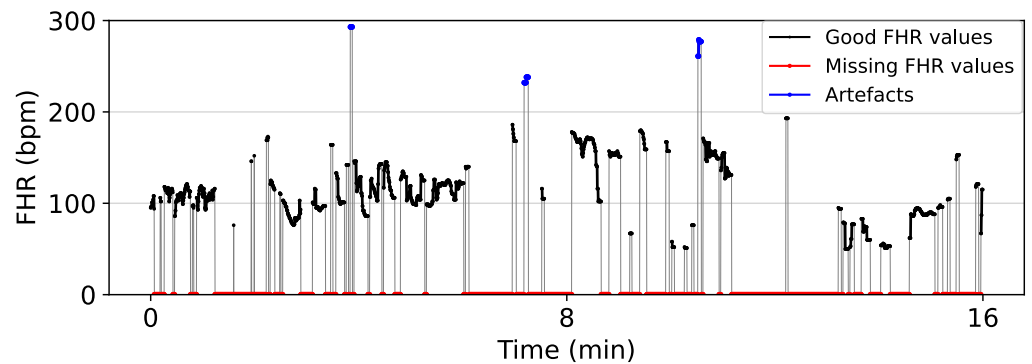


Figure 3. Segment of CTG 1087 from CTU-UHB showing the artefacts and missing values of a raw FHR signal. FHR values in the range 50–200 bpm are considered good values, while FHR values below 50 bpm and above 200 bpm are considered outliers (artefacts). Zero values represent the missing FHR values.

In addition to these conventional and straightforward steps, more sophisticated methods also have been proposed for FHR noise removal and the recovery of missing samples [90,91]. An example of typical pre-processing steps applied to an FHR signal is shown in Figure 4, as well as a summary of different pre-processing steps used by prior studies given in Table 2, demonstrating large differences in the pre-processing implementation across studies.

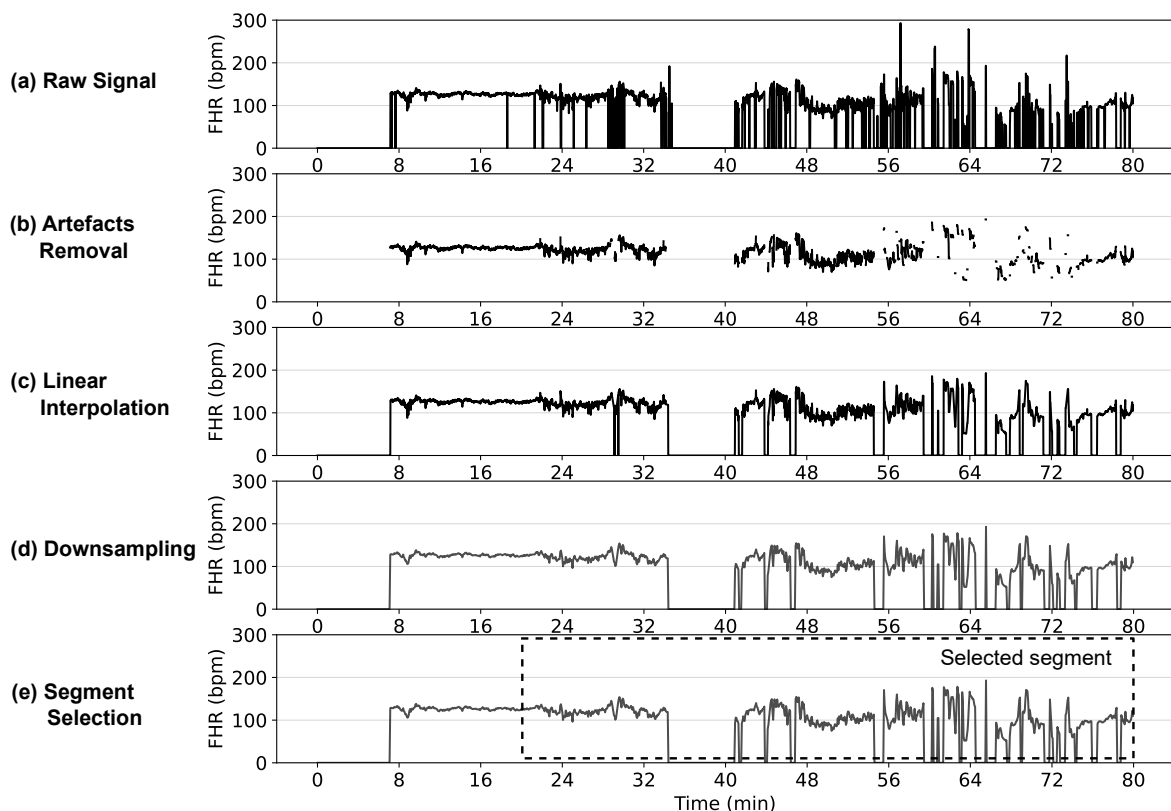


Figure 4. Typical pre-processing steps on a FHR signal showing: (a) raw FHR signal at original 4 Hz sampling rate; (b) artefact removal of FHR values outside 50–200 bpm and non-physiological values with variation among adjacent FHR values > 25 bpm; (c) linear interpolation to fill the missing gaps—only gaps < 15 s are interpolated, the gaps > 15 s are shown as zeros, which will be either skipped or removed in subsequent feature extraction or classification stages; (d) downsampled signal at 0.25 Hz; and (e) last 60 min segment selected for analysis.

Table 2. Overview of different FHR pre-processing techniques used in terms of segment selection, artefact removal, signal interpolation, downsampling and signal detrending.

Year	Authors	Segment Selection	Artefact Removal	Signal Interpolation	Downsampling Frequency	Smoothing and Detrending
2013, 2014	Spilka et al. [35,36]	Not specified	FHR < 50 bpm or FHR > 210 bpm corrected Long gaps > 15 s not included	Hermite spline	None	None
2016	Stylios et al. [39]	Last 30 min of 1st stage of labour	Not specified	Hermite spline	None	None
2017	Georgoulas et al. [40]	Not specified	FHR < 50 bpm or FHR > 200 bpm corrected Long gaps > 15 s not included	Hermite spline	None	None
2018	Cömert and Kocamaz [47]	Last 15 min of 2nd stage of labour	FHR < 50 bpm or FHR > 200 bpm corrected Long gaps > 15 s not included	Cubic Hermite spline	None	None
2018	Petrozziello et al. [87]	Last 60 min of labour	Not specified	Linear	None	None
2018	Zhao et al. [43]	Not specified	Adjacent 5 FHR values with variability < 10 bpm and signal with zeros for >10 s not included FHR ≤ 50 bpm or FHR ≥ 200 bpm corrected FHR values with differences in adjacent values exceeding 25 bpm corrected	Linear and spline	None	None
2018	Cömert et al. [44]	Last 30 min of 1st stage of labour	Outliers and artefacts corrected	Cubic spline	None	Detrending: 2nd order polynomial
2018	Cömert et al. [88]	Last part of the 1st stage of labour	Long gaps > 15 s not included Outliers corrected	Cubic Hermite spline	None	Smoothing: Median filter Detrending: Not specified
2019	Fuentealba et al. [92]	Last 25 min of labour	FHR < 50 bpm or FHR > 200 bpm corrected	Hermite spline	None	None
2019	Petrozziello et al. [37]	Last 60 min regardless of labour stage	Not specified	Linear	0.25 Hz	None
2019	Zhao et al. [45,86]	Not specified	FHR < 50 bpm or FHR > 200 bpm corrected Long gaps >15 s not included FHR values with differences in adjacent values exceeding 25 bpm are corrected	Linear and cubic spline	None	None
2021	Liang and Li [89]	Last 30 min of labour	FHR< 50 bpm or FHR > 200 bpm corrected Long gaps >15 s not included	Hermite spline	1 Hz	Smoothing: Median filter
2021	Liu et al. [46]	Last 20 min of FHR	FHR < 50 bpm or FHR > 200 bpm corrected Long gaps > 15 s not included FHR values with differences in adjacent values exceeding 25 bpm are corrected	Linear and Hermite spline	1 Hz	None
2021	O'Sullivan et al. [93]	Not specified	FHR < 50 bpm or FHR > 210 bpm are corrected Patients with more than 30% traces missing are removed from the study FHR values with a change greater than 30% from the moving average are corrected	Not specified	None	None

7. UC Pre-Processing Techniques

Uterine contractions are commonly linked with FHR decelerations [94] and may provide vital information about the fetus during labour. This requires capturing the intensity, duration and shape of uterine contractions to make an assessment. However, the UC signals obtained from CTG often exhibit poor quality due to technical constraints due to the majority of signals being acquired using TOCO [87]. Clinical practice typically takes this into account by only utilising UC signals when they are reliable [37]. Therefore, for computerised CTG analysis methods to avoid utilising poor quality information, the majority of studies exclude UC signals from their approaches [95]. In limited studies such as [37,87], the raw UC signals are pre-processed by assessing the quality using an

established autoregressive model [96]. Then, UC signals of poor quality are replaced by zeros before using them in the computerised CTG analysis. Therefore, pre-processing approaches for UC signals in the literature involve one of two methods: (1) the complete exclusion of UC signals, where they are not considered or analyzed in the study or (2) the replacement of poor-quality UC values with zeros, effectively removing or discounting unreliable or noisy data points from the analysis.

8. Fetal Compromise Classification Criteria

Before training computerised CTG analysis approaches, it is necessary to assign labels to the CTG data to represent the status of the fetus at birth. Different studies have used a variety of classification criteria. The predominant objective indicator of fetal compromise that has been used in the literature is the umbilical artery pH. The pH value indicates the degree of fetal acidemia resulting from the dispersion of accumulated lactic acid in the fetal circulation [26]. The fetal pH is typically measured from a sample, which may be acquired from fetal scalp, umbilical arterial or umbilical venous blood. As fetal carbon dioxide is removed through the umbilical arterial blood, it typically has a slightly higher pH reading than venous blood. When arterial pH is used to define the class labels, it is important that we only use the pH readings that are validated to prevent errors in the labels. An arterial pH value that is measured is considered valid if it is at least 0.02 more than the venous pH [25]. BDecf and Apgar scores are two other predictors of fetal compromise used in the existing literature. Typically, pH values below 7.05 and BDecf values above 10 mmol/L are strongly correlated with adverse neonatal outcomes [26,27], whereas a 5-min Apgar score of less than or equal to 6 indicates an increased risk of neonatal death [28]. Consequently, umbilical artery pH, BDecf and Apgar scores are widely employed as common clinical endpoints to define fetal compromise. In addition, Petroziello et al. [37] have used other fetal outcomes such as stillbirth, neonatal death, neonatal encephalopathy, neonatal intubation, and admission to neonatal intensive care for 48 h or more to define fetal compromise. Table 3 shows different classification criteria and thresholds used in the literature.

Table 3. Overview of different classification criteria used to define classes and their corresponding class distribution.

