IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 9, 2019, accepted October 24, 2019, date of publication October 31, 2019, date of current version November 13, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950798

Cardiotocographic Signal Feature Extraction Through CEEMDAN and Time-Varying Autoregressive Spectral-Based Analysis for Fetal Welfare Assessment

PATRICIO FUENTEALBA[1,3], (Student Member, IEEE), ALFREDO ILLANES[2], AND FRANK ORTMEIER[1]

[1]Faculty of Computer Science, Institute for Intelligent Cooperating Systems, Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany
[2]Faculty of Electrical Engineering and Information Technology, Institute of Medical Engineering, Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany
[3]Facultad de Ciencias de la Ingeniería, Instituto de Electricidad y Electrónica, Universidad Austral de Chile, Valdivia 5111187, Chile

Corresponding author: Patricio Fuentealba (patricio.fuentealba@ovgu.de)

This research was supported by the National Commission for Scientific and Technological Research CONICYT through the Chilean Scholarship Program for Graduate Studies.

ABSTRACT Cardiotocograph (CTG) is a widely used tool for fetal surveillance during labor, which provides the joint recording of fetal heart rate (FHR) and uterine contraction data. Unfortunately, the CTG interpretation is difficult because it involves a visual analysis of highly complex signals. Recent clinical research indicates that a correct CTG assessment requires a good understanding of the fetal compensatory mechanisms modulated by the autonomic nervous system. Certainly, this modulation reflects variations in the FHR, whose characteristics can involve significant information about the fetal condition. The main contribution of this work is to investigate these characteristics by a new approach combining two signal processing methods: the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and time-varying autoregressive (TV-AR) modeling. The idea is to study the CEEMDAN intrinsic mode functions (IMFs) in both the time-domain and the spectral-domain in order to extract information that can help to assess the fetal condition. For this purpose, first, the FHR signal is decomposed, and then for each IMF, the TV-AR spectrum is computed in order to study their spectral dynamics over time. In this paper, we first explain the foundations of our proposed features. Then, we evaluate their performance in CTG classification by using three machine learning classifiers. The proposed approach has been evaluated on real CTG data extracted from the CTU-UHB database. Results show that by using only conventional FHR features, the classification performance achieved 78, 0%. Then, by including the proposed CEEMDAN spectral-based features, it increased to 81, 7%.

INDEX TERMS Biomedical signal processing, cardiotocograph, empirical mode decomposition, fetal heart rate, spectral analysis, time-varying autoregressive modeling.

I. INTRODUCTION

The main aim of fetal surveillance during labor is to timely identify potential acidotic fetuses without unnecessary interventions. During this process, a fetus can repeatedly suffer from decreased oxygen insufficiency, which is a natural phenomenon, but fetuses with weakened defense mechanisms could develop metabolic acidosis. As a consequence, it can lead to neuro-development disability, cerebral palsy, or in some cases, even death [1]. For this reason, fetal monitoring during labor is essential, which is commonly performed by using a cardiotocograph (CTG), which provides the joint recording of fetal heart rate (FHR) and uterine contraction (UC) signals. The CTG assessment is currently performed by a visual analysis of several morphological FHR signal patterns based on proposed medical guidelines [2]–[4]. However, the CTG interpretation by this methodology has demonstrated a wide intra- and inter-observer disagreement, lack of objectivity, and poor interpretation reproducibility [1], [5], [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

In this context, several computer-based support systems have been proposed for assisting clinicians in CTG assessment. Unfortunately, there is no evidence that those systems have an effect on the incidence of newborn acidosis without excessive obstetrical intervention compared with the conventional CTG analysis [7]–[9]. That is why several approaches based on signal processing techniques have been proposed in order to extract hidden FHR characteristics that can help to distinguish between a normal and an acidotic fetal condition. However, the obtained results are not satisfactory enough for their use in clinical practice so far [5].

Current advances in clinical research indicate that correct identification of hypoxemia requires a good understanding of the fetal compensatory mechanisms, whose condition depends on how the fetus is compensating itself over time. These compensatory mechanisms are modulated by the autonomic nervous system (ANS), which prepares the fetus for intense activity after a perceived oxygen insufficiency [10], [11]. Certainly, this modulation reflects variations in the beat-to-beat FHR, whose time-varying dynamics can involve significant information about the fetal condition. Considering this phenomenon, conventional signal processing methods that do not integrate these physiological characteristics could not be appropriate for a correct CTG assessment, because they do not consider the nonlinear and non-stationary characteristics involved in the FHR as a result of the ANS modulation.

The main contribution of this work is to study those physiological characteristics in both the time-domain and the spectral-domain, based on the modulated FHR characteristics. The idea is to extract significant information from the FHR signal that helps to assess the fetal condition. For this purpose, we propose an innovative approach that combines two signal processing methods: the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and time-varying autoregressive (TV-AR) modeling. On the one hand, considering that the FHR dynamics of interest are a result of the ANS modulation, CEEMDAN can help to demodulate them by the direct extraction of the energy associated with the FHR signal oscillations. As a result, we obtain a finite number of intrinsic mode functions (IMFs), which allow extracting information about the FHR dynamics of interest in the time-domain. On the other hand, considering that each IMF is a function modulated in both amplitude and frequency, we propose to study their characteristics in the frequency domain by TV-AR modeling. The purpose of this spectral-based analysis is to extract more information about the studied phenomenon by characterizing the spectral dynamics involved in each IMF. The main objective of our proposed approach is to evaluate whether features based on CEEMDAN in combination with TV-AR modeling allow improving the CTG classification performance compared with traditional features proposed in the literature for CTG analysis.

Up to our knowledge, in other areas of engineering a previous version of CEEMDAN, empirical mode decomposition (EMD) [12], has been used together with the stationary version of the AR modeling for a fault diagnosis of a roller bearing [13], looseness identification for rotating machinery [14], or classification of Magnetic Resonance Brain Images [15]. Likewise, a hybrid EMD-AR modeling approach has been proposed for wave forecasting in the time-domain [16]. However, these approaches do not consider a time-varying analysis, i.e., they do not allow the study of the spectral dynamics involved in the IMFs over time as performed by the time-varying version of the AR modeling.

The main advantage of using the TV-AR modeling for the IMFs analysis is that now it is possible to track the IMFs dynamics progressively based on their spectral information over time. In summary, first, the FHR signal is decomposed into the IMFs. Then, for each IMF, we extract its spectral information by using TV-AR modeling. Finally, we compute signal features based on such information, whose classification performance is evaluated by using three machine learning classifiers: Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and k-Nearest Neighbor (k-NN). The hypothesis is that the proposed features improve the CTG classification performance compared with conventional FHR features.

The rest of the paper is organized as follows. Section II gives a brief overview of previous research related to processing techniques used for FHR signal feature extraction. Then, Section III presents the proposed methodology and details the used signal processing techniques. Continuing, Section IV presents the obtained results and discussion by a qualitative and quantitative analysis. Finally, Section V concludes this approach and discusses directions for future work.

II. STATE OF THE ART

In recent years, several approaches based on signal processing techniques have been proposed in order to extract information from the FHR signal [5]. These proposed approaches for signal feature extraction can be roughly divided into two main categories: stationary and time-variant techniques.

A. STATIONARY TECHNIQUES

The stationary techniques can be subdivided into three categories: time-domain, frequency-domain, and nonlinear features.

1) TIME-DOMAIN FEATURES

This first category usually consists of statistical indicators computed from short and long term analysis in order to characterize the standard morphological CTG patterns described on guidelines [17]–[22], or statistical models to reproduce the clinician's interpretation of such patterns [23].

2) FREQUENCY-DOMAIN FEATURES

The FHR signal involves different frequency contributions [24]–[26]: the average of the FHR (DC component); slow signal dynamics lying on a very low frequency (VLF) band (approximately $0 - 0.03$Hz) and presenting

nonlinear characteristics; a low frequency (LF) band ($0.03 - 0.15$Hz), mainly related to the stimuli of the neural sympathetic fetal system; fetal breathing lying on a high frequency (HF) component ($0.5 - 1.0$Hz); and also fetal movements and maternal breathing associated with a frequency movements (MF) band between 0.15 and 0.5Hz.

In order to analyze the contribution of those frequency bands, several approaches have been proposed. They usually depend on operations performed over different energy spectral components since it is assumed that differences in spectral bands can be correlated to the fetal condition [27]. Most of these methods are based on Fast Fourier Transform (FFT) [24], [28]–[33] and AR-based parametric modeling, which allows the extraction of quantitative spectral parameters [34]–[36].

3) NONLINEAR FEATURES

In this category, several methods have been proposed in order to connect different nonlinear FHR signal characteristics to the fetal condition. Mutual information (MI) has been used to design new efficient features for FHR analysis [37] and study the UC and FHR coupling [38]. Multiscale entropy (MSE) has been employed to analyze the complexity [39], [40] or regularity [41] of the FHR signal. A multivariate analysis based on linear and nonlinear features has been proposed for discrimination of normal and Intra Uterine Growth Restricted fetuses [42]. Likewise, EMD has been used for different purposes in CTG analysis, as it is explained in Section III-B5.

B. TIME-VARIANT TECHNIQUES

Several time-varying frequency approaches have been proposed, which are mainly based on Short Time Fourier Transform [43], quadratic time-frequency distributions [44], and TV-AR modeling [26], [45], [46]. Amplitude-frequency characterization based on continuous and discrete wavelet transform has been employed in order to consider the transient behavior of the UC [47]. Likewise, continuous wavelet transform has been used for a spectral-based analysis of FHR and UC signals [48].

However, the existing limitation of these approaches lies in the fact that most of them focus only on fetal reactivity, i.e., they analyze the FHR as a response to a UC stimulus, without considering the FHR dynamics as a result of the ANS modulation over time.

The main difference of this work compared with the CTG analysis already proposed in the literature is that it involves the study of those modulated FHR dynamics by the extraction of *modal-spectral* features. The proposed hypothesis is that the decomposition and subsequent spectral-based analysis of the IMFs allows a better understanding of the described physiological phenomenon. In other words, we postulate that a particular FHR dynamic observed in the time-domain, which results from the ANS modulation, reflects different spectral dynamics that can be related to the fetal condition. As a consequence, the analysis of such dynamics could improve the interpretation and subsequent classification of non-reassuring CTG recordings during labor.

III. METHODOLOGY

The main idea behind the proposed methodology is to investigate FHR signal dynamics resulting from the ANS modulation (signal dynamics of interest) and analyze their connection with the fetal health condition. As described above, the FHR signal involves highly complex characteristics mainly associated with nonlinearities as a result of the physiological regulation mechanisms contributing to the fetal cardiac activity modulated by the ANS [10], [49]. Under this concept, we postulate that if such characteristics are not previously decomposed or demodulated, the FHR signal can still involve high complexity dynamics, which are difficult to analyze over time. In order to study these characteristics, conventional signal processing methods no longer can be used because they cannot describe many processing conditions involved in this physiological phenomenon.

In this perspective, we propose a new approach for processing the FHR signal by combining two methods: CEEMDAN [50] and TV-AR modeling [51].

On the one hand, the CEEMDAN is an adaptive method that allows to decompose decomposing nonlinear and non-stationary time series into a finite number of IMFs. This method provides information about the signal dynamics in the time-domain based on local properties of the signal data itself by the direct extraction of the energy associated with signal oscillations. Considering that its decomposition operation has similar principles of signal demodulation in amplitude, this method can be appropriate for extracting information about the FHR dynamics modulated by the ANS activity. In consequence, the main advantage of CEEMDAN is its dependency on the data-driven mechanism, and it does not require a priori known basis as the case of other traditional methods such as wavelet and Fourier transform-based decomposition [12], [52]–[54].

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

The classical stationary AR modeling [51] can adequately describe the peaks of a narrow-band power spectrum [55] and requires only a fraction of the signal samples that are required by standard methods, such as the FFT, in order to obtain the same spectral resolution. It is a technique for time series analysis in which a mathematical model is fitted to a sampled signal. Therefore, AR modeling has several advantages for the FHR signal analysis since this parametric

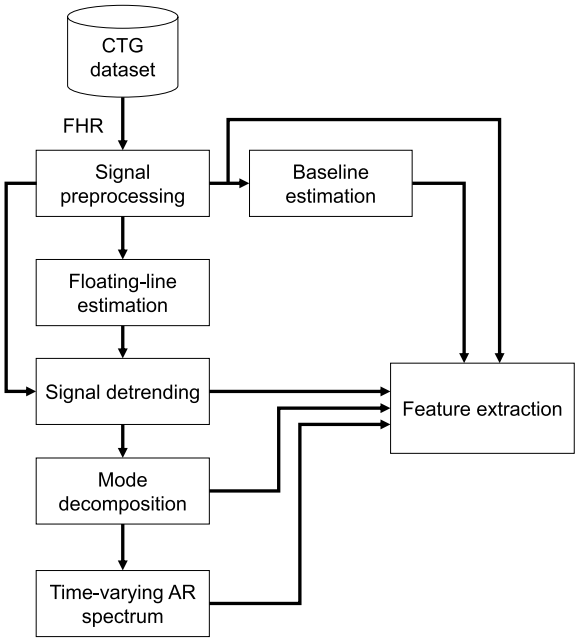

FIGURE 1. Feature extraction strategy.

modeling provides a signal description simpler to analyze by a few model parameters. Besides, it allows the extraction of quantitative spectral parameters versus time, which are better suited for quantitative spectral analysis [36].

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

In summary, for the proposed signal feature extraction operation, we first decompose the FHR signal by using the CEEMDAN method. Then, we compute the AR spectrum for each extracted IMF by using TV-AR modeling. After that, we extract a set of features based on the spectral energy behavior. Finally, as it is explained in Section IV, we evaluate their performance in CTG classification by using SVM, LDA, and *k*-NN as classifiers.

Fig. 1 shows the main steps of our strategy proposed for the FHR signal feature extraction (these steps are detailed below in Section III-B). Firstly, a preprocessing step involves both FHR signal outliers removal and signal interpolation procedures, whose resulting signal (preprocessed FHR signal) is used for the FHR baseline and the *floating-line* estimation. Then, a signal detrending step involves the subtraction operation between the preprocessed FHR signal and the floating-line. This resulting signal (detrended FHR signal) is decomposed into the IMFs using CEEMDAN, and subsequently, for each IMF, the TV-AR spectrum is computed.

Finally, the feature extraction operation (explained in Section IV) is based on the TV-AR spectra extracted from the IMFs. In addition, we propose to include another set of

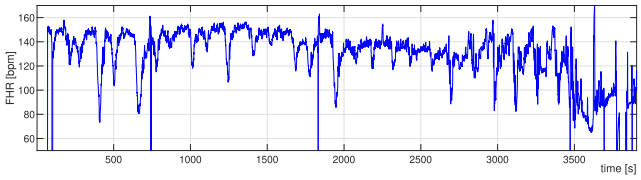

FIGURE 2. Raw FHR signal, recording no. 1022.

time-domain features generally used for CTG classification in order to compare their performance with each other.

A. CTG DATABASE

For the analysis and evaluation of the proposed methodology, the CTU-UHB Intrapartum Cardiotocography Database [56] is used. It is freely available on the PhysioNet website (*http://physionet.org/physiobank/database/ctu-uhb-ctgdb/*) and is the unique open-access database for research purposes on intrapartum CTG signal processing so far.

This database [56], from the Czech Technical University (CTU) in Prague and the University Hospital in Brno (UHB), contains 552 CTG recordings, which were selected from 9164 recordings collected between 2010 and 2012 at the UHB. Those 552 CTG recordings provide the FHR and the UC signals sampled at 4Hz, where 506 cases correspond to vaginal delivery and 46 to cesarean section. Besides, for each recording, the database includes different labor outcome parameters such as:

- Maternal data: age; parity; gravidity;
- Fetal outcome data: analysis of umbilical artery blood sample (i.e., pH; pCO2; pO2; base excess and computed BDecf) and Apgar score.

B. FHR SIGNAL PROCESSING STRATEGY

In this section, we present the proposed signal processing strategy and its principles, which are explained by using the FHR signal shown in Fig. 2. This signal belongs to the recording no. 1022, extracted from the CTU-UHB database.

1) FHR SIGNAL PREPROCESSING

The FHR signal is usually subject to different types of artifacts such as loss of data and outliers, mainly produced by the loss of sensor's contact that can temporarily interrupt the acquisition. For this reason, the artifact rejection method proposed in [57] is applied, which consists of two main steps: outliers removal and signal interpolation. In the outliers removal step, the signal values considered physiologically inconsistent in amplitude, i.e., outside the range between 50bpm and 210bpm, are removed from the signal. Then, in the signal interpolation step, loss of signal data corresponding to segments of length equal or less than 75s are interpolated by using a Hermite spline method. In Fig. 3(a), we can observe the preprocessed FHR signal.

It is important to note that the preprocessed FHR signal is used only for the FHR baseline and floating-line estimation. The subsequent processing operations are performed by using the raw FHR signal only after the outliers removal step.

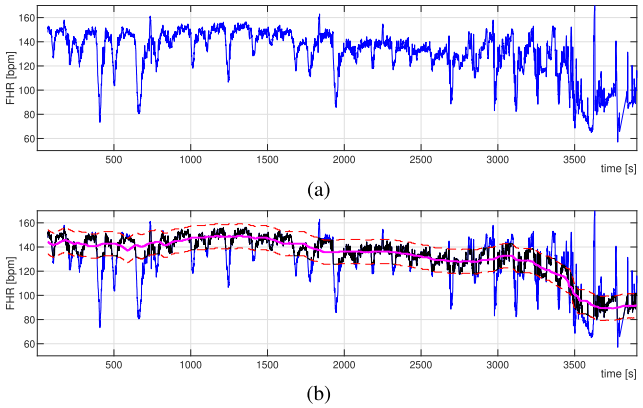

FIGURE 3. (a) Preprocessed FHR signal from Fig. 2; (b) FHR signal (blue), the FHR_{LH} (black), L and H traces (red) and PBL (magenta).

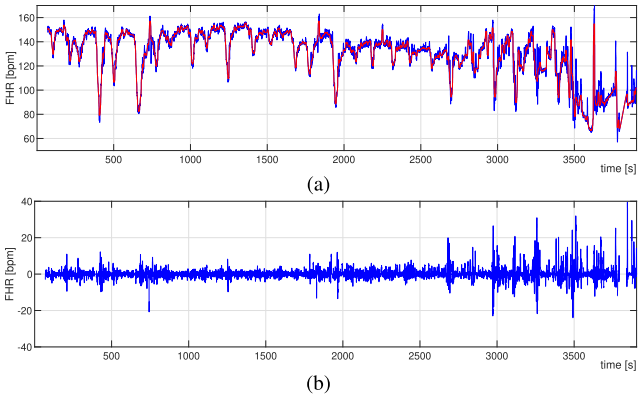

FIGURE 4. (a) Floating-line (red); (b) detrended FHR signal.

2) PROGRESSIVE FHR BASELINE ESTIMATION

According to CTG guidelines [2], the FHR baseline (BL) is considered as the mean level of the most horizontal and less oscillatory FHR segments in absence of accelerations or decelerations. In the conventional CTG analysis, the BL is particularly important because the assessment of other time-domain patterns rely on it [17]. Several algorithms have been proposed for the BL estimation [20], [58]; however, considering that there is no precise definition of how to identify it, they cannot guarantee a correct baseline assessment.

In this work, we propose to estimate a progressive BL (PBL). For this purpose, first, a virtual BL (VBL) is extracted by filtering the FHR signal using a nonlinear median filter [59]. Based on [22], this filter is computed over a sliding window of 400s length. Then, this VBL is used to define the low (L) and high (H) traces (see Fig. 3(b) red lines), which delimit the range of interest that will be considered for the PBL computation. The L and H traces are represented by:

$$L(n) = \mathrm{VBL}(n) - \triangle\mathrm{FHR} \tag{1}$$

$$H(n) = \mathrm{VBL}(n) + \triangle\mathrm{FHR} \tag{2}$$

where n is the sample number and $\triangle$FHR is set to 10bpm following [17].