Year	Authors	Datasets	Classification Criteria (Denoting All Unhealthy Classes as Compromised and Healthy Classes as Normal)	Number of Recordings
2013	Spilka et al. [35]	CTU-UHB	pH $\leq$ 7.05 as Compromised	Total = 552 Normal = 508 Compromised = 44
2013	Georgieva et al. [97]	Subset of Oxford Dataset	Training: 7.27 < pH < 7.33 as Normal pH < 7.1 as Compromised Testing: 7.22 < pH < 7.27 as Normal pH < 7.1 as Compromised	Training set: Total = 124 Normal = 62 Compromised = 62 Testing set: Total = 252 Normal = 126 Compromised = 126
2016	Spilka et al. [38]	Lyon Dataset CTU-UHB	pH $\leq$ 7.05 as Compromised	Training: Lyon DB Total = 1288 Normal = 1251 Compromised = 37 Testing: CTU-UHB Total = 420 Normal = 400 Compromised = 20
2016	Stylios et al. [39]	CTU-UHB	pH $\leq$ 7.05 as Compromised	Total = 552 Normal = 508 Compromised = 44

Table 3. Cont.

Year	Authors	Datasets	Classification Criteria (Denoting All Unhealthy Classes as Compromised and Healthy Classes as Normal)	Number of Recordings
2017	Georgoulas et al. [40]	CTU-UHB	pH $\leq$ 7.05 as Compromised	Total = 552 Normal = 508 Compromised = 44
2017	Spilka et al. [98]	Lyon Dataset	pH $\leq$ 7.05 as Compromised	Total = 1288 Normal = 1251 Compromised = 37
2018	Cömert and Kocamaz [47]	CTU-UHB	pH $<$ 7.2 as Compromised	Total = 552 Normal = 375 Compromised = 177
2018	Feng et al. [41]	CTU-UHB	pH $>$ 7.2 as Normal pH $<$ 7.1 as Compromised	Total = 447 Normal = 358 Compromised = 62
2018	Petrozziello et al. [87]	Oxford Dataset	pH $<$ 7.05 as Compromised	Total = 35,429 Normal = 33,959 Compromised = 1470
2018	Zhao et al. [43]	CTU-UHB	pH $<$ 7.15 as Compromised	Total = 552 Normal = 447 Compromised = 105
2018	Cömert et al. [44]	CTU-UHB	pH $\leq$ 7.15 as Compromised	Total = 552 Normal = 439 Compromised = 113
2018	Cömert et al. [88]	CTU-UHB	pH $\leq$ 7.15 as Compromised	Total = 552 Normal = 439 Compromised = 113
2019	Fuentealba et al. [92]	CTU-UHB	pH $>$ 7.2 and BDecf $<$ 12 as Normal pH $<$ 7.05 and BDecf $\geq$ 12 as Compromised	Total = 372 Normal = 354 Compromised = 18
2019	Petrozziello et al. [37]	Oxford Dataset CTU-UHB Lyon Dataset	Normal: pH $\geq$ 7.15 Severe Compromise: pH $<$ 7.05 and a composite outcome of stillbirth; neonatal death; neonatal encephalopathy; intubation or cardiac massage followed by admission to neonatal intensive care for $\geq$ 48 h Moderate Compromise: pH $<$ 7.05 Intermediate: 7.05 $\leq$ pH $<$ 7.15	Oxford Dataset Training: 30,115 Testing: 4429 Normal = 4249 Moderate/Severe compromise with pH $<$ 7.05 = 180 Testing: CTU-UHB Total = 552 Normal = 512 Compromised = 40
2019	Zhao et al. [86]	CTU-UHB	pH $\geq$ 7.15 as Normal pH $<$ 7.15 as Compromised	Normal = 447, Compromised = 105 After CWT: Normal = 2682 Compromised = 630
2019	Zhao et al. [45]	CTU-UHB	pH $<$ 7.15 as Compromised	Normal = 105, Compromised = 105 After RP 2D: Normal = 21,000 Compromised = 21,000
2020	Furuya et al. [99]	Private	5-min Apgar score $<$ 8 or pH $<$ 7.1 as Compromised	Total = 1301 Normal = 1184 Compromised = 117
2021	Liang and Li [89]	CTU-UHB	pH $\leq$ 7.05 as Compromised	Total = 552 Normal = 508 Compromised = 44
2021	Liu et al. [46]	CTU-UHB	pH $\leq$ 7.15 as Compromised	Total = 552 Normal = 439 Compromised = 113
2021	O'Sullivan et al. [93]	CTU-UHB	pH $\geq$ 7.15 and Apgar 5 $\geq$ 9 as Normal pH $\leq$ 7.0 or Apgar 5 $\leq$ 6 as Compromised	Total = 333 Normal = 310 Compromised = 23

9. Automated Fetal Compromise Classification Methods

The development of automated fetal compromise classification approaches began with classical feature-based machine learning methods [35,36] and later evolved to modern deep learning methods [37,86]. Classical methods use additional steps like feature extraction and selection to identify the most relevant features from the raw CTG signals, while modern deep learning methods work directly on the raw CTG signals. In this section, we review the basis of each stream. Figure 5 provides an overview of how the pre-processing steps and classification methods work in sequence.

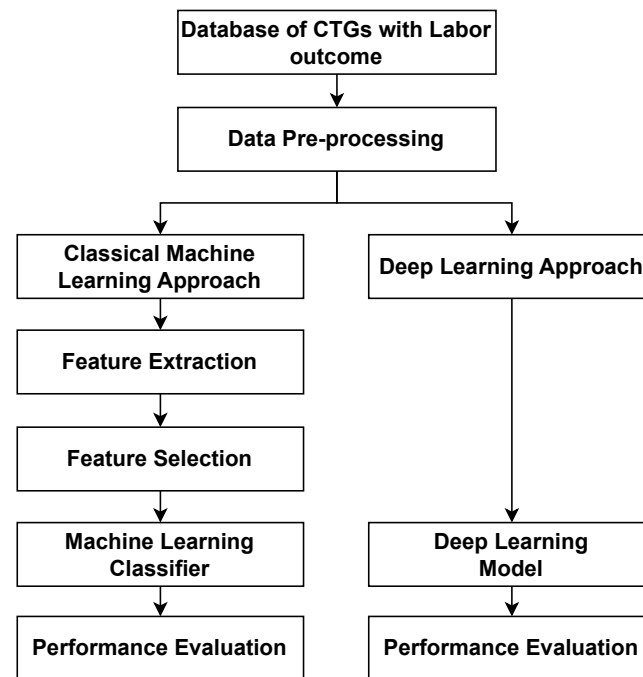


Figure 5. An overview of the steps for the computerised analysis of fetal compromise showing the classical machine learning approach and deep learning approach.

9.1. Feature Extraction

Feature extraction plays a crucial role in fetal compromise detection. This process is aimed at extracting relevant information or characteristics from the raw CTG signals. These extracted features serve as indicators of fetal health, allowing healthcare professionals or computerised methods to monitor and evaluate the condition of the fetus. The extracted features used in classification algorithms in existing studies can be broadly categorised into the following categories: (1) morphological and time domain, (2) frequency domain and (3) non-linear.

9.1.1. Morphological and Time Domain

The basic morphological features defined by the FIGO guidelines for evaluating CTG signals are baseline, acceleration, deceleration and variability [11]. These are characterised by the structure of the FHR waveform and are easily visible to the naked eye. Therefore, these are routinely used by obstetricians and midwives in the visual assessment of CTGs in clinical practice [43]. Inspired by this, initial automatic classification systems used statistical techniques to compute new time domain features such as short-term variability (STV) and long-term irregularity (LTI), which quantify the variability in the FHR in the short term and long term, respectively, as well as the delta, total delta, mean and standard deviation of FHR [100].

9.1.2. Frequency Domain

Frequency domain methods assess the spectral energy content in each frequency component of the FHR and use them as features. The range of the FHR signal in the frequency domain is often divided into four bands: very low frequency (VLF: 0–0.03 Hz, associated with a long time period and non-linear contributions), low frequency (LF: 0.03–0.15 Hz, related to fetal sympathetic nervous system activity), high frequency (HF: 0.5–1 Hz, reflecting fetal breathing) and movement frequency (MF: 0.15–0.5 Hz, correlated with fetal movements and maternal activity) [101]. The power spectral density is estimated in each band using Fourier transform, autoregressive or wavelet transform-based models [49]. Inspired by the usage of the ratio of LF/HF in the analysis of adult heart rate variability (HRV), a similar ratio computed as $LF/(MF+HF)$ has been used as another fetal HRV feature as it quantifies the balance of parts of the fetal autonomic nervous system [102]. Although the frequency domain features can inspect periodic trends in heart rate variations, they are sensitive to artefacts and do not identify non-periodic trends in the FHR variations [49].

9.1.3. Non-Linear Domain

Non-linear features were introduced to investigate and quantify non-periodic variations of the FHR signals. Almost all non-linear features used for FHR analysis are introduced from adult HRV research. Among them, the most frequently used are approximate entropy and sample entropy, which contain information on the fetal state [103], as well as Lempel Ziv Complexity, which is a non-linear feature used to examine the recurring patterns in a continuous signal [104]. Phase-rectified signal averaging is another non-linear method that is used to calculate the decelerative capacity of FHR. This feature quantifies the downward movement of FHR and has shown significantly better performance for acidosis prediction compared to STV [105]. Symbolic dynamics analysis, fractal analysis, detrended fluctuation analysis and Poincaré maps are other techniques used in the literature to compute other non-linear features for FHR analysis [49]. However, analysing FHR using these features has two major limitations: (1) the accuracy of these measures relies on the quality of the FHR signal and (2) a specific data length needs to be selected to obtain a reliable estimate of the values for non-linear features [106].

9.2. Feature Selection

All features calculated across these categories might not be as informative as expected and might contain overlapping information. Therefore, a feature selection or dimensionality reduction step is often used to determine the most influential features that contain useful information for classification. An optimal selection of features will in return improve the computational efficiency and potentially improve the discriminative capability of the classifier [35]. Principal component analysis (PCA) [107], information gain [95], and relevance in estimation features (RELIEF) [35] are feature selection methods often used in literature.

9.3. Classical Machine Learning Classifiers

Initially, research in automated fetal compromise detection was based on simple logic-based computer software algorithms that resembled the clinical decision-making process based on the FHR baseline, accelerations and deceleration [29]. Later, with the advancement of machine learning algorithms, these methods utilised feature extraction and feature selection steps prior to use in classical ML algorithms such as Bayesian models [108], support vector machines (SVM) [39,98,109], adaptive boosting (AdaBoost) [43], random forest (RF) [36,110], decision trees (DT) [43], deep Gaussian processes (DGP) [41], logistic regression (LR) [93] and artificial neural networks (ANN) [111,112] for detecting fetal compromise. The following sections summarise feature-based machine-learning methods used for fetal compromise detection.

9.3.1. AdaBoost

Adaboost is an ensemble learning algorithm that combines multiple weak classifiers to create a strong classifier to reduce overfitting and improve performance. Spilka et al. [35] utilised the nearest mean classifier with AdaBoost to classify pathological cases of FHR using more than 50 features taken from morphological, frequency and non-linear domain features. This method utilised the RELIEF technique for feature selection and the synthetic minority oversampling technique (SMOTE) to tackle the class imbalance of the data before classification. Zhao et al. [43] proposed new software for computerised analysis of the FHR signal (CAS-FHR) that extracts 47 comprehensive features covering morphological, time, frequency and nonlinear domains. The same study compared decision trees, SVM and AdaBoost machine learning algorithms for assessing fetal state and found that AdaBoost had a stronger performance for classification.