Then, the FHR data of interest (FHR_{LH}) is described by:

$$\mathrm{FHR}_{LH}(n) = \begin{cases} H(n) & \mathrm{FHR}(n) > H(n) \\ \mathrm{FHR}(n) & L(n) \leq \mathrm{FHR}(n) \leq H(n) \\ L(n) & \mathrm{FHR}(n) < L(n) \end{cases} \tag{3}$$

whose graphic representation is shown in Fig. 3(b) in black.

Finally, the PBL is computed considering only the FHR_{LH} data by the same nonlinear median filter used for the VBL extraction. The PBL is plotted in Fig. 3(b) in magenta color.

3) FLOATING-LINE COMPUTATION

As presented in Section II-A2, the FHR dynamics related to the sympathetic ANS modulation (information of interest) lies in frequencies higher than 0.03Hz. For this reason, before the signal decomposition operation, the VLF trend ($0-0.03$Hz) is attenuated. This VLF trend corresponds to the morphological behavior of FHR decelerations and accelerations and involves nonlinear characteristics [5], [53]. Following [60], in order to track these characteristics, we calculate a floating-line. This floating-line is extracted by filtering the FHR signal using the same filter applied in Section III-B2 but with a different sliding window length, whose size was set to 10s, which was determined as follows:

1) We randomly chose a set of ten FHR signals from the CTU-UHB database.
2) Each FHR signal was filtered by the employed nonlinear filter using different window lengths between the range of 6 and 12s.
3) The extracted preliminary floating-lines (7 for each signal) were superimposed on the corresponding FHR signal in order to examine which one tracks better the morphological characteristics of decelerations.
4) After a visual analysis, we selected the floating-line computed by a sliding window of 10s length.

Fig. 4(a) shows the FHR signal (blue) and the corresponding floating-line (red).

4) FHR SIGNAL DETRENDING

For the subsequent modal and spectral analysis, the signal trend described by the floating-line is extracted from the FHR signal. This detrended FHR signal is computed by the subtraction operation between the FHR signal and the floating-line. Fig. 4(b) shows the detrended FHR signal.

5) FHR SIGNAL DECOMPOSITION

As explained above, for the FHR signal decomposition, the CEEMDAN method is used. The previous version of this technique, empirical mode decomposition (EMD), has been used to analyze data in different fields of biomedical engineering [52]. Particularly, in CTG signal analysis EMD has been utilized for FHR estimation from Doppler Ultrasound signals by using EMD-kurtosis method [61], analysis of the FHR signal components in order test the reliability of the EMD performance by using simulated FHR signals [62],

FHR baseline estimation with analysis of fetal movements by using EMD and Kohonen neural network [63], assessment of the high frequency information of FHR in different conditions of fetal activity [64], enhancement of the CTG signal quality by reducing signal artifacts [53], and FHR signal feature extraction and classification by using EMD and SVM [52], [65]. These works have concluded that EMD could be suited for the extraction and analysis of the FHR signal components in the time-domain.

Since EMD was originally proposed [12], it has evolved and developed in order to overcome different problems associated with signal decomposition [50], [66], [67]. The main drawback of the original EMD method is the *mode mixing*, i.e., more than one mode of oscillation may contribute to one IMF, or one mode can spread across different IMFs. In consequence, this phenomenon can provide an incorrect IMFs extraction, which can lead to an unreliable analysis of the FHR signal. In order to solve this problem, in [50], the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method has been proposed, which solves the mode mixing problem and subsequent drawbacks such as residual noise and *spurious* modes generated by the operation itself. Therefore, in this work and following [65], the FHR signal decomposition is performed by using the CEEMDAN, whose IMFs are computed by using a noise standard deviation (Nstd) set to 0.03, and both the number realizations (NR) and the maximum number of sifting iterations (NI) are set to 50. For a more in-depth explanation of this technique, please refer to [50].

In the next step, the IMFs are analyzed individually, and the TV-AR spectrum is computed for each IMF.

6) TIME-VARYING AR SPECTRUM ESTIMATION

As explained at the beginning of Section III, the classical stationary AR modeling analysis has been studied in different approaches, and it shows several advantages compared with non-parametrical spectral-based methods [68].

Considering that the IMFs extracted from the FHR signal involve characteristics strongly variant in time, the classical AR modeling method [51] is no longer suited for their analysis. That is why, in this work, we use the TV-AR modeling, which can be appropriate for the IMFs analysis since its parameters are now time-dependent.

An AR model assumes that the current signal sample $y[n]$ at sample number n in a data sequence $y[1], y[2], \ldots, y[N]$, can be modeled as a linearly weighted sum of the p most recent sample values $y[n-1], y[n-2], \ldots, y[n-p]$ and a white zero-mean noise $e[n]$ of variance σ^2, and its time-varying parametric model can be represented by:

$$y[n] = -\sum_{k=1}^{p} a_k(n) y[n-k] + e[n] \tag{4}$$

where $a_k(n)\{k = 1, 2, ..., p\}$ are the time-dependent AR parameters, which correspond to a set of values a_k that is updated sample-by-sample n. This results in a time-varying AR model transfer function represented by:

$$H[z, n] = \frac{Y[z, n]}{E[z, n]} = \frac{1}{1 + \sum_{k=1}^{p} a_k(n) z^{-k}} \tag{5}$$

where $Y[z, n]$ and $E[z, n]$ are the time-dependent z-transforms of the time series $y[n]$ and $e[n]$, respectively. Then, the TV-AR spectrum can be expressed as:

$$S_{AR}[f, n] = \frac{1}{\left|1 + \sum_{k=1}^{p} a_k(n) e^{-j2\pi fk}\right|^2} \tag{6}$$

This last equation allows us to perform a time-variant spectral analysis since it not only depends on the frequency-domain but also on the time-domain.

In this work and following [45], the AR coefficients $a_k(n)$ were computed by using a recursive least squares algorithm with a forgetting factor (λ) set to 0.99, which is appropriate to consider not only the faster but also, the slower signal dynamics of the frequency band of interest.

In order to choose an appropriate AR model order p, we considered the characteristics of the IMFs and their expected spectral behavior. By definition, each IMF consists of a non-overlapping function, which is modulated in both amplitude and frequency [50]. Thus, we assume that in the spectral-domain, only one absolute maximum peak contains the information of interest, which needs to be represented in the AR spectrum updated at each sample n.

With this in mind, we analyzed a set of 30 IMFs (selected randomly from different FHR signals). For each IMF, the AR spectrum was computed by using different AR model orders in the range of 4^{th} and 10^{th}. Then, we examined which model order offers a better spectral representation of the frequency component of interest, taking into account that only one and marked component should be distinguished. After this visual analysis and following [69], we decided to use a 6^{th} AR model order p.

Fig. 5 shows a decomposition of the detrended FHR signal plotted in Fig. 4(b) by using the CEEMDAN method. Each row depicts one IMF together with its corresponding TV-AR spectrum at the left and right sides, respectively. In this figure, we can observe the dynamics of the IMFs in both the time-domain and the frequency-domain, which is represented by the TV-AR spectrum. In this spectral representation, the x-axis is the time axis in seconds and the y-axis is the frequency axis in Hz. The spectral energy is represented by a color map, where the blue and yellow colors are the lowest and highest energy level, respectively. In the graph, the energy values are normalized between 0 and 1 for each sample n for better visualization of the frequency dynamics. It is important to note that the proposed features are based on this spectral representation.

IV. RESULTS AND DISCUSSION

This section presents the obtained results of the proposed CTG signals analysis. First, a qualitative analysis explains