9.3.2. Artificial Neural Networks

Artificial neural networks are highly flexible and powerful models inspired by the human brain, capable of learning complex patterns and relationships in data through interconnected layers of neurons. Georgieva et al. [97] used six clinical features and six features from the FHR as input data to a feed-forward ANN. PCA was utilised for feature selection prior to using it as input to the ANN. Classifier performance was improved by averaging the output of 10 independently trained ANNs and testing on a larger dataset consisting of 7568 cases. Comert and Kocamaz [113] showed that ANNs gave better results in classifying FHR signals as hypoxic or normal compared to other ML algorithms like SVMs, extreme learning machines, radial basis function networks, and RFs. In another study, Comert and Kocamaz [111] analysed the effects of linear and non-linear features of FHR on the performance of detecting fetal compromise for three stages of labour. By using ANN as the classifier, they identified that using linear and non-linear features together gave the best performance and that the contribution of non-linear features was greatest for the second stage of labour.

9.3.3. Bayesian Models

Bayesian models provide a probabilistic framework for modelling uncertainty and incorporating prior knowledge. Dash et al. [108] employed Bayesian models to classify fetal status based on selected FHR features and found that it demonstrated better performance compared to SVMs.

9.3.4. Decision Trees

Decision trees are simple yet powerful algorithms that provide an interpretable and effective approach for fetal compromise detection. Zhao et al. [43] compared decision trees with other ML algorithms and demonstrated their efficacy in assessing fetal state. However, its performance was lower than the AdaBoost model compared in the same work.

9.3.5. Deep Gaussian Processes

Deep Gaussian Processes are flexible models that can capture complex relationships while also providing uncertainty estimates to aid decision-making. Feng et al. [41] used a different approach by utilizing the UC signal with the corresponding FHR signal in DGPs to improve the performance of the classification of fetal well-being.

9.3.6. Logistic Regression

Logistic regression is a linear classifier that can be used to model the relationship between input features and the probability of fetal compromise. O'Sullivan et al. [93] showed that including electronic health records and the duration of the last two stages of labour in addition to the CTG features improved the performance of the LR model in identifying fetal compromise.

9.3.7. Random Forest

Random forest is an ensemble method that can combine multiple decision trees, providing robustness against overfitting and the ability to capture complex relationships. Spilka et al. [36] used latent class analysis to define classes and employed the RF algorithm for classification as it provides faster training than the boosting classifiers. Afridi et al. [110] also showed the effectiveness of using RF models for discriminating fetal status.

9.3.8. SVM

Support vector machines are binary classifiers that find the best hyperplane to separate classes by maximizing the margin between them, allowing for effective handling of both linear and non-linear decision boundaries. Georgoulas et al. [114] proposed a novel method to extract scale dependant features of FHR signal using discrete wavelet transform (DWT) and used SVMs for classification. In [107], Georgoulas et al. used SVMs to predict the risk of metabolic acidosis using FHR features extracted from the time and frequency domain in addition to its morphological features. The experimental results of this work indicated that the SVM classifier with the radial basis function (RBF) kernel shows better classification performance than using SVMs with polynomial kernels. SVM is also used in [109] to detect normal or at-risk fetuses based on FHR features extracted through empirical mode decomposition. A different study by Stylios et al. [39] showed that the least square support vector machine (LS-SVM) can be used to discriminate compromised from healthy fetuses. In another study [44], LS-SVM showed increased classification accuracy of the fetal hypoxic assessment when features found using the novel approach called image-based time-frequency (ITBF) analysis were used. Similarly, sparse SVM classification managed to outperform clinical practice in fetal acidosis detection by selecting only a small number of features [98].

9.4. Deep Learning-Based Classifiers

Deep learning-based classifiers are powerful modern approaches that greatly benefit from the availability of well-defined, open-access databases as they allow for the learning of complex characteristics independently from the raw data without relying on the features derived from human knowledge. In prior works on automatic fetal compromise detection, the most typically used deep learning algorithms are convolutional neural networks (CNN) and long short-term memory networks (LSTM). CNNs can capture hidden characteristics of both spatial and temporal data, while LSTMs are capable of capturing both long- and short-term dependencies in time-series data [87].

These methods have high computational complexity during the training process, but once trained they can be configured to predict fetal compromise in real-time. However, to achieve this, deep learning models should be trained using diverse and large datasets for the model to generalise better. Otherwise, these methods can overfit the training data and perform poorly on unseen data. The following sections summarise the recent deep learning-based methods used for fetal compromise detection.

9.4.1. Convolutional Neural Networks

CNNs are deep learning models specifically designed to automatically extract relevant features from raw data, enabling the detection of fetal compromise by capturing intricate patterns and spatial dependencies. One of the main benefits of CNNs is their ability to simplify the model through weight sharing and subsampling [115]. Nevertheless, CNNs primarily learn local features and lack knowledge about the global relationships within the data. Additionally, CNNs require a substantial number of training samples to enhance the model's generalisation. In the domain of FHR classification, the effectiveness of 1D CNN has been demonstrated. This type of CNN can process a time series of FHR signal data, extracting features by applying various kernels to classify the signal as normal or compromised. Petrozziello et al. [87] compared both LSTM and CNN models using a dataset of over 35,000 labouring CTGs. The results indicated that both LSTM and CNN outperformed

clinical practice in predicting fetal compromise and concluded that CNNs surpassed conventional classification methods based on feature extraction. Petrozziello et al. [37] further explored the use of a multimodal CNN (MCNN) for predicting fetal acidemia at birth. This approach employed a quality vector, UC, and FHR signals as inputs to the MCNN model. Their MCNN outperformed the current clinical practice and OxSys1.5 [34] prototype system in evaluating fetal compromise during labour. Additionally, a stacked MCNN was proposed, where the output from the first MCNN, computed on the last 60 min of the first stage of labour CTG signals, served as input to a second MCNN. The second MCNN also used the last 30 min of the second stage of labour CTG signals as input, ultimately providing the final classification outcome. However, the best performance was reported by the MCNN when analysed on the last 60 min of CTG signal regardless of labour. In a different study, Li et al. [116] divided a 20-min FHR segment into 10 window segments and used a CNN to process them in parallel. A majority voting technique was used to provide the final outcome. They showcased the superiority of the CNN approach by comparing it with a feature-based SVM and a multilayer perceptron classification. Zhao et al. [86] presented a different approach using a 2D CNN with 2D images obtained from continuous wavelet transform (CWT) to automatically predict fetal acidemia. Furthermore, the same authors utilised a 2D CNN with 2D images constructed using a recurrence plot (RP) to capture the non-linear characteristics of the FHR signals, aiming to predict fetal hypoxia.

9.4.2. Long Short-Term Memory Networks

LSTMs are a type of recurrent neural network that can effectively model temporal dependencies in sequential data, making them suited for fetal compromise detection by capturing long-term dependencies and recognizing patterns over time. The study by Petrozziello et al. [87] examined CNN and LSTM as two standalone models and demonstrated that CNNs outperform LSTM in predicting fetal compromise. However, work by Liu et al. [46] shows that a novel attention-based method combining CNN and a bi-directional LSTM (CNN-BiLSTM) can achieve satisfactory performance in detecting fetal compromise. Its CNN-BiLSTM captures complex non-linear, spatial and temporal characteristics of the FHR, while the attention mechanism focuses on the important features of the input. However, this architecture showed overfitting, and DWT was used to obtain another feature to reduce the issue of overfitting.

A summary of computerised fetal compromise detection methods along with their strengths and limitations is presented in Table 4.

9.5. Performance Evaluation

Fetal compromise detection using CTG data suffers from a high imbalance in classes. For example, only about 7% of CTU-UHB traces fall under compromised cases when a pH threshold of 7.05 is used as the outcome measure. Several approaches like under-sampling the majority class [117], oversampling the minority class [35] and using class weights in the loss function [37] are used to tackle the disproportion in classes. Under-sampling is usually not desirable as it may result in the loss of important data [118].

When evaluating an ML or DL model, the whole dataset is separated into two sets: a training set and a test set. When only a single division of a training set and a test set is allocated for evaluation, this is called a holdout. Despite the faster evaluation of this method, the model performance has a significant bias towards the data points allocated to the test set and therefore will depend on how the division is made. Especially when smaller datasets are used, the generalizability of the trained model will be lower. Hence, to overcome this limitation k-fold cross-validation is typically used. Here, the dataset is divided into k subsets and the holdout method is repeated k times. At each time, one of the subsets is kept for testing while the remaining subsets are used for training. Finally, the average performance across all k trials is reported.

Table 4. Summary of computer-based fetal compromise detection methods and their strengths and limitations.

Computerised Method		Strengths	Limitations	Computational Cost
Features	Morphological and Time Domain	Macroscopic features suitable for visual inspection Recognised clinical value for several features	Some features based on statistical computation with no direct link to fetal physiology	NA*
	Frequency Domain	Capture periodic trends in FHR variations	Difficult to observe via visual inspection Sensitive to artefacts Does not identify non-periodic trends in FHR variations	
	Non-Linear	Quantify complex non-periodic variations of FHR	Sensitive to artefacts Values depend on the choice of parameters Some highly depend on the FHR signal length	
Classical Machine Learning		Internal operation is more easily understandable	Human involvement needed for feature extraction and selection	Low–High
ML Classifiers	Adaboost	Reduces the risk of overfitting by combining multiple weak classifiers	Computationally expensive due to the iterative nature of the algorithm	Medium
	ANN	Learns complex relationships among features	Prone to overfitting when a higher number of layers used	Medium
	Bayesian Models	Can incorporate prior knowledge and domain expertise through prior distributions	Computationally expensive for large datasets	High
	Decision Trees	Easy to interpret and visualise	Prone to overfitting if not properly regularised	Low
	Deep Gaussian Processes	Can learn complex, non-linear relationships	Can be challenging to interpret and visualise Computationally expensive for large datasets	High
	Logistic Regression	Simple and efficient Provides probabilistic outputs useful for interpretation	May not capture complex interactions between features	Low
	Random Forest	Robust against overfitting due to ensemble of decision trees	Computationally expensive for large datasets	Medium
	SVM	Effective in high-dimensional spaces Works well on small datasets Can handle linear and non-linear decision boundaries	Can be sensitive to the choice of kernel function and hyperparameters Computationally expensive for large datasets	Medium
Deep Learning		Feature extraction and selection are not needed Learns complex features from raw data	Lack of interpretability and transparency of operation High computational complexity	High
DL Classifiers	CNN	Reduces the complexity of the model by weight sharing and subsampling	Does not learn global relationships Requires more training samples for generalisation	High
	LSTM	Widely used for time series forecasting Learns temporal features	Takes longer time to train Requires more training samples for generalisation	Higher

* NA = Not applicable.

To measure the performance of computerised fetal compromise detection algorithms, various metrics are utilised in the literature. Traditional metrics like accuracy and error rate are not suitable for evaluating the performance of the classifiers on imbalanced data as they are sensitive to the distribution of the data. Alternative measures, including sensitivity, specificity, F-measure, Matthew’s correlation coefficient and geometric mean derived from the confusion matrix, are typically used to present their performance [118].

Sensitivity and specificity are preferred metrics as the objective of computerised fetal compromise detection is to maximise the number of the correctly predicted positive class while minimising the number of the incorrectly predicted negative class. However, sensitivity and specificity are often inversely correlated, so some trade-off in specificity has to be

tolerated. Often, receiver operating characteristics (ROC) curves are formed by plotting true positive rates (equivalent to sensitivity) against false positive rates (equivalent to 100%—specificity) for different thresholds. These curves provide a graphical representation of the performance of the classifier across different sensitivity and specificity thresholds. A single metric called area under the curve (AUC) is computed after plotting the ROC to assess the performance of different classifiers. AUC values greater than 0.5 mean that the model is performing better than random chance, while an AUC value of 1.0 indicates perfect classification. A summary of the evaluation metrics used in the literature is given in Table 5. The performance of computerised CTG analysis methods according to these performance metrics is given in Table 6.

Table 5. Summary of evaluation metrics used in the literature.

Metric Name	Equation	Description
Accuracy (Acc)	$Acc = \frac{TP+TN}{TP+TN+FP+FN}$	The simple ratio between the number of correctly predicted points to the total number of points (probability of correct predictions) Not suitable for imbalanced datasets
Sensitivity (Se)	$Se = \frac{TP}{TP+FN}$	The proportion of the correctly predicted positive instances from the total positive instances
Specificity (Sp)	$Sp = \frac{TN}{TN+FP}$	The proportion of the correctly predicted negative instances from the total negative instances
Precision	$precision = \frac{TP}{TP+FP}$	The proportion of the correctly predicted positive instances from the total classified positive instances
Geometric mean (g – mean)	$g - mean = \sqrt{Se \times Sp}$	Measure of the balance between classification performances in both the majority and minority classes
Harmonic mean (F – measure)	$F - measure = \frac{2 \times Se \times precision}{Se + precision}$	A measure of the effectiveness of classification
Matthew’s correlation coefficient (MCC)	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$	Minimally influenced by imbalanced data, the correlation coefficient between the observed and predicted classifications (range from –1 to +1), +1 : perfect prediction 0 : no better than random prediction –1 : worst prediction
Area under the receiver operating characteristic curve (AUC)	Plot of the true positive rate vs. the false positive rate at all possible thresholds	Higher the AUC, the better the performance of the model at distinguishing between the classes Used to compare and evaluate different classification algorithms

Table 6. Performance of computer-based approaches for fetal compromise detection on the open-source CTU-UHB dataset. Performance is reported per classifier type when multiple approaches are used in a single work.

Year	Authors	Training Method [†]	Sensitivity (%)	Specificity (%)	AUC	Independent Test/Train Data	Complete CTU-UHB at Threshold pH < 7.05 or pH ≤ 7.05
2013	Spilka et al. [35]	ID: FHR features augmented by SMOTE CT: Nearest mean classifier with AdaBoost CV: 44-fold	64.00	65.00	Not specified	Yes	Yes
2014	Spilka et al. [36]	ID: FHR features CT: RF CV: 2-fold repeated 5 times	72.00	78.00	Not specified	Yes	No, subset of dataset used and labelled using clinical annotations.
2016	Cömert and Kocamaz [112]	ID: FHR features CT: ANN CV: 5-fold	88.70	85.10	Not specified	Yes	No, subset of 100 records randomly chosen.
2016	Cömert and Kocamaz [111]	ID: FHR features CT: ANN, for 3 stages of labour CV: 10-fold	I = 95.89 II = 87.06 III = 85.87	I = 74.75 II = 75.90 III = 72.73	Not specified	Yes	No, complete dataset used but labelled using clinical annotations.
2016	Spilka et al. [38]	ID: FHR features CT: Sparse SVM CV: Train—Lyon Dataset, Test—CTU-UHB	40.00	86.00	0.79	Yes	No, subset of dataset which has less than 50% signal loss.

Table 6. Cont.

Year	Authors	Training Method [†]	Sensitivity (%)	Specificity (%)	AUC	Independent Test/Train Data	Complete CTU-UHB at Threshold pH < 7.05 or pH ≤ 7.05
2016	Stylios et al. [39]	ID: FHR features CT: LS SVM with RBF kernel CV: 44-fold repeated 15 times	68.50	77.70	Not specified	Yes	Yes
2017	Georgoulas et al. [40]	ID: FHR features CT: LS SVM CV: 44-fold	72.12	65.30	Not specified	Yes	Yes
2018	Cömert and Kocamaz [47]	ID: FHR features CT: ANN, SVM, k-NN CV: 10-fold repeated 30 times	ANN = 68.52 SVM = 76.83 k-NN = 53.28	ANN = 70.29 SVM = 78.27 k-NN = 66.80	ANN = 0.76 SVM = 0.84 k-NN = 0.64	Yes	No, complete dataset used but with compromise defined by pH < 7.2.
2018	Feng et al. [41]	ID: FHR and UC features CT: Supervised two-layer DGP network CV: No CV but repeated 5 times	FHR = 73.00 FHR+UC = 91.00	FHR = 91.00 FHR+UC = 82.00	Not specified	Yes	No, subset of dataset used with compromise defined by pH < 7.1.
2018	Petrozziello et al. [87]	ID: Raw FHR and UC CT: CNN, LSTM CV: Train—Oxford dataset, Test—CTU-UHB	Not specified	Not specified	CNN = 0.82 LSTM = 0.81	Yes	Yes
2018	Zhao et al. [43]	ID: FHR features CT: AdaBoost CV: 10-fold CV	92.00	90.00	0.91	No, unclear whether data samples have been used multiple times.	No, complete dataset used with compromise defined by pH < 7.15.
2018	Cömert et al. [44]	ID: FHR features CT: LS SVM with RBF kernel CV: 10-fold repeated 100 times	63.45	65.88	0.65	Yes	No, complete dataset used with compromise defined by pH < 7.15.
2018	Cömert et al. [88]	ID: FHR features CT: SVM with RBF kernel CV: 10-fold repeated 30 times	57.42	70.11	Not specified	Yes	No, complete dataset used with compromise defined by pH < 7.15.
2019	Petrozziello et al. [37]	ID: Raw FHR and UC, FHR quality score CT: MCNN, stacked MCNN CV: Train—Oxford dataset, Test—CTU-UHB	MCNN * 33.00 48.00 58.00 65.00	MCNN * 95.00 90.00 85.00 80.00	MCNN = 0.81	Yes	Yes
			Stacked MCNN * 33.00 45.00 58.00 65.00	Stacked MCNN * 95.00 90.00 85.00 80.00	Stacked MCNN = 0.82		
2019	Zhao et al. [86]	ID: Raw FHR transformed to 2D using CWT CT: 8-layer CNN CV: 10-fold	98.22	94.84	0.97	No, data augmented before data split.	No, complete dataset used with compromise defined by pH < 7.15.
2019	Zhao et al. [45]	ID: Raw FHR transformed to 2D image using RP CT: 8-layer CNN CV: 10-fold	99.29	98.1	0.98	No, data augmented before data split.	No, subset of dataset used with compromise defined by pH < 7.15.
2021	Liang and Li [89]	ID: Raw FHR CT: CNN based on a weighted voting mechanism CV: No CV but holdout test set	80.93	79.85	0.90	Yes	No, performance is reported on a small holdout set of the dataset.
2021	Liu et al. [46]	ID: Raw FHR CT: CNN + BiLSTM + Attention + DWT CV: 10-fold repeated 10 times	75.23	70.82	Not specified	Yes	No, complete dataset used with compromise defined by pH < 7.15.
2021	O'Sullivan et al. [93]	ID: FHR and UC features, EHR, duration of stage I and II labour CT: Logistic regression CV: 5-fold	82.60	77.70	0.81	Yes	No, subset of dataset used with compromise defined by pH < 7.0 or Apgar ≤ 6.

[†] ID = input data, CT = classifier type and CV = cross validation. * Multiple sensitivity/specificity pairs given.

10. Discussion

This review highlights that a comprehensive understanding of the complex and heterogeneous nature of cardiotocography signals for detecting fetal compromise remains an open challenge. In addition to questions about the most relevant physiological features of CTG signals for interpreting fetal state, we draw attention to several additional technical challenges, as follows.

It can be seen that a common challenge faced by all computerised CTG analysis methods is the quality of the CTG data acquired. Petrozziello et al. [37] showed that quantification of signal quality is important and poor signal quality results in the adverse performance of the models. They achieved the current state-of-the-art AUC of 0.82 on the entire CTU-UHB database in fetal compromise prediction with pH threshold < 7.05 by using a quality parameter as an input along with FHR and UC in their MCNN model. Although direct fECG provides superior signal quality, its invasiveness prevents its widespread use. The performance of the non-invasive Doppler ultrasound-based CTG and NI-fECG methods also differ significantly from the current gold standard of direct fECG measurement. According to prior studies, NI-fECG demonstrates a mean positive percent agreement with the fetal scalp electrode between 81 and 83%, compared to 62–73% for Doppler ultrasound CTG [61,62]. Despite the lower agreement, Doppler ultrasound-based CTG is still the most widely used monitoring technique. For instance, about 75% of the CTU-UHB dataset is

acquired from Doppler ultrasound-based CTG. This raises questions of whether the performance of all computerised CTG analysis methods are impacted by the quality of Doppler ultrasound-based records. We suggest that future work should look at the viability of using only direct fetal ECG or NI-fECG records for training computerised CTG algorithms for fetal compromise prediction and compare these to ultrasound-based methods.

Further, it is important to note that the direct fetal ECG or NI-fECG terms in the context of these studies are more closely related to QRS detection and heart rate trends rather than true fECG, which entails more comprehensive cardiac waveform display and time intervals. Therefore, we suggest that future works may also evaluate the potential of the raw fECG signal and its features for detecting fetal compromise.

As highlighted by Table 3, different classification criteria like base deficit, Apgar scores, and umbilical cord pH are used to define the fetal compromise where the latter is the most widely used. According to [119], the umbilical artery base deficit is comparable or inferior to pH as a perinatal outcome measure. Further, the results of [119,120] show that low cord pH is strongly associated with poor perinatal outcomes and therefore is a strong candidate for verifying the performance of computerised CTG analyses. Different pH thresholds such as 7.05 [35,37–40], 7.10 [41,42], 7.15 [43–46] and 7.20 [47] have been used in existing studies to define the fetal compromised babies. This use of multiple pH thresholds hinders the comparison, improvement and criticism among different proposed methods. Therefore, agreeing on common criteria by the research community is essential to drive this field forward. As can be seen in Table 3, a pH threshold of 7.05 is the most widely used criterion to define fetal compromise. Therefore, we propose that future works should adopt this criterion as the standard to define fetal compromise and use it when reporting the performance of new computerised methods for fetal compromise detection.