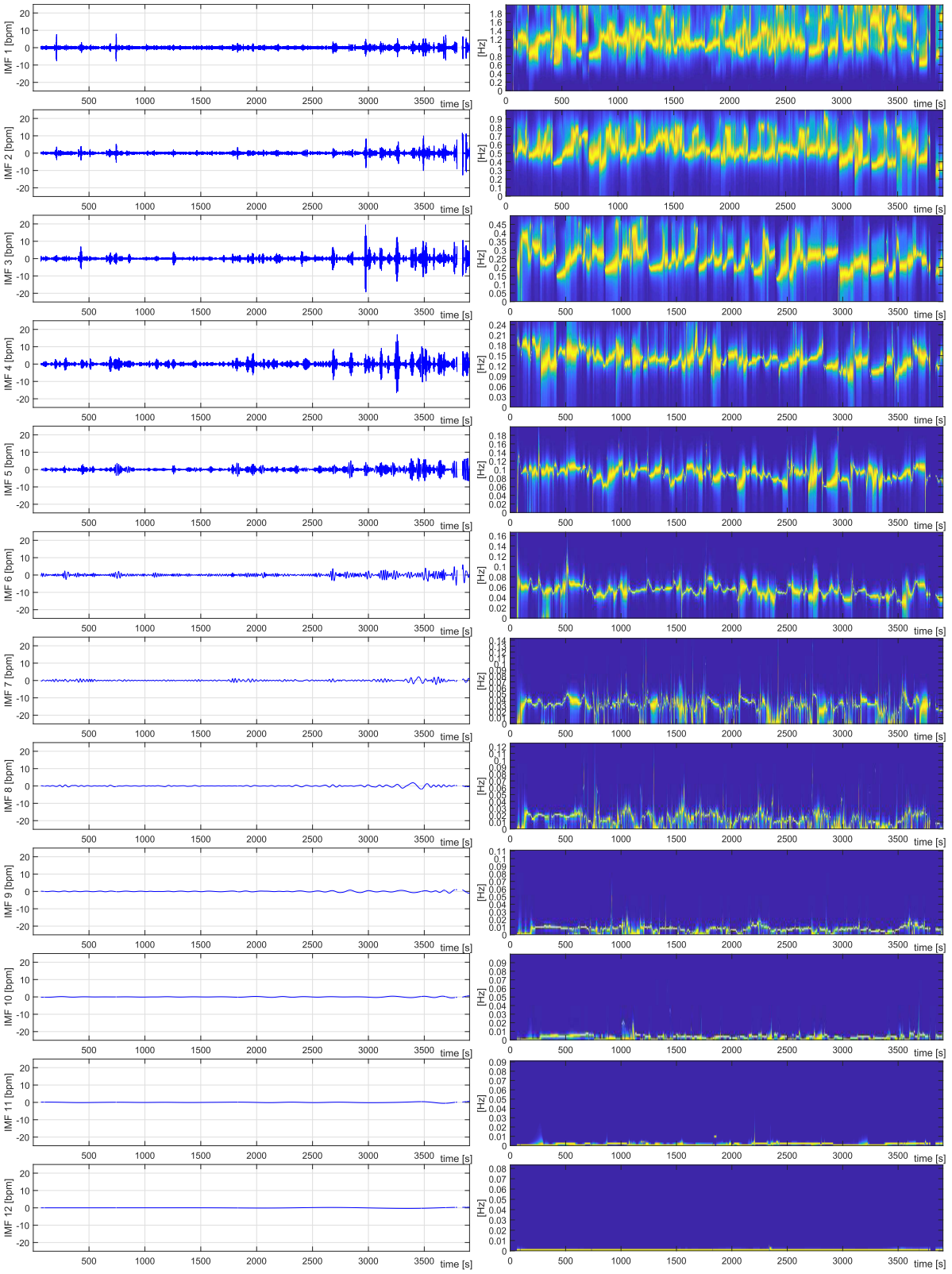

FIGURE 5. CEEMDAN of FHR recording no. 1022; (left) IMFs extracted from the detrended FHR signal presented in Fig. 4(b); (right) TV-AR spectrum of each IMF.

the foundations on which the feature extraction is based. Then, by a quantitative analysis, we evaluate the classification performance of the proposed features by using SVM, LDA, and k-NN as classifiers.

TABLE 1. Demographic information and outcome statistics of the selected dataset extracted from the CTU-UHB database.

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

A. DATASET SELECTION

The analysis was performed by using a dataset of CTG recordings extracted from the CTU-UHB database (see Section III-A).

In order to obtain comparable results with related studies (as presented in Section IV-D2), the class formation criteria for the dataset involves two groups: normal and acidotic. These two groups were selected according to the fetal outcome information, specifically pH and BDecf values. According to the literature [70], in umbilical cord test sampling at birth, values of pH $<$ 7.05 and BDecf $\geq$ 12 commonly indicate a fetal metabolic acidosis, whereas values of pH $>$ 7.20 and BDecf $<$ 12 indicate a normal fetal health condition. Under these criteria, 372 recordings were selected for the dataset; 354 labeled as examples of normal fetuses and 18 labeled as examples of acidotic fetuses. Table 1 shows the main demographic information and statistics concerning the labor outcome parameters of the selected dataset.

B. QUALITATIVE ANALYSIS

In this subsection, first, a visual analysis is performed by using two representative cases (see Fig. 6) extracted from the dataset. Then, the graphic information obtained by the analysis is evaluated in the complete dataset. For this qualitative analysis, we decided to study one of the IMFs that describe spectral dynamics inside the frequency band of interest. As explained in Section II-A2, this band lies in the frequency range $0.03 - 0.15$Hz, thereby after a visual evaluation of the spectral contribution of each IMF (see Fig. 5), we arbitrarily select the IMF 6.

It is important to note that in this qualitative analysis, only one selected IMF is studied. Nevertheless, the subsequent quantitative analysis (explained in Section IV-C) involves the examination of the complete band by considering the first ten IMFs.

Fig. 6 shows the selected representative cases, one for each column. The left column corresponds to an example of a normal fetal condition and the right column to a fetal metabolic acidosis, belonging to the recordings no. 1167 and 2011 of the CTU-UHB database, respectively. The first row shows the raw FHR signal for each case. The second row plots the preprocessed FHR signal and PBL in blue and magenta color, respectively. The third row shows the floating-line in red color. The fourth and fifth rows show the detrended FHR signal and the extracted IMF 6, respectively. The sixth row exhibits the TV-AR spectrum computed from the IMF 6, whose spectral energy values were normalized between 0 and 1 for each sample n for better visualization of the frequency dynamics (instead of components energies). The last row shows an energy indicator (E) computed from the TV-AR spectrum, which allows examining the spectral energy changes involved in the studied signal dynamics over time.

Our assumption is that the spectral energy represented by E describes different dynamics, whose behavior differs between a normal and an acidotic fetal condition. This indicator is calculated from the total frequency band ($0 - 2$Hz) of the AR spectrum for each sample n as described in eq.(7).

$$E[n] = \sum_{f=0}^{2Hz} (S_{AR}[f, n]) \tag{7}$$

Fig. 6 shows that in both cases, the FHR variations in the time-domain reflects important spectral dynamics over time, whose behavior differs between a normal and an acidotic fetal condition. Particularly, in the case of normal condition, the AR spectrum (Fig. 6(k)) exhibits several dynamical changes, represented by its E (Fig. 6(m)), which describes pronounced variations in amplitude over time. In contrast, the case corresponding to an acidotic fetus (right column) shows a completely different spectral behavior (see Fig. 6(n)). In this example, the spectral energy level is considerably lower compared with the first case. Likewise, the E does not show significant variations in amplitude, i.e., the AR spectrum exhibits a less marked response, whose behavior completely differs with respect to the normal case.

This phenomenon can be explained by the fact that the capacity of the fetal response, modulated by the sympathetic ANS, decreases for an acidotic fetus compared with a normal

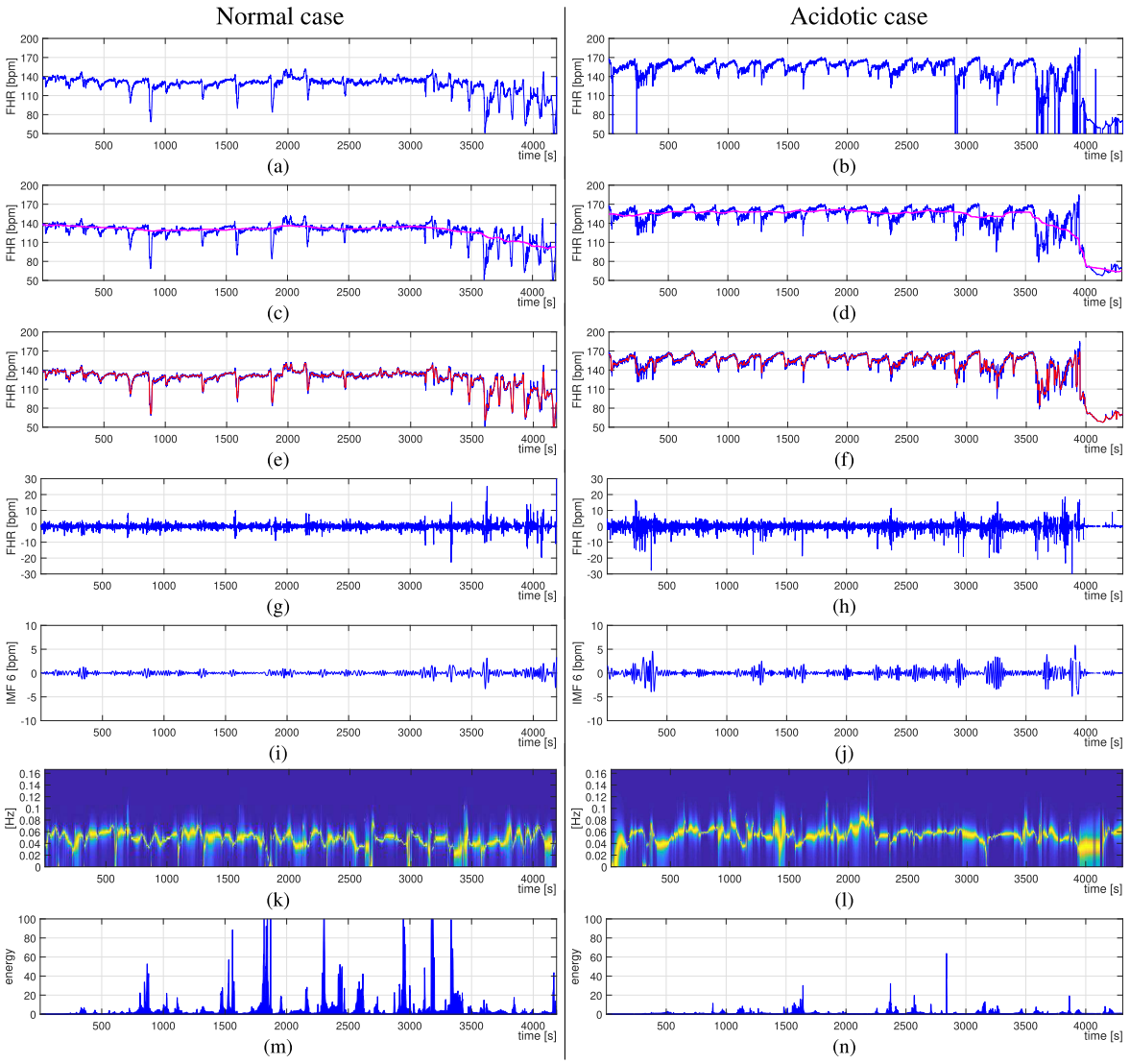

FIGURE 6. Representative cases; (a) raw FHR signal of recording no. 1167, pH = 7.36, BDecf = 5.03; (b) raw FHR signal of recording no. 2011, pH = 7.01, BDecf = 12.10; (c-d) preprocessed FHR signal (blue) and PBL (magenta); (e-f) floating-line (red); (g-h) detrended FHR signal from (e-f); (i-j) IMF 6; (k-l) TV-AR spectrum of the IMF 6; (m-n) E from (k-l).

fetus [45]. Therefore, the sympathetic path corresponding to an acidotic fetal condition might not reflect high activity in the FHR compared with the FHR activity of a normal condition.