As the performance of the algorithms depends on the data itself, pre-processing plays a large role in the comparison of different computerised analysis methods. Comparing different studies without a standard pre-processing procedure may lead to incorrect conclusions about whether a particular classifier or feature set is superior to a previously identified work. We have summarised the typical pre-processing steps in Figure 4, but there is yet no conclusive evidence of the optimum signal length, artefact removal, interpolation method, downsampling frequency or smoothing technique that should be used. Future work should try to address these questions to define a standard pre-processing methodology. Moreover, data manipulation, like data augmentation, should be carefully carried out such that no information is leaked into the test set. If this happens, the results will be subject to bias and the conclusions could be misleading [121]. For example, we believe that the performances achieved by the two studies [45,86] are likely not representative of real-world performance as the data is augmented using CWT and RP before the data is split into training and test sets. Therefore, care should be taken when pre-processing and splitting the data so that the independence of the test set is maintained.

Furthermore, existing analysis methods primarily use static and long segment lengths taken from the end of the CTG recording for detecting fetal compromise [37,89,98]. This results in predictions only made close to the time of delivery, leaving inadequate time for the clinicians to make clinical interventions. Therefore, studies such as [98] have suggested that future work should assess the evolution of FHR over time rather than evaluating a static segment. A method of this nature will be more clinically relevant as this could potentially detect fetal compromise events much earlier and enable clinical intervention.

In considering prediction at earlier time points, currently a single label is given to the overall CTG trace based on a clinical metric typically taken after birth (i.e., pH < 7.05), making it a ‘weak label’ where it is unclear whether a particular section of the CTG trace contains an abnormality or not. In particular, if it is assumed that the labels are correct, the information regarding the specific location of the abnormality on the CTG trace is not known. Therefore, there is a significant risk of introducing noisy labels unless the distress is chronic and prevails throughout the extent of the trace. However, different types of fetal hypoxia exist such as acute, subacute, evolving and chronic, and they generally

occur in different forms [56]. Some cases show evidence of distress in the CTG but no acidemia in blood gas analysis. One reason could be the treatment paradox. A treatment paradox is when the strong predictor of a complication results in rapid intervention and improved outcomes as a result of effective treatment [122], thereby making the class labels unreliable. An approach to making the labels more reliable in theory would be to use clinical experts to annotate the CTGs. This consideration has led to the creation of two new sets of expert annotated labels for the CTU-UHB dataset, which are also available for public access [21,123]. Conversely, this would again lead to the primary issue of human bias in deciding the labels for each recording.

Another aspect studied by researchers is the usage of electronic health records (EHR) as input to the classifier. The study [93] used clinical variables, including gestation, parity, hypertension and the duration of stages of labour to show that the classifier performance can be enhanced by using EHR information. These records can provide valuable information on the baby and the mother, but it is essential that we only include health records that are known prior to or during labour because the end goal is the real-time detection of fetal compromise during labour.

In evaluating the performance of the computerised approaches for fetal compromise detection, various metrics have been used in the literature. However, the main objective of these methods should be to achieve a high sensitivity while maintaining high specificity. Since specificity decreases when sensitivity increases, a certain degree of specificity decrease has to be tolerated. The current clinical sensitivity or true positive rate (TPR) in detecting fetal compromise is about 31–48% at a 16–21% false positive rate (FPR) [34,37,124]. Therefore, the research community should agree to present their performance in terms of TPR at approximately 10–20% FPR such that it can be compared with current clinical performance. This then raises an important question of what would be a necessary level to use in clinical practice. According to Georgieva et al. [34], the performance should exceed 60% TPR at 15% FPR to gain tangible clinical benefits. Nevertheless, this is still debatable, and consensus must be reached to achieve clinical translation in the future.

As can be seen from Table 6, the current best performance achieved on the complete CTU-UHB dataset with pH threshold < 7.05 is 33–65% TPR at 5–20% FPR by Petroziello et al. [37]. Although there are other studies that have reported higher performance, they have either used a different class definition criteria or have only used a subset of the dataset to report performance. A subset of the dataset is chosen by some studies [38,45] either to balance the distribution of classes or to remove records based on signal loss. We believe that the imbalanced nature of the compromised and normal classes should be accepted as it reflects the real world, and choosing a subset of data to balance the classes is not ideal. Rather, other means of tackling the class imbalance should be sought such as using class weights [37]. Furthermore, all studies should cross-validate their results on the complete CTU-UHB dataset such that the reported performances are generalised and unbiased. However, the validation on CTU-UHB is still small in terms of the number of recordings compared to large clinical studies as presented in Section 4, which range between 7730–46,042 participants [33,77].

Another possible reason for the low performance of computer-based approaches could be the limited number of public CTG datasets. Currently, accessible databases have a very low number of compromised cases and are typically taken from a single institution. The CTU-UHB with 552 raw FHR and UC records with pH values is available for public access as a benchmark dataset to evaluate and compare the performance of new algorithms. Even so, much larger, multi-centre datasets with well-documented outcomes, including obstetric clinical information, are needed for the automatic computer-based classification methods to be robust and reliable in clinical use [125]. This cannot be achieved by a single stakeholder, and the field should look to create collaborations between multinational and multidisciplinary research groups to solve this challenge collectively [126]. This will help further demonstrate the performance of these data-driven methods prior to prospective clinical validation. Furthermore, a consensus on the pre-processing steps,

outcome measures and evaluation metrics must be reached to allow for transparency, reproducibility and comparison among different approaches.

11. Conclusions

To our knowledge, this is the first review to summarise the physiological basis for CTG, FHR and UC monitoring technologies; computerised CTG analysis systems in clinical practice; public and private CTG datasets; pre-processing steps; and classification methods used in automated data-driven approaches for fetal compromise detection. The quality of the CTG data is one of the main obstacles in achieving the performance required for widespread clinical translation. Therefore, further research on emerging monitoring technologies such as NI-fECG and EHG is warranted to improve the underlying data quality for building automated methods. We also believe that a standardised pre-processing workflow and criteria for the classification of fetal compromise would address criticisms that impact progress in this field while allowing for reproducibility and comparability. Furthermore, rather than working on private datasets, we recommend, where possible, making datasets public as advancing the performance of deep learning methods depends on large multi-center datasets to improve generalisability.

Author Contributions: Conceptualisation, L.M. and E.K.; writing—original draft preparation, L.M.; writing—review and editing, E.K., M.P. and F.B.; and supervision, E.K., M.P. and F.B. All authors have read and agreed to the published version of the manuscript.

Funding: L.M. is supported by the Melbourne Research Scholarship and a scholarship from the Graeme Clark Institute for Biomedical Engineering at the University of Melbourne. E.K. is supported by an Early Career Research Fellowship from the Department of Obstetrics and Gynaecology at the University of Melbourne. F.B. is supported by a NHMRC Early Career Fellowship (NHMRC #1142636) and a Norman Beischer Clinical Research Fellowship.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: E.K., M.P. and F.B. are co-founders and shareholders in Kali Healthcare Pty Ltd. L.M. has no conflict of interest to declare.

References

1. United Nations Inter-agency Group for Child Mortality Estimation (UN IGME). *A Neglected Tragedy: The Global Burden of Stillbirths*; Report; United Nations Children's Fund: New York, NY, USA, 2020.
2. Lawn, J.E.; Lee, A.C.; Kinney, M.; Sibley, L.; Carlo, W.A.; Paul, V.K.; Pattinson, R.; Darmstadt, G.L. Two million intrapartum-related stillbirths and neonatal deaths: Where, why, and what can be done? *Int. J. Gynecol. Obstet.* **2009**, *107*, S5–S19. [\[CrossRef\]](#)
3. Lawn, J.E.; Gravett, M.G.; Nunes, T.M.; Rubens, C.E.; Stanton, C.; the GAPPS Review Group. Global report on preterm birth and stillbirth (1 of 7): Definitions, description of the burden and opportunities to improve data. *BMC Pregnancy Childbirth* **2010**, *10*, S1. [\[CrossRef\]](#)
4. Vogel, J.; Souza, J.; Mori, R.; Morisaki, N.; Lumbiganon, P.; Laopaiboon, M.; Ortiz-Panoso, E.; Hernandez, B.; Pérez-Cuevas, R.; Roy, M.; et al. Maternal complications and perinatal mortality: Findings of the World Health Organization Multicountry Survey on Maternal and Newborn Health. *Bjog Int. J. Obstet. Gynaecol.* **2014**, *121*, 76–88. [\[CrossRef\]](#)
5. Bhutta, Z.A.; Das, J.K.; Bahl, R.; Lawn, J.E.; Salam, R.A.; Paul, V.K.; Sankar, M.J.; Blencowe, H.; Rizvi, A.; Chou, V.B.; et al. Can available interventions end preventable deaths in mothers, newborn babies, and stillbirths, and at what cost? *Lancet* **2014**, *384*, 347–370. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Goldenberg, R.L.; Harrison, M.S.; McClure, E.M. Stillbirths: The Hidden Birth Asphyxia—US and Global Perspectives. *Clin. Perinatol.* **2016**, *43*, 439–453. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Low, J.A. Intrapartum fetal asphyxia: Definition, diagnosis, and classification. *Am. J. Obstet. Gynecol.* **1997**, *176*, 957–959. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Bennet, L.; Gunn, A.J. The Fetal Heart Rate Response to Hypoxia: Insights from Animal Models. *Clin. Perinatol.* **2009**, *36*, 655–672. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Dilenge, M.E.; Majnemer, A.; Shevell, M.I. Topical Review: Long-Term Developmental Outcome of Asphyxiated Term Neonates. *J. Child Neurol.* **2001**, *16*, 781–792. [\[CrossRef\]](#)