In order to study if this phenomenon is reflected in the complete dataset, we compute the average of the spectral energy ($\overline{E}$) for every FHR signal. Then a Wilcoxon rank-sum test [71] was employed to evaluate if this feature shows a statistically significant difference between normal and acidotic cases. It was performed under the hypothesis that the median values of the $\overline{E}$ differ between normal and acidotic cases.

Table 2 present the obtained results of the statistical test, whose boxplots are exhibited in Fig. 7. Here, we can observe that the median values of $\overline{E}$ are 3.62 and 2.24 for the normal and acidotic cases, respectively.

As a result of the statistical test, the feature $\overline{E}$ was significantly higher (p-value < 0.01) for the group of normal fetuses compared with the acidotic fetuses, showing that the studied phenomenon is also reflected in the other cases included in the dataset.

TABLE 2. Results of the statistical test of the extracted spectral feature; values correspond to the $50^{th}[25^{th} - 75^{th}]$ percentiles of the data, respectively.

	Normal cases	Acidotic cases	Significance (p-value)
$\overline{E}$	$3.62[2.42 - 5.41]$	$2.24[2.07 - 3.24]$	0.0064

C. QUANTITATIVE ANALYSIS

1) FEATURES COMPUTATION

The qualitative analysis shows that it is possible to find significant spectral energy differences between normal and acidotic cases, which opens perspectives for the extraction

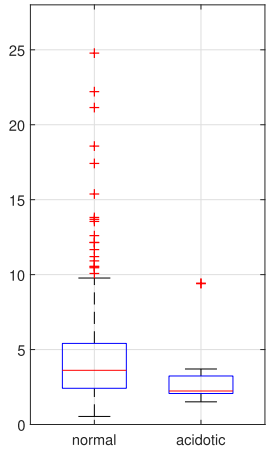

FIGURE 7. Boxplots for normal and acidotic cases of feature $\bar{E}$. The borders of the box are the 25^{th} and 75^{th} percentiles of the data. The red line represents the median of the data, and red crosses are the outliers. For the normal cases, the outlier of 80.1 is not plotted in the graph.

and subsequent classification of a larger set of features. The idea is to extract features based not only on the spectral energy but also on the frequency dynamics computed from the AR spectrum of each IMF.

For this operation, first, from the AR spectrum, we compute the traces of total spectral energy (E) (eq.7), the energy of the main component (E_{mc}), and frequency of the main component (f_{mc}) (see eq.8).

$$E_{mc}[n] = S_{AR}[f, n], \quad f = f_{mc} \tag{8}$$

where f_{mc} corresponds to the frequency value at which the AR spectrum presents the maximal energy level for each sample n.

Then, for each computed trace we calculate a set of seven statistical coefficients that have been commonly used in CTG analysis. The first five coefficients correspond to standard coefficients: arithmetic mean (mean), median, standard deviation (std), mean absolute deviation (mad), and root mean squares (RMS). The other two coefficients correspond to the sample entropy (SampEn) [72] and approximate entropy (ApEn) [73], which allow to quantify the complexity of the signal and have shown better performance in classification compared with the conventional CTG signal analysis [41], [74], [75]. Moreover, the entropy measures might be related to the interaction of the sympathetic and parasympathetic ANS response [76], which is the information of interest to be investigated in this work. Following [73], [77], [78], for the SampEn and ApEn coefficients computation, we use an embedding dimension $m = 2$ and a tolerance $r = 0.2 \times \sigma$, where σ is the standard deviation.

As described above, we have selected only features that have been classically used in the CTG analysis in order to obtain comparable results with related works proposed in the literature.

It is important to note that the number of decomposed IMFs of the 372 signals of the dataset varies from 12 to 17. Nevertheless, after a visual analysis of their spectral information, we decided to consider only the first ten IMFs, because they involve all of the frequency band of interest for the analysis of the FHR signal (see Section II-A2).

2) SELECTION OF AN INFORMATIVE FHR SIGNAL SEGMENT

Due to the FHR signal quality and considering that the fetal condition can change in the course of labor, it is usually proposed in the literature to extract an informative segment as close as possible to the delivery where features can be computed (also called as *epoch*). However, there is no precise definition about an optimal epoch length for the FHR signal analysis, because it depends mainly on the type of analysis (e.g., time-invariant, time-variant, short term, or long term analysis) [5].

In this work, for the evaluation of our extracted features, we propose to estimate an optimal epoch length before delivery between the range of 20 and 60 minutes in steps of 5 minutes. The lower limit of this range (20 minutes) was selected according to the literature [38], [79], and the higher limit was set to 60 minutes, which corresponds to the length of the shortest CTG recording available in the database. For each studied epoch, we calculated the number of statistically significant features (p-value < 0.05), and we selected the epoch providing the highest amount of statistically significant features. Considering our class formation criteria involve two groups (normal and acidotic), and following [80]–[82], a Wilcoxon rank-sum test [71] was used as an appropriate method for the evaluation of our features. Its main advantage is that it does not require the data normality assumption, as in the case of a t-test [83].

In summary, the length of an optimal epoch for our analysis was determined as follows:

1) We compute the proposed features from different FHR segment lengths before delivery between the range of 20 and 60 minutes in steps of 5 minutes (nine sets of features in complete).
2) For each extracted feature, we apply a Wilcoxon rank-sum test [71] under the hypothesis that the median values of the classes significantly differ from each other.
3) For each set of features, we calculate the amount of statistically significant features (p-value < 0.05).
4) The segment length corresponding to the set of features with the highest amount of statistically significant features is chosen as the optimal epoch.

The results obtained from the analysis described above is presented in Fig. 8. Here, we can clearly observe that the number of significant features depends on the length of the studied epoch. Particularly, for epochs in the range between 20 and 35 minutes length, this number is increasing from 21 to 30 significant features. Then, by considering a longer epoch length ($\geq$ 40 min), we obtain a lower number of significant features. According to this information, the optimal epoch corresponds to a segment of 35 minutes before delivery.

TABLE 3. Feature sets and corresponding significant features.

Feature set	Significant features		Type
1	IMF 1 - E	mean, median, RMS	Modal-spectral
2	IMF 1 - E_{mc}	mean, median, RMS	
3	IMF 2 - E_{mc}	ApEn, SampEn	
4	IMF 4 - E_{mc}	ApEn, SampEn	
5	IMF 6 - E	mean, std, mad, RMS	
6	IMF 6 - E_{mc}	mean, std, mad, RMS	
7	IMF 8 - E	mean, std, mad, RMS	
8	IMF 8 - E_{mc}	mean, std, mad, RMS	
9	IMF 10 - E	ApEn, SampEn	
10	IMF 10 - E_{mc}	ApEn SampEn	
11	Raw FHR signal	std, mad, ApEn, SampEn	Conventional
12	Progressive baseline	std, mad	
13	IMF 6 time-domain	ApEn	
14	IMF 7 time-domain	ApEn	

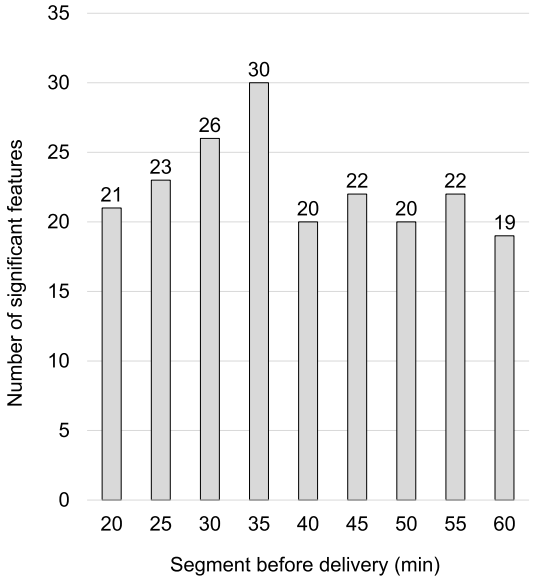

FIGURE 8. Selection of an optimal epoch.

D. EVALUATION METHODOLOGY

For the performance evaluation of the extracted features, we include another set of features already proposed in the literature in order to compare with each other. For this operation, we compute the same seven statistical coefficients described above but now applied to the raw FHR signal, PBL, detrended FHR signal, and IMFs in the time-domain. However, considering that the central tendency measures are not informative for the analysis of the detrended FHR signal and the IMFs, the arithmetic mean and median coefficients are not computed for such traces.

In the sequel, the features computed from the AR spectrum of each IMF are denoted as the *modal-spectral* features. In contrast, the features extracted from the raw FHR signal, PBL, detrended FHR signal, and the IMFs in the time-domain are denoted as the *conventional* features. In summary, the complete set of features consists of 210 modal-spectral and 69 conventional features. Once the features are extracted, the data are standardized using z-score based normalization. As a result, each feature has a mean value equal to zero and a standard deviation equal to one.

1) FEATURE ELIMINATION

In order to study the discriminant capability of the conventional features, as we performed above for the informative epoch selection, a Wilcoxon rank-sum test [71] was used. This test was applied independently for each feature, considering the hypothesis that the median of the acidotic group distribution significantly differs from the median of the normal group distribution. All the features that present a statistically significant difference (p-value < 0.05) were selected, and the others were excluded for the analysis. As a result, after the feature elimination step, the dataset consists of 38 features: 30 modal-spectral features and 8 conventional features, which are presented in Table 3. It is important to note that the significant modal-spectral features include only features computed from the energy of the spectral components (E or E_{mc}). In contrast, all the features computed from the frequency component (f_{mc}) were identified as non-significant by the applied test and therefore excluded. This information indicates that the frequency behavior of the IMFs does not provide significant information about the fetal condition; nevertheless, the energy associated with such frequency components can help to discriminate between a normal or an acidotic fetal status.