10. Kawagoe, Y.; Sameshima, H. Hypoxia: Animal experiments and clinical implications. *J. Obstet. Gynaecol. Res.* **2017**, *43*, 1381–1390. [[CrossRef](#)]
11. Ayres-de Campos, D.; Spong, C.Y.; Chandrachan, E.; FIGO Intrapartum Fetal Monitoring Expert Consensus Panel. FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *Int. J. Gynecol. Obstet.* **2015**, *131*, 13–24. [[CrossRef](#)]
12. Ayres-de Campos, D. Electronic fetal monitoring or cardiotocography, 50 years later: what's in a name? *Am. J. Obstet. Gynecol.* **2018**, *218*, 545–546. [[CrossRef](#)] [[PubMed](#)]
13. Reinhard, J.; Hayes-Gill, B.R.; Schiermeier, S.; Hatzmann, W.; Herrmann, E.; Heinrich, T.M.; Louwen, F. Intrapartum signal quality with external fetal heart rate monitoring: a two way trial of external Doppler CTG ultrasound and the abdominal fetal electrocardiogram. *Arch. Gynecol. Obstet.* **2012**, *286*, 1103–1107. [[CrossRef](#)] [[PubMed](#)]
14. Bakker, P.; Colenbrander, G.; Verstraeten, A.; Van Geijn, H. The quality of intrapartum fetal heart rate monitoring. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2004**, *116*, 22–27. [[CrossRef](#)]
15. Bakker, P.C.A.M.; Rijsiwijk, S.V.; Geijn, H.P.v. Uterine activity monitoring during labor. *J. Perinat. Med.* **2007**, *35*, 468–477. [[CrossRef](#)] [[PubMed](#)]
16. Euliano, T.Y.; Nguyen, M.T.; Darmanjian, S.; McGowrray, S.P.; Euliano, N.; Onkala, A.; Gregg, A.R. Monitoring uterine activity during labor: A comparison of three methods. *Am. J. Obstet. Gynecol.* **2013**, *208*, 66.e1–66.e6. [[CrossRef](#)]
17. Lempersz, C.; Noben, L.; van Osta, G.; Wassen, M.L.H.; Meershoek, B.P.J.; Bakker, P.; Jacquemyn, Y.; Cuerva, M.J.; Vullings, R.; Westerhuis, M.E.M.H.; et al. Intrapartum non-invasive electrophysiological monitoring: A prospective observational study. *Acta Obstet. Gynecol. Scand.* **2020**, *99*, 1387–1395. [[CrossRef](#)]
18. Alfrevic, Z.; Gyte, G.M.; Cuthbert, A.; Devane, D. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. *Cochrane Database Syst. Rev.* **2017**. [[CrossRef](#)]
19. Grivell, R.M.; Alfrevic, Z.; Gyte, G.M.; Devane, D. Antenatal cardiotocography for fetal assessment. *Cochrane Database Syst. Rev.* **2015**, *2015*, CD007863. [[CrossRef](#)]
20. Devane, D.; Llor, J.G.; Daly, S.; McGuire, W.; Cuthbert, A.; Smith, V. Cardiotocography versus intermittent auscultation of fetal heart on admission to labour ward for assessment of fetal wellbeing. *Cochrane Database Syst. Rev.* **2017**. [[CrossRef](#)]
21. Hruban, L.; Spilka, J.; Chudáček, V.; Janků, P.; Huptych, M.; Burša, M.; Hudec, A.; Kacerovský, M.; Koucký, M.; Procházka, M.; et al. Agreement on intrapartum cardiotocogram recordings between expert obstetricians. *J. Eval. Clin. Pract.* **2015**, *21*, 694–702. [[CrossRef](#)]
22. Devane, D.; Llor, J. Midwives' visual interpretation of intrapartum cardiotocographs: intra- and inter-observer agreement. *J. Adv. Nurs.* **2005**, *52*, 133–141. [[CrossRef](#)] [[PubMed](#)]
23. Chauhan, S.P.; Klausner, C.K.; Woodring, T.C.; Sanderson, M.; Magann, E.F.; Morrison, J.C. Intrapartum nonreassuring fetal heart rate tracing and prediction of adverse outcomes: Interobserver variability. *Am. J. Obstet. Gynecol.* **2008**, *199*, 623.e1–623.e5. [[CrossRef](#)] [[PubMed](#)]
24. Steer, P.J. Has electronic fetal heart rate monitoring made a difference. *Semin. Fetal Neonatal Med.* **2008**, *13*, 2–7. [[CrossRef](#)]
25. Armstrong, L.; Stenson, B.J. Use of umbilical cord blood gas analysis in the assessment of the newborn. *Arch. Dis. Child. Fetal Neonatal Ed.* **2007**, *92*, F430–F434. [[CrossRef](#)] [[PubMed](#)]
26. Ayres-de Campos, D.; Arulkumaran, S.; Panel, F.I.F.M.E.C. FIGO consensus guidelines on intrapartum fetal monitoring: Physiology of fetal oxygenation and the main goals of intrapartum fetal monitoring. *Int. J. Gynecol. Obstet.* **2015**, *131*, 5–8. [[CrossRef](#)] [[PubMed](#)]
27. Wayenberg, J.L. Threshold of metabolic acidosis associated with neonatal encephalopathy in the term newborn. *J. Matern.-Fetal Neonatal Med.* **2005**, *18*, 381–385. [[CrossRef](#)]
28. Casey, B.M.; McIntire, D.D.; Leveno, K.J. The Continuing Value of the Apgar Score for the Assessment of Newborn Infants. *N. Engl. J. Med.* **2001**, *344*, 467–471. [[CrossRef](#)]
29. Chung, T.K.H.; Mohajer, M.P.; Yang, Z.J.; Chang, A.M.Z.; Sahota, D.S. The prediction of fetal acidosis at birth by computerised analysis of intrapartum cardiotocography. *BJOG Int. J. Obstet. Gynaecol.* **1995**, *102*, 454–460. [[CrossRef](#)]
30. Dawes, G.S.; Moulden, M.; Redman, C.W. System 8000: Computerised antenatal FHR analysis. *J. Perinat. Med.* **1991**, *19*, 47–51. [[CrossRef](#)]
31. Bernardes, J.; Moura, C.; de Sa, J.P.; Leite, L.P. The Porto system for automated cardiotocographic signal analysis. *J. Perinat. Med.* **1991**, *19*, 61–65. [[CrossRef](#)]
32. Keith, R.D.; Greene, K.R. Development, evaluation and validation of an intelligent system for the management of labour. *Bailliere's Clin. Obstet. Gynaecol.* **1994**, *8*, 583–605. [[CrossRef](#)] [[PubMed](#)]
33. Group, T.I.C. Computerised interpretation of fetal heart rate during labour (INFANT): A randomised controlled trial. *Lancet* **2017**, *389*, 1719–1729. [[CrossRef](#)]
34. Georgieva, A.; Redman, C.W.G.; Papageorgiou, A.T. Computerised data-driven interpretation of the intrapartum cardiotocogram: A cohort study. *Acta Obstet. Gynecol. Scand.* **2017**, *96*, 883–891. [[CrossRef](#)]
35. Spilka, J.; Georgoulas, G.; Karvelis, P.; Oikonomou, V.P.; Chudáček, V.; Stylios, C.; Lhotská, L.; Janků, P. Automatic Evaluation of FHR Recordings from CTU-UHB CTG Database. In *Information Technology in Bio- and Medical Informatics, Proceedings of the 4th International Conference, ITBAM 2013, Prague, Czech Republic, 28 August 2013*; Bursa, M., Khuri, S., Renda, M.E., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 47–61. [[CrossRef](#)]

36. Spilka, J.; Georgoulas, G.; Karvelis, P.; Chudáček, V.; Stylios, C.D.; Lhotská, L. Discriminating Normal from “Abnormal” Pregnancy Cases Using an Automated FHR Evaluation Method. In *Artificial Intelligence: Methods and Applications, Proceedings of the 8th Hellenic Conference on AI, SETN 2014, Ioannina, Greece, 15–17 May 2014*; Likas, A., Blekas, K., Kalles, D., Eds.; Springer: Cham, Switzerland, 2014; pp. 521–531. [\[CrossRef\]](#)
37. Petrozziello, A.; Redman, C.W.G.; Papageorgiou, A.T.; Jordanov, I.; Georgieva, A. Multimodal Convolutional Neural Networks to Detect Fetal Compromise During Labor and Delivery. *IEEE Access* **2019**, *7*, 112026–112036. [\[CrossRef\]](#)
38. Spilka, J.; Chudáček, V.; Huptych, M.; Leonarduzzi, R.; Abry, P.; Doret, M. Intrapartum Fetal Heart Rate Classification: Cross-Database Evaluation. In *Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, Paphos, Cyprus, 31 March–2 April 2016; Kyriacou, E., Christofides, S., Pattichis, C.S., Eds.; Springer: Cham, Switzerland, 2016; pp. 1199–1204. [\[CrossRef\]](#)
39. Stylios, C.D.; Georgoulas, G.; Karvelis, P.; Spilka, J.; Chudáček, V.; Lhotska, L. Least Squares Support Vector Machines for FHR Classification and Assessing the pH Based Categorization. In *Proceedings of the XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*, Paphos, Cyprus, 31 March–2 April 2016; Kyriacou, E., Christofides, S., Pattichis, C.S., Eds.; Springer: Cham, Switzerland, 2016; pp. 1211–1215. [\[CrossRef\]](#)
40. Georgoulas, G.; Karvelis, P.; Spilka, J.; Chudáček, V.; Stylios, C.D.; Lhotská, L. Investigating pH based evaluation of fetal heart rate (FHR) recordings. *Health Technol.* **2017**, *7*, 241–254. [\[CrossRef\]](#)
41. Feng, G.; Quirk, J.G.; Djurić, P.M. Supervised and Unsupervised Learning of Fetal Heart Rate Tracings with Deep Gaussian Processes. In *Proceedings of the 2018 14th Symposium on Neural Networks and Applications (NEUREL)*, Belgrade, Serbia, 20–21 November 2018; pp. 1–6. [\[CrossRef\]](#)
42. Gonçalves, H.; Rocha, A.P.; Ayres-de Campos, D.; Bernardes, J. Linear and nonlinear fetal heart rate analysis of normal and academic fetuses in the minutes preceding delivery. *Med. Biol. Eng. Comput.* **2006**, *44*, 847. [\[CrossRef\]](#)
43. Zhao, Z.; Zhang, Y.; Deng, Y. A Comprehensive Feature Analysis of the Fetal Heart Rate Signal for the Intelligent Assessment of Fetal State. *J. Clin. Med.* **2018**, *7*, 223. [\[CrossRef\]](#)
44. Cömert, Z.; Kocamaz, A.F.; Subha, V. Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Comput. Biol. Med.* **2018**, *99*, 85–97. [\[CrossRef\]](#)
45. Zhao, Z.; Zhang, Y.; Comert, Z.; Deng, Y. Computer-Aided Diagnosis System of Fetal Hypoxia Incorporating Recurrence Plot With Convolutional Neural Network. *Front. Physiol.* **2019**, *10*, 255. [\[CrossRef\]](#)
46. Liu, M.; Lu, Y.; Long, S.; Bai, J.; Lian, W. An attention-based CNN-BiLSTM hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification. *Expert Syst. Appl.* **2021**, *186*, 115714. [\[CrossRef\]](#)
47. Cömert, Z.; Kocamaz, A.F. Open-access software for analysis of fetal heart rate signals. *Biomed. Signal Process. Control* **2018**, *45*, 98–108. [\[CrossRef\]](#)
48. Tournier, A.; Beacom, M.; Westgate, J.A.; Bennet, L.; Garabedian, C.; Ugwumadu, A.; Gunn, A.J.; Lear, C.A. Physiological control of fetal heart rate variability during labour: implications and controversies. *J. Physiol.* **2022**, *600*, 431–450. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Ponsiglione, A.M.; Cosentino, C.; Cesarelli, G.; Amato, F.; Romano, M. A Comprehensive Review of Techniques for Processing and Analyzing Fetal Heart Rate Signals. *Sensors* **2021**, *21*, 6136. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Ribeiro, M.; Monteiro-Santos, J.; Castro, L.; Antunes, L.; Costa-Santos, C.; Teixeira, A.; Henriques, T.S. Non-linear Methods Predominant in Fetal Heart Rate Analysis: A Systematic Review. *Front. Med.* **2021**, *8*, 661226. [\[CrossRef\]](#)
51. Castro, L.; Loureiro, M.; Henriques, T.S.; Nunes, I. Systematic Review of Intrapartum Fetal Heart Rate Spectral Analysis and an Application in the Detection of Fetal Acidemia. *Front. Pediatr.* **2021**, *9*, 661400. [\[CrossRef\]](#)
52. O’Sullivan, M.E.; Considine, E.C.; O’Riordan, M.; Marnane, W.P.; Rennie, J.M.; Boylan, G.B. Challenges of Developing Robust AI for Intrapartum Fetal Heart Rate Monitoring. *Front. Artif. Intell.* **2021**, *4*, 765210. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Ben M’Barek, I.; Jauvion, G.; Ceccaldi, P. Computerised cardiotocography analysis during labor—A state-of-the-art review. *Acta Obstet. Gynecol. Scand.* **2022**, *102*, 130–137. [\[CrossRef\]](#)
54. Bernardes, J. Computerised analysis of cardiotocograms in clinical practice and the SisPorto® system thirty-two years after: Technological, physiopathological and clinical studies. *J. Perinat. Med.* **2023**, *51*, 145–160. [\[CrossRef\]](#)
55. Nageotte, M.P. Fetal heart rate monitoring. *Semin. Fetal Neonatal Med.* **2015**, *20*, 144–148. [\[CrossRef\]](#)
56. Pinas, A.; Chandharan, E. Continuous cardiotocography during labour: Analysis, classification and management. *Best Pract. Res. Clin. Obstet. Gynaecol.* **2016**, *30*, 33–47. [\[CrossRef\]](#)
57. Ugwumadu, A. Understanding cardiotocographic patterns associated with intrapartum fetal hypoxia and neurologic injury. *Best Pract. Res. Clin. Obstet. Gynaecol.* **2013**, *27*, 509–536. [\[CrossRef\]](#)
58. Divon, M.Y.; Torres, F.P.; Yeh, S.Y.; Paul, R.H. Autocorrelation techniques in fetal monitoring. *Am. J. Obstet. Gynecol.* **1985**, *151*, 2–6. [\[CrossRef\]](#) [\[PubMed\]](#)
59. Carter, M.C. Signal processing and display—cardiotocographs. *Bjog Int. J. Obstet. Gynaecol.* **1993**, *100*, 21–23. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Cohen, W.R.; Hayes-Gill, B. Influence of maternal body mass index on accuracy and reliability of external fetal monitoring techniques. *Acta Obstet. Gynecol. Scand.* **2014**, *93*, 590–595. [\[CrossRef\]](#)
61. Cohen, W.R.; Ommanni, S.; Hassan, S.; Mirza, F.G.; Solomon, M.; Brown, R.; Schiffrin, B.S.; Himsworth, J.M.; Hayes-Gill, B.R. Accuracy and reliability of fetal heart rate monitoring using maternal abdominal surface electrodes. *Acta Obstet. Gynecol. Scand.* **2012**, *91*, 1306–1313. [\[CrossRef\]](#) [\[PubMed\]](#)