2) FEATURES EVALUATION

After the feature elimination step, following [84], the significant features are divided by their category, denoted in the sequel as the *feature sets*. As presented in Table 3, the modal-spectral features are divided into ten feature sets,

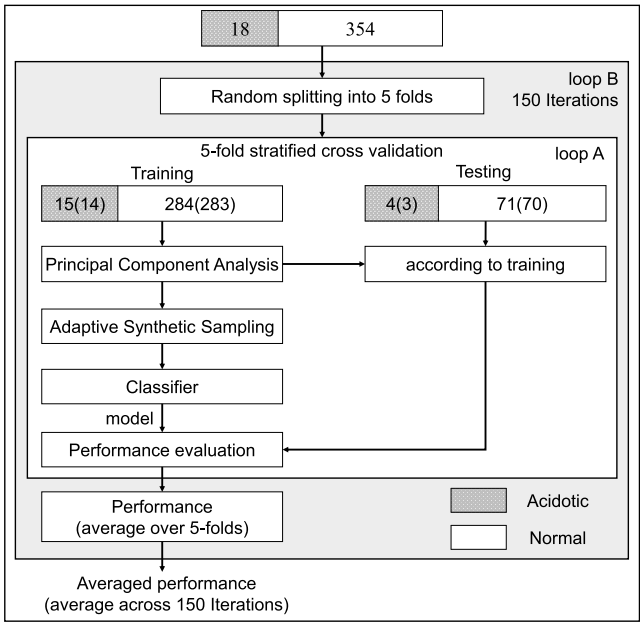

FIGURE 9. Proposed features performance evaluation strategy.

one for each IMF and for each trace computed from the AR spectrum (E and E_{mc}). Likewise, the conventional features are divided into four feature sets, one set for the Raw FHR signal, PBL, and each IMF in the time-domain, independently. As a result, the significant features are divided into 14 feature sets.

For the features performance evaluation, we proposed the strategy presented by the diagram of Fig. 9, which consists of two main loops (*loop A* and *loop B*). The loop A is based on 5-folds stratified cross-validation using a computer-based classifier. First, the features data are randomly split into five non-overlapping folds keeping the original proportion of normal and acidotic cases. Then, for each iteration, one different fold is used as a testing data subset, and the other remaining four folds are used as a training data subset. Considering that the studied features can convey redundant information to each other, for the training subset, we apply a feature preprocessing step based on principal component analysis (PCA). This method allows obtaining attributes that are less correlated to each other and preserving as much of the significant information as possible [84].

In order to attenuate the bias generated by the imbalanced input data to the classifier, we employ the adaptive synthetic sampling (ADASYN) technique computed by a factor of 19 using 5 k-neighbors. Its main advantage compared with conventional methods is that it can generate synthetic data for oversampling the imbalance class by considering the nearest data of the majority class. For a more in-depth explanation of the ADASYN method, please refer to [85].

It is important to note that the PCA is computed only from the training subset and then the testing subset is transformed accordingly (see Fig. 9). Likewise, ADASYN is applied only in the training subset, i.e., the testing subset does not include synthetic data, only real data.

a: MACHINE LEARNING CLASSIFIERS

For the evaluation of our proposed features, we selected three machine learning classifiers: SVM, LDA, and k-NN, which have been commonly used in CTG classification [52], [75], [77], [80], [84], [86]. It is important to note that the main objective of this approach is not the selection of the best classifier, but rather the performance evaluation of our proposed features based on the combination of CEEMDAN and TV-AR modeling. Therefore, these three classification methods, usually employed for discriminating two classes, can provide representative results that are comparable with related works that involve a similar type of classifiers.

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

- A Support Vector Machine classifier consists of a supervised learning model based on hyperplanes in the feature space, whose dimension depends on the number of features. It is based on dividing the space into two half subspaces, minimizing the empirical error and maximizing the margin between the nearest instances and the hyperplanes [87]. The employed SVM classifier was implemented by using a linear kernel function and a penalty factor for classification $C = 1$. For more detailed information about SVM please refer to [87].
- Linear Discriminate Analysis is a simple but powerful technique based on the difference between classes compared to a linear combination of the most significant features. This combination depends on the observations of predefined groups, which is used to predict the new observations. This widely used statistical method is based on a linear projection of the feature vector onto an optimal vector, whose classification is performed in this feature space. This operation allows minimizing the variance of each class while maximizing the class separation represented by the mean of each class. A more detailed explanation of the statistical operation involved in this technique can be found in [88].
- The k-Nearest Neighbor is a nonparametrical method, whose classification technique is based on the k closest points or set of points in the feature space. The k-NN uses distance metrics, usually Euclidean distance, to find the optimal *neighborhood* of attributes in relation to the class labels of the training data. In this work, in order to determine an appropriate number of k nearest neighbors, we experiment with different values of k: 1, 3, 5, 7, 9, and 11. As a result, we found that the highest classification performance was achieved by using 7-nearest neighbors. For a detailed explanation of the k-NN classifier, please refer to [89].

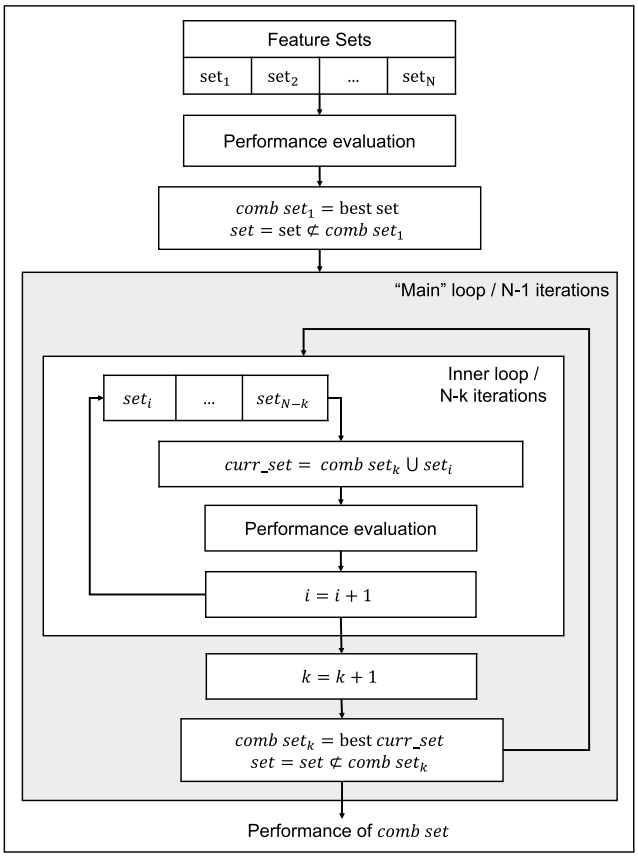

FIGURE 10. Combined sets performance evaluation strategy.

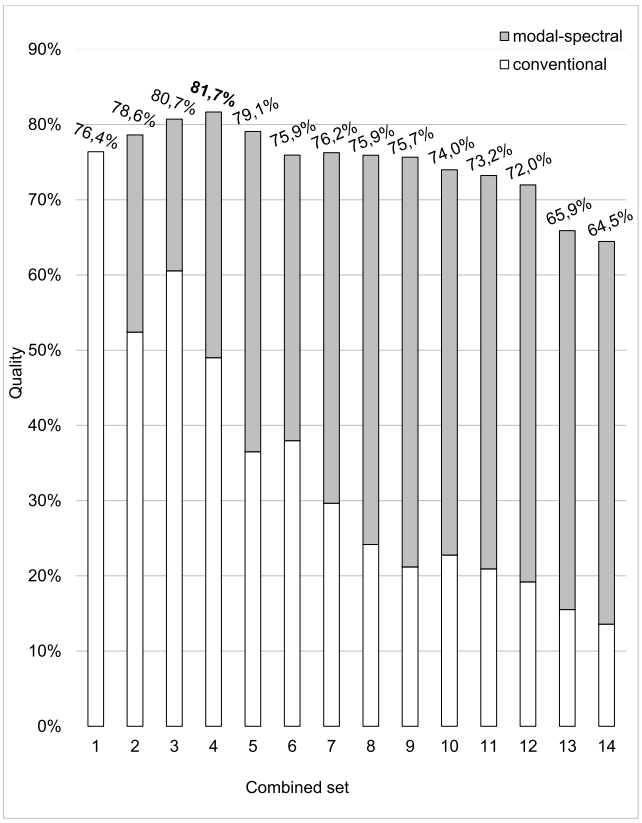

FIGURE 11. Performance of combined sets by using both modal-spectral and conventional features and the SVM classifier.

b: PERFORMANCE EVALUATION

As a measure of classification performance, we employ the geometric mean (9), since it has been considered as an appropriate quality metric (QI) used in CTG classification [84].

$$QI = \sqrt{Se \cdot Sp} \tag{9}$$

This statistical metric is based on the statistical metrics of sensitivity (Se) and specificity (Sp), obtained from the classification step, and defined in (10) and (11), respectively. In those equations, TP corresponds to the number of true predicted as acidotic cases, and FN to the false predicted as normal cases. Analogously, TN corresponds to the number of true predicted as normal cases, and the FP to the false predicted as acidotic cases.

$$Se = \frac{TP}{TP + FN} \tag{10}$$

$$Sp = \frac{TN}{TN + FP} \tag{11}$$

The main idea is to find a combination of features sets (denoted as the *combined set*) showing the best classification performance. Moreover, we examine the feature sets required for achieving such performance to study their contribution according to their category. For this operation, we propose the strategy described in Fig. 10.

In Fig. 10, N corresponds to the number of features sets to be tested (14 features sets), and the performance evaluation block represents a function that returns the set from the input sets that achieves the best classification performance. This function is based on the performance evaluation strategy described in Fig. 9. The proposed strategy first computes the classification performance of the 14 feature sets independently. Second, the feature set achieving the best performance is selected as the combined set 1 ($comb\ set_1$). Then, in the main loop, the $comb\ set_1$ is tested in combination with each remaining feature set independently, whose operation is described by the inner loop of N-k iterations, i.e., 13 sets ($curr\_sets$) to be tested. Now, the $curr\_set$ set achieving the best performance is selected as the $comb\ set_2$. Continuing, the $comb\ set_2$ is tested together with each remaining feature set independently (12 $curr\_sets$ to be tested) and the $curr\_set$ set achieving the best performance is selected as the $comb\ set_3$. After that, we repeat this process until we have tested all of the combined sets.