62. Euliano, T.Y.; Darmanjian, S.; Nguyen, M.T.; Busowski, J.D.; Euliano, N.; Gregg, A.R. Monitoring Fetal Heart Rate during Labor: A Comparison of Three Methods. *J. Pregnancy* **2017**, *2017*, 8529816. [\[CrossRef\]](#)
63. Declercq, E.R.; Sakala, C.; Corry, M.P.; Applebaum, S.; Herrlich, A. Major Survey Findings of Listening to MothersSM III: Pregnancy and Birth. *J. Perinat. Educ.* **2014**, *23*, 9–16. [\[CrossRef\]](#)
64. Wood, P.L.; Dobbie, H.G. Equipment. In *Electronic Fetal Heart Rate Monitoring: A Practical Guide*; Wood, P.L., Dobbie, H.G., Eds.; Macmillan Education: London, UK, 1989; pp. 6–11. [\[CrossRef\]](#)
65. Neilson, D.R.; Freeman, R.K.; Mangan, S. Signal ambiguity resulting in unexpected outcome with external fetal heart rate monitoring. *Am. J. Obstet. Gynecol.* **2008**, *198*, 717–724. [\[CrossRef\]](#)
66. Behar, J.; Andreotti, F.; Zaunseder, S.; Oster, J.; Clifford, G.D. A practical guide to non-invasive foetal electrocardiogram extraction and analysis. *Physiol. Meas.* **2016**, *37*, R1–R35. [\[CrossRef\]](#)
67. Agostinelli, A.; Grillo, M.; Biagini, A.; Giuliani, C.; Burattini, L.; Fioretti, S.; Di Nardo, F.; Giannubilo, S.R.; Ciavattini, A.; Burattini, L. Noninvasive Fetal Electrocardiography: An Overview of the Signal Electrophysiological Meaning, Recording Procedures, and Processing Techniques. *Ann. Noninvasive Electrocardiol. Off. J. Int. Soc. Holter Noninvasive Electrocardiol. Inc.* **2015**, *20*, 303–313. [\[CrossRef\]](#)
68. Keenan, E.; Karmakar, C.; Brownfoot, F.C.; Palaniswami, M. Personalised anatomic modeling for non-invasive fetal ECG: methodology and applications. *Eur. Trans. Instrum. Meas.* **2021**, *70*, 4004312. [\[CrossRef\]](#)
69. Keenan, E.; Karmakar, C.K.; Palaniswami, M. The effects of asymmetric volume conductor modeling on non-invasive fetal ECG extraction. *Physiol. Meas.* **2018**, *39*, 105013. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Bhogal, K. Focus on cardiotocography: Intrapartum monitoring of uterine contractions. *Br. J. Midwifery* **2017**, *25*, 491–497. [\[CrossRef\]](#)
71. Euliano, T.Y.; Nguyen, M.T.; Marossero, D.; Edwards, R.K. Monitoring Contractions in Obese Parturients: Electrohysterography Compared With Traditional Monitoring. *Obstet. Gynecol.* **2007**, *109*, 1136–1140. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Wilmsink, F.A.; Wilms, F.F.; Heydanus, R.; Mol, B.W.J.; Papatsonis, D.N.M. Fetal complications after placement of an intrauterine pressure catheter: A report of two cases and review of the literature. *J. Matern.-Fetal Neonatal Med.* **2008**, *21*, 880–883. [\[CrossRef\]](#)
73. Cohen, W.R. Clinical assessment of uterine contractions. *Int. J. Gynecol. Obstet.* **2017**, *139*, 137–142. [\[CrossRef\]](#)
74. Ayres-de Campos, D.; Bernardes, J.; Garrido, A.; Marques-de Sá, J.; Pereira-Leite, L. Sisporto 2.0: A program for automated analysis of cardiotocograms. *J. Matern.-Fetal Med.* **2000**, *9*, 311–318. [\[CrossRef\]](#)
75. Ayres-de Campos, D.; Sousa, P.; Costa, A.; Bernardes, J. Omniview-SisPorto® 3.5—A central fetal monitoring station with online alerts based on computerised cardiotocogram+ST event analysis. *J. Perinat. Med.* **2008**, *36*, 260–264. [\[CrossRef\]](#)
76. Ayres-de Campos, D.; Rei, M.; Nunes, I.; Sousa, P.; Bernardes, J. SisPorto 4.0—Computer analysis following the 2015 FIGO Guidelines for intrapartum fetal monitoring. *J. Matern.-Fetal Neonatal Med.* **2017**, *30*, 62–67. [\[CrossRef\]](#)
77. Nunes, I.; Ayres-de Campos, D.; Ugwumadu, A.; Amin, P.; Banfield, P.; Nicoll, A.; Cunningham, S.; Sousa, P.; Costa-Santos, C.; Bernardes, J.; et al. Central Fetal Monitoring With and Without Computer Analysis: A Randomised Controlled Trial. *Obstet. Gynecol.* **2017**, *129*, 83. [\[CrossRef\]](#)
78. Balayla, J.; Shrem, G. Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: A systematic review and meta-analysis. *Arch. Gynecol. Obstet.* **2019**, *300*, 7–14. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Campanile, M.; D'Alessandro, P.; Della Corte, L.; Saccone, G.; Tagliaferri, S.; Arduino, B.; Esposito, G.; Esposito, F.G.; Raffone, A.; Signorini, M.G.; et al. Intrapartum cardiotocography with and without computer analysis: A systematic review and meta-analysis of randomised controlled trials. *J. Matern.-Fetal Neonatal Med.* **2020**, *33*, 2284–2290.
80. Campos, D.; Bernardes, J. Cardiotocography. UCI Machine Learning Repository. 2010. Available online: <https://archive.ics.uci.edu/dataset/193/cardiotocography> (accessed on 10 July 2023). [\[CrossRef\]](#)
81. Chudáček, V.; Spilka, J.; Burša, M.; Janků, P.; Hruban, L.; Huptych, M.; Lhotská, L. Open access intrapartum CTG database. *BMC Pregnancy Childbirth* **2014**, *14*, 16. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Spencer, J.A.; Belcher, R.; Dawes, G.S. The influence of signal loss on the comparison between computer analyses of the fetal heart rate in labour using pulsed Doppler ultrasound (with autocorrelation) and simultaneous scalp electrocardiogram. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **1987**, *25*, 29–34. [\[CrossRef\]](#)
83. Jezewski, J.; Wrobel, J.; Matonia, A.; Horoba, K.; Martinek, R.; Kupka, T.; Jezewski, M. Is Abdominal Fetal Electrocardiography an Alternative to Doppler Ultrasound for FHR Variability Evaluation? *Front. Physiol.* **2017**, *8*, 305. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Gudivada, V.; Apon, A.; Ding, J. Data Quality Considerations for Big Data and Machine Learning: Going Beyond Data Cleaning and Transformations. *Int. J. Adv. Softw.* **2017**, *10*, 1–20.
85. Chudáček, V.; Huptych, M.; Koucký, M.; Spilka, J.; Bauer, L.; Lhotská, L. Fetal heart rate data pre-processing and annotation. In Proceedings of the 2009 9th International Conference on Information Technology and Applications in Biomedicine, Larnaka, Cyprus, 4–7 November 2009; pp. 1–4. ISSN: 2168–2208. [\[CrossRef\]](#)
86. Zhao, Z.; Deng, Y.; Zhang, Y.; Zhang, Y.; Zhang, X.; Shao, L. DeepFHR: Intelligent prediction of fetal Acidemia using fetal heart rate signals based on convolutional neural network. *BMC Med. Inform. Decis. Mak.* **2019**, *19*, 286. [\[CrossRef\]](#)
87. Petrozziello, A.; Jordanov, I.; Aris Papageorgiou, T.; Christopher Redman, W.G.; Georgieva, A. Deep Learning for Continuous Electronic Fetal Monitoring in Labor. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Honolulu, HI, USA, 18–21 July 2018; Volume 2018, pp. 5866–5869. [\[CrossRef\]](#)