The results obtained from the previous operation are presented in Fig. 11, 12, and 13, which show the classification performance represented by the QI metric in percentage, for the SVM, LDA, and k-NN classifiers, respectively. Besides, for each classifier, Tables 4, 5, and 6 show in detail the features automatically selected and included in each combined set to achieve the highest classification performance. In this display of results, the contribution of the modal-spectral and

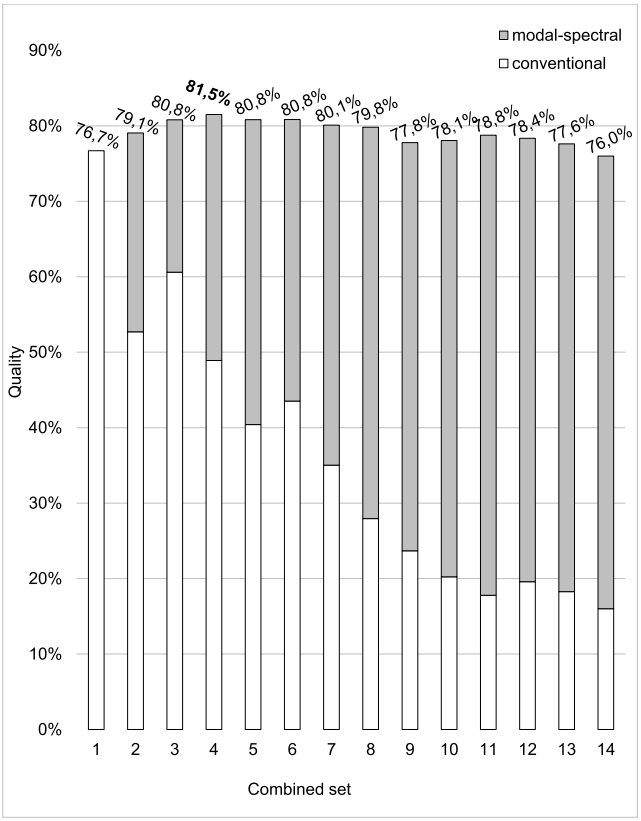

FIGURE 12. Performance of combined sets by using both modal-spectral and conventional features and the LDA classifier.

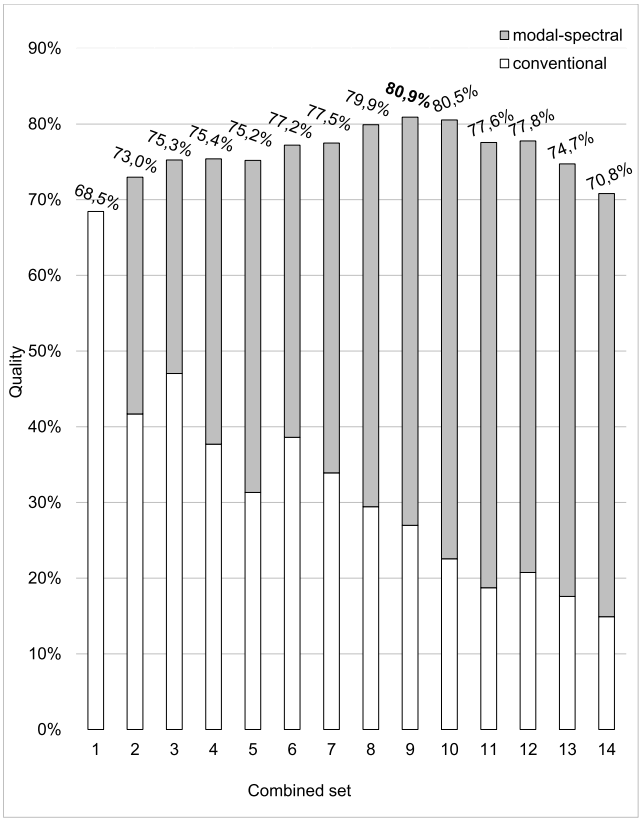

FIGURE 13. Performance of combined sets by using both modal-spectral and conventional features and the *k*-NN classifier.

TABLE 4. Features automatically selected to achieve the highest performance by using the SVM classifier.

Combined set k	Features			QI
1	raw FHR		std	$76,4\%$
	raw FHR		mad	
	raw FHR		ApEn	
	raw FHR		SampEn	
2	IMF 4	E_{mc}	ApEn	$78,6\%$
	IMF 4	E_{mc}	SampEn	
3	PBL		std	$80,7\%$
	PBL		mad	
4	IMF 10	E_{mc}	ApEn	$81,7\%$
	IMF 10	E_{mc}	SampEn	

TABLE 5. Features automatically selected to achieve the highest performance by using the LDA classifier.

Combined set k	Features			QI
1	raw FHR		std	$76,7\%$
	raw FHR		mad	
	raw FHR		ApEn	
	raw FHR		SampEn	
2	IMF 4	E_{mc}	ApEn	$79,1\%$
	IMF 4	E_{mc}	SampEn	
3	PBL		std	$80,8\%$
	PBL		mad	
4	IMF 10	E_{mc}	ApEn	$81,5\%$
	IMF 10	E_{mc}	SampEn	

conventional features is illustrated in gray and white colors, respectively.

In Fig. 11, 12, and 13, we can observe that the classification performance achieved by the SVM, LDA, and *k*-NN classifiers were 81, 7%, 81, 5% and 80, 9% of quality, respectively. Moreover, for the three classifiers, such performance was achieved by the contribution of both modal-spectral and conventional features. In particular, as presented in Tables 4, 5, and 6, the highest performance obtained by the SVM and LDA classifiers was achieved by four features sets, whereas for the *k*-NN classifier, nine feature sets were required.

For the evaluation and comparison of the proposed modal-spectral features, we use the highest performance achieved by using only conventional features as a reference. In order to calculate such performance, we use the same strategy described in Fig. 10, but now considering the feature sets that include only conventional features. The results of this analysis are presented in Table 7, which shows the highest classification performance by using only conventional features and then by including the modal-spectral features, for each classifier. Here, we can observe that the best performance achieved by using only conventional features

TABLE 6. Features automatically selected to achieve the highest performance by using the *k*-NN classifier.

Combined set k	Features			QI
1, 2, 3, 4, 5, 6, 7, 8, 9	raw FHR		std	$68,5\%$
	raw FHR		mad	
	raw FHR		ApEn	
	raw FHR		SampEn	
2, 3, 4, 5, 6, 7, 8, 9	IMF 1	E_{mc}	mean	$73,0\%$
	IMF 1	E_{mc}	median	
	IMF 1	E_{mc}	RMS	
3, 4, 5, 6, 7, 8, 9	IMF 6		ApEn	$75,3\%$
4, 5, 6, 7, 8, 9	IMF 10	E_{mc}	ApEn	$75,4\%$
	IMF 10	E_{mc}	SampEn	
5, 6, 7, 8, 9	IMF 4	E_{mc}	ApEn	$75,2\%$
	IMF 4	E_{mc}	SampEn	
6, 7, 8, 9	PBL		std	$77,2\%$
	PBL		mad	
7, 8, 9	IMF 2	E_{mc}	ApEn	$77,5\%$
	IMF 2	E_{mc}	SampEn	
8, 9	IMF 1	E	mean	$79,9\%$
	IMF 1	E	median	
	IMF 1	E	RMS	
9	IMF 10	E	ApEn	$80,9\%$
	IMF 10	E	SampEn	

TABLE 7. Classification performance achieved by using only conventional, and then including the modal-spectral features, for each classifier.

Classifier	Conventional	Including modal-spectral
SVM	$78,0\%$	$81,7\%$
LDA	$77,9\%$	$81,5\%$
k-NN	$71,4\%$	$80,9\%$

was 78, 0%, and by including modal-spectral, it improved to 81, 7%, obtained by using the SVM classifier.

Considering that the best classification performance was achieved by using the SVM classifier, the analysis and evaluation of our proposed features are based on the results obtained with this classifier. As presented in detail in Table 4, the best performance was achieved by including four feature sets (two modal-spectral and two conventional feature sets). Concerning the combined set 2, we can observe that by adding only one set of modal-spectral features, the performance improved from 76, 4% to 78, 6% of quality. Then, the QI obtained by the combined set 3 achieved 80, 7% by incorporating one set of conventional features. Finally, the combined set 4 (best classification performance: $QI = 81,7\%$) consists of two sets of modal-spectral and two sets of conventional features. Particularly, the selected sets of modal-spectral features are extracted from IMF 4 and IMF 10, and the selected sets of conventional features are extracted from the raw FHR signal and PBL.

It is important to note that the combined sets from 5 to 14 (see Fig. 11) show that by including more features as input to the classifier, the classification performance may decrease. This phenomenon can be explained by the fact that even such features are considered as significant (criterion explained in Section IV-D1), they are less informative for the studied biological process compared with the features included in the combined set 4.

Conclusively, from the complete set of signal features studied in this work, as presented in Table 4, the optimal set of features consists of four modal-spectral and six conventional features. Therefore, if we compare the performance achieved by using only conventional features (see Table 7) with the performance achieved by adding the proposed modal-spectral features, it increases from 78, 0% to 81, 7% of quality, which proves the proposed hypothesis.

In order to analyze the significance of our results, we compare them with approaches already published in the literature. Table 8 presents the results of recent related studies that involve a similar approach and use the pH value as a criterion for class formation.