88. Cömert, Z.; Yang, Z.; Velappan, S.; Boopathi, A.M.; Kocamaz, A.F. Performance evaluation of Empirical Mode Decomposition and Discrete Wavelet Transform for computerised hypoxia detection and prediction. In Proceedings of the 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, Turkey, 2–5 May 2018; pp. 1–4. [\[CrossRef\]](#)
89. Liang, S.; Li, Q. Automatic Evaluation of Fetal Heart Rate Based on Deep Learning. In Proceedings of the 2021 2nd Information Communication Technologies Conference (ICTC), Nanjing, China, 7–9 May 2021; pp. 235–240. [\[CrossRef\]](#)
90. Krupa, B.N.; Ali, M.A.M.; Zahedi, E. The application of empirical mode decomposition for the enhancement of cardiotocograph signals. *Physiol. Meas.* **2009**, *30*, 729–743. [\[CrossRef\]](#)
91. Oikonomou, V.P.; Spilka, J.; Stylios, C.; Lhostka, L. An adaptive method for the recovery of missing samples from FHR time series. In Proceedings of the Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems, Porto, Portugal, 20–22 June 2013; pp. 337–342. ISSN: 1063–7125. [\[CrossRef\]](#)
92. Fuentealba, P.; Illanes, A.; Ortmeier, F. Cardiotocograph Data Classification Improvement by Using Empirical Mode Decomposition*. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 5646–5649, ISSN: 1558–4615. [\[CrossRef\]](#)
93. O’Sullivan, M.; Gabruseva, T.; Boylan, G.; O’Riordan, M.; Lightbody, G.; Marnane, W. Classification of fetal compromise during labour: Signal processing and feature engineering of the cardiotocograph. In Proceedings of the 2021 29th European Signal Processing Conference (EUSIPCO), Dublin, Ireland, 23–27 August 2021; pp. 1331–1335. ISSN: 2076–1465. [\[CrossRef\]](#)
94. Lear, C.A.; Galinsky, R.; Wassink, G.; Yamaguchi, K.; Davidson, J.O.; Westgate, J.A.; Bennet, L.; Gunn, A.J. The myths and physiology surrounding intrapartum decelerations: The critical role of the peripheral chemoreflex. *J. Physiol.* **2016**, *594*, 4711–4725. [\[CrossRef\]](#)
95. Chudáček, V.; Spilka, J.; Janků, P.; Koucký, M.; Lhotská, L.; Huptych, M. Automatic evaluation of intrapartum fetal heart rate recordings: A comprehensive analysis of useful features. *Physiol. Meas.* **2011**, *32*, 1347–1360. [\[CrossRef\]](#)
96. Cazares, S.; Moulden, M.; Redman, C.W.G.; Tarassenko, L. Tracking poles with an autoregressive model: A confidence index for the analysis of the intrapartum cardiotocogram. *Med. Eng. Phys.* **2001**, *23*, 603–614. [\[CrossRef\]](#) [\[PubMed\]](#)
97. Georgieva, A.; Payne, S.J.; Moulden, M.; Redman, C.W.G. Artificial neural networks applied to fetal monitoring in labour. *Neural Comput. Appl.* **2013**, *22*, 85–93. [\[CrossRef\]](#)
98. Spilka, J.; Frecon, J.; Leonarduzzi, R.; Pustelnik, N.; Abry, P.; Doret, M. Sparse Support Vector Machine for Intrapartum Fetal Heart Rate Classification. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 664–671. [\[CrossRef\]](#)
99. Furuya, N.; Hasegawa, J.; Imai, H.; Homma, C.; Kurasaki, A.; Kondo, H.; Suzuki, N. Accuracy of predicting neonatal distress using a five-level classification of fetal heart rate monitoring. *J. Obstet. Gynaecol. Res.* **2020**, *47*, 254–261. [\[CrossRef\]](#)
100. Georgoulas, G.; Stylios, C.; Nokas, G.; Groumpos, P. Classification of fetal heart rate during labour using hidden Markov models. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), Budapest, Hungary, 25–29 July 2004; Volume 3, pp. 2471–2475, ISSN: 1098–7576. [\[CrossRef\]](#)
101. Signorini, M.; Magenes, G.; Cerutti, S.; Arduini, D. Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 365–374. [\[CrossRef\]](#)
102. Gonçalves, H.; Bernardes, J.; Paula Rocha, A.; Ayres-de Campos, D. Linear and nonlinear analysis of heart rate patterns associated with fetal behavioral states in the antepartum period. *Early Hum. Dev.* **2007**, *83*, 585–591. [\[CrossRef\]](#) [\[PubMed\]](#)
103. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [\[CrossRef\]](#)
104. Ferrario, M.; Signorini, M.; Magenes, G. Complexity analysis of the fetal heart rate for the identification of pathology in fetuses. In Proceedings of the Computers in Cardiology, Lyon, France, 25–28 September 2005; pp. 989–992. ISSN: 2325–8853. [\[CrossRef\]](#)
105. Georgieva, A.; Papageorgiou, A.; Payne, S.; Moulden, M.; Redman, C. Phase-rectified signal averaging for intrapartum electronic fetal heart rate monitoring is related to acidemia at birth. *BJOG Int. J. Obstet. Gynaecol.* **2014**, *121*, 889–894. [\[CrossRef\]](#)
106. Spilka, J.; Chudáček, V.; Koucký, M.; Lhotská, L.; Huptych, M.; Janků, P.; Georgoulas, G.; Stylios, C. Using nonlinear features for fetal heart rate classification. *Biomed. Signal Process. Control* **2012**, *7*, 350–357. [\[CrossRef\]](#)
107. Georgoulas, G.; Stylios, D.; Groumpos, P. Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 875–884. [\[CrossRef\]](#)
108. Dash, S.; Quirk, J.G.; Djurić, P.M. Fetal Heart Rate Classification Using Generative Models. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 2796–2805. [\[CrossRef\]](#) [\[PubMed\]](#)
109. Krupa, N.; MA, M.A.; Zahedi, E.; Ahmed, S.; Hassan, F.M. Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine. *Biomed. Eng. Online* **2011**, *10*, 6. [\[CrossRef\]](#)
110. Afridi, R.; Iqbal, Z.; Khan, M.; Ahmad, A.; Naseem, R. Fetal Heart Rate Classification and Comparative Analysis Using Cardiotocography Data and Known Classifiers. *Int. J. Grid Distrib. Comput.* **2019**, *12*, 31–42. [\[CrossRef\]](#)
111. Cömert, Z.; Kocamaz, A. Evaluation of Fetal Distress Diagnosis during Delivery Stages based on Linear and Nonlinear Features of Fetal Heart Rate for Neural Network Community. *Int. J. Comput. Appl.* **2016**, *156*, 26–31. [\[CrossRef\]](#)
112. Cömert, Z.; Kocamaz, A. A Study Based on Gray Level Co-Occurrence Matrix and Neural Network Community for Determination of Hypoxic Fetuses. In Proceedings of the International Artificial Intelligence and Data Processing Symposium, Malatya, Turkey, 17–18 September 2016.
113. Cömert, Z.; Kocamaz, A. Comparison of Machine Learning Techniques for Fetal Heart Rate Classification. *Acta Phys. Pol. A* **2017**, *132*, 451–454. [\[CrossRef\]](#)

114. Georgoulas, G.; Stylios, C.; Groumpos, P. Classification of fetal heart rate using scale dependent features and support vector machines. *IFAC Proc. Vol.* **2005**, *38*, 313–318. [[CrossRef](#)]
115. Kiranyaz, S.; Avci, O.; Abdeljaber, O.; Ince, T.; Gabbouj, M.; Inman, D.J. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **2021**, *151*, 107398. [[CrossRef](#)]
116. Li, J.; Chen, Z.Z.; Huang, L.; Fang, M.; Li, B.; Fu, X.; Wang, H.; Zhao, Q. Automatic Classification of Fetal Heart Rate Based on Convolutional Neural Network. *IEEE Internet Things J.* **2019**, *6*, 1394–1401. [[CrossRef](#)]
117. Yefei, Z.; Yanjun, D.; Xiaohong, Z.; Lihuan, S.; Zhidong, Z. Bidirectional Long Short-term Memory-based Intelligent Auxiliary Diagnosis of Fetal Health. In Proceedings of the 2021 IEEE Region 10 Symposium (TENSYP), Jeju, Republic of Korea, 23–25 August 2021; pp. 1–5. ISSN: 2642–6102. [[CrossRef](#)]
118. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [[CrossRef](#)]
119. Georgieva, A.; Moulden, M.; Redman, C.W.G. Umbilical cord gases in relation to the neonatal condition: The EveREst plot. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **2013**, *168*, 155–160. [[CrossRef](#)]
120. Malin, G.L.; Morris, R.K.; Khan, K.S. Strength of association between umbilical cord pH and perinatal and long term outcomes: Systematic review and meta-analysis. *BMJ Br. Med. J.* **2010**, *340*, 1121–1121. [[CrossRef](#)]
121. Wen, J.; Thibeau-Sutre, E.; Diaz-Melo, M.; Samper-González, J.; Routier, A.; Bottani, S.; Dormont, D.; Durrleman, S.; Burgos, N.; Colliot, O. Convolutional neural networks for classification of Alzheimer’s disease: Overview and reproducible evaluation. *Med. Image Anal.* **2020**, *63*, 101694. [[CrossRef](#)] [[PubMed](#)]
122. Cheong-See, F.; Allotey, J.; Marlin, N.; Mol, B.; Schuit, E.; ter Riet, G.; Riley, R.; Moons, K.; Khan, K.; Thangaratinam, S. Prediction models in obstetrics: Understanding the treatment paradox and potential solutions to the threat it poses. *BJOG Int. J. Obstet. Gynaecol.* **2016**, *123*, 1060–1064. [[CrossRef](#)] [[PubMed](#)]
123. Romagnoli, S.; Sbrillini, A.; Burattini, L.; Marcantoni, I.; Morettini, M.; Burattini, L. Annotation dataset of the cardiotocographic recordings constituting the “CTU-CHB intra-partum CTG database”. *Data Brief* **2020**, *31*, 105690. [[CrossRef](#)] [[PubMed](#)]
124. Abry, P.; Spilka, J.; Leonarduzzi, R.; Chudáček, V.; Pustelnik, N.; Doret, M. Sparse learning for Intrapartum fetal heart rate analysis. *Biomed. Phys. Eng. Express* **2018**, *4*, 034002. [[CrossRef](#)]
125. Georgieva, A.; Abry, P.; Chudáček, V.; Djurić, P.M.; Frasch, M.G.; Kok, R.; Lear, C.A.; Lemmens, S.N.; Nunes, I.; Papageorgiou, A.T.; et al. Computer-based intrapartum fetal monitoring and beyond: A review of the 2nd Workshop on Signal Processing and Monitoring in Labor (October 2017, Oxford, UK). *Acta Obstet. Gynecol. Scand.* **2019**, *98*, 1207–1217. [[CrossRef](#)]
126. Georgieva, A.; Abry, P.; Nunes, I.; Frasch, M.G. Editorial: Fetal-maternal monitoring in the age of artificial intelligence and computer-aided decision support: A multidisciplinary perspective. *Front. Pediatr.* **2022**, *10*, 1007799. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.