Table 8 shows the related work (reference), used database, number of recordings (abnormal/normal), selection criteria for class formation, sensitivity (*Se*), specificity (*Sp*), and quality (QI), for each column, respectively. Some of these works do not provide QI; thus, for their comparison, QI was computed based on their Se and Sp by using (9). It is important to note that, as can be observed in the last row of this table, the values of *Se*, *Sp*, and QI obtained in our approach do not satisfy the relation presented in (9). It is because QI was not computed based on the resulting average of (*Se*) and (*Sp*), but rather it was calculated for each iteration and then averaged, as explained in Section IV-D2.

TABLE 8. Results of recent related studies using pH for class formation.

Reference	Database	# of recordings (abnormal/normal)	CTG selection criteria	Se (%)	Sp (%)	QI (%)
Spilka et al. 2012 [75]	Private	217 (94/123)	pH<7.15	$73,4$	$76,3$	$74,8$
Georgieva et al. 2013 [90]	Private	252 (126/126)	pH<7.10 $7.27<$pH<7.33	$60,3$	$67,5$	$63,8$
Xu et al. 2014 [91]	Private	510 (255/255)	pH<7.05 $7.27<$pH<7.33	83	66	74
Spilka et al. 2017 [82]	Private	1288 (37/1251)	pH≤ 7.05	73	75	74
Warmerdam et al. 2018 [92]	Private	100 (20/80)	pH<7.05	81	77	79
Rotariu et al. 2014 [93]	CTU-UHB	289 (25/264)	pH<7.20; BDecf>8	$96,0$	$87,6$	$91,7$
Rotariu et al. 2014 [94]	CTU-UHB	552 (168/384)	pH<7.20; apgar<6	$73,2$	$88,2$	$80,4$
Georgoulas et al. 2017 [80]	CTU-UHB	552 (44/508)	pH≤ 7.05	$68,5$	$77,7$	$72,9$
Yu et al. 2017 [81]	CTU-UHB	88 (44/44) 122 (61/61)	pH≤ 7.05; pH>7.20 pH≤ 7.10; pH>7.20	$75,3$ $65,4$	$84,4$ $75,4$	$79,7$ $70,2$
Zarmehri et al. 2019 [95]	CTU-UHB	552 (44/508)	pH≤ 7.05	$63,6$	$80,1$	$71,4$
Authors' approach	CTU-UHB	372 (18/354)	pH<7.05; BDecf≥ 12 pH>7.20; BDecf<12	$79,5$	$86,45$	$81,7$

The presented table shows that, in comparison with most of the related studies, our approach achieves a higher classification performance ($81,7\%$). However, in the particular case of results achieved in Rotariu et al. 2014 [93], where they have used the same database (CTU-UHB), our classification performance might seem considerably lower. This higher performance presented in [93], can be explained by the fact that they use different criteria for the dataset formation. Despite they consider a threshold based on pH value (pHn < 7.20), they excluded from the CTG database 263 recordings considering a signal-to-noise ratio under 20%. As a result, they used only 289 recordings for the analysis, whose smaller proportion of the original database makes these results not comparable to our approach.

It is important to note that the class formation criteria based on the pH value can significantly differ from one approach to another (see Table 8). This phenomenon can be explained by the fact that, although the pH value is considered as a gold standard measure for postpartum evaluation, the association of a precise value of pH at birth with the neonatal outcome is still inconclusive [96]. As a consequence, the number of abnormal/normal recordings can also differ from one work to another, which can make the comparison of results a difficult task, or in some cases, even impossible.

In summary, according to the information presented in Table 8, the obtained results reveal that our approach based on CEEMDAN in combination with TV-AR modeling is a promising strategy for FHR signal analysis, and thereby estimate the fetal distress during labor. However, considering the wide variety of class formation criteria and the number of recordings used in recent related works, more investigation is needed to evaluate our results. For this operation, together with the evaluation methodology proposed in this work, it would be necessary the study of different class formation criteria, different algorithms for automatic classification, and different CTG databases.

V. CONCLUSION

According to the performed analysis, the FHR signal can describe different dynamics in both the time-domain and the spectral-domain strongly variant in time, whose behavior can be related to the fetal condition. Particularly, the IMFs computed from the FHR signal involve significant information in the spectral-domain that can help to discriminate between normal and acidotic fetuses.

Results reveal that the highest classification performance for the SVM, LDA, and k-NN classifiers were $81,7\%$, $81,5\%$ and $80,9\%$ of quality, respectively. Interestingly, for the three employed classifiers, such performance was achieved by the contribution of both modal-spectral and conventional features. Particularly, the best performance was achieved by a combination of four features sets that include four modal-spectral features and six conventional features and using the SVM classifier. As a result, the classification performance increased from $78,0\%$ of quality, by using only conventional features, to $81,7\%$ of quality by including the proposed modal-spectral features. In other words, those results show that by including a selected set of modal-spectral features, the CTG classification performance improves compared with the performance achieved by using only conventional features, which proves the proposed hypothesis.

It is important to note that most of the features that contribute to achieve the best classification performance (six of ten features) correspond to entropy-based features. These results open perspectives for a more in-depth study of this type of features by including, together with our proposed features, other coefficients based on entropy such as spectral entropy, Rényi entropy, and permutation entropy.

As a result of the feature elimination step, the modal-spectral features computed from the frequency of the spectral component (f_{mc}) were identified as non-significant. This information indicates that the frequency behavior of the IMFs does not provide significant information about the fetal condition. Nevertheless, the energy associated with such frequency components, described by the traces E and E_{mc}, involves significant information that can help to discriminate between a normal or an acidotic fetal status.

Considering that the proposed method based on CEEMDAN and TV-AR modeling can provide significant information about fetal distress, it could be an important approach for future research that involves the study of the fetal ANS activity.

As a future step, in addition to the features proposed in this work, we propose to analyze the spectral dynamics of the IMFs during FHR decelerations and resting periods, independently. The idea is to study the contribution of such information in order to investigate if those signal patterns, analyzed independently, can provide significant information about the fetal condition that improves the CTG classification.

REFERENCES

[1] R. K. Freeman, T. J. Garite, M. P. Nageotte, and L. A. Miller, *Fetal Heart Monitoring*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2012.
[2] D. Ayres-de-Campos, C. Y. Spong, and E. Chandraharan, "FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography," *Int. J. Gynecol. Obstetrics*, vol. 131, no. 1, pp. 13–24, 2015.
[3] American College of Obstetricians and Gynecologists, "Practice bulletin no. 116: Management of intrapartum fetal heart rate tracings," *Obstetrics Gynecology*, vol. 116, no. 5, pp. 1232–1240, Nov. 2010.
[4] G. Rooth, A. Huch, and R. Huch, "Figo news: Guidelines for the use of fetal monitoring," *Int. J. Gynecol. Obstet.*, vol. 25, pp. 159–167, 1987.
[5] M. Haritopoulos, A. Illanes, and A. K. Nandi, "Survey on cardiotocography feature extraction algorithms for foetal welfare assessment," in *Proc. 14th Medit. Conf. Med. Biol. Eng. Comput.* Cham, Switzerland: Springer, 2016, pp. 1193–1198.
[6] S. Santo and D. Ayres-de Campos, "Human factors affecting the interpretation of fetal heart rate tracings: An update," *Current Opinion Obstetrics Gynecology*, vol. 24, no. 2, pp. 84–88, Mar. 2012.
[7] A. Ugwumadu, P. Steer, B. Parer, B. Carbone, C. Vayssiere, G. Maso, and S. Arulkumaran, "Time to optimise and enforce training in interpretation of intrapartum cardiotocograph," *BJOG, Int. J. Obstetrics Gynaecology*, vol. 123, no. 6, pp. 866–869, May 2016.
[8] I. Nunes and D. Ayres-de Campos, "Computer analysis of foetal monitoring signals," *Best Pract. Res. Clin. Obstetrics Gynaecology*, vol. 30, pp. 68–78, Jan. 2016.
[9] J. E. Lutomski, S. Meaney, R. A. Greene, A. C. Ryan, and D. Devane, "Expert systems for fetal assessment in labour," *Cochrane Database Syst. Rev.*, Apr. 2013, pp. 1–10.
[10] C. Garabedian, J. De Jonckheere, L. Butruille, P. Deruelle, L. Storme, and V. Houfflin-Debarge, "Understanding fetal physiology and second line monitoring during labor," *J. Gynecology Obstetrics Hum. Reproduction*, vol. 46, no. 2, pp. 113–117, Feb. 2017.
[11] A. Ugwumadu, "Are we (mis)guided by current guidelines on intrapartum fetal heart rate monitoring? Case for a more physiological approach to interpretation," *BJOG, Int. J. Obstetrics Gynaecology*, vol. 121, no. 9, pp. 1063–1070, Aug. 2014.
[12] N. E. Huang, Z. Shen, Steven R. Long, M. C. Wu, Hsing H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London Ser. A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[13] C. Junsheng, Y. Dejie, and Y. Yu, "A fault diagnosis approach for roller bearings based on EMD method and AR model," *Mech. Syst. Signal Process*, vol. 20, pp. 350–362, Feb. 2006.
[14] T.-Y. Wu, H.-C. Hong, and Y.-L. Chung, "A looseness identification approach for rotating machinery based on post-processing of ensemble empirical mode decomposition and autoregressive modeling," *J. Vibrat. Control*, vol. 18, no. 6, pp. 796–807, 2012.
[15] O. Sahu, V. Anand, V. Kanhangad, and R. B. Pachori, "Classification of magnetic resonance brain images using bi-dimensional empirical mode decomposition and autoregressive model," *Biomed. Eng. Lett.*, vol. 5, no. 4, pp. 311–320, Dec. 2015.
[16] W.-Y. Duan, L.-M. Huang, Y. Han, and D.-T. Huang, "A hybrid EMD-AR model for nonlinear and non-stationary wave forecasting," *J. Zhejiang Univ. Sci. A*, vol. 17, no. 2, pp. 115–129, Feb. 2016.
[17] A. Agostinelli, E. Braccili, E. Marchegiani, R. Rosati, A. Sbrollini, L. Burattini, M. Morettini, F. Di Nardo, S. Fioretti, and L. Burattini, "Statistical baseline assessment in cardiotocography," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 3166–3169.
[18] S.-Y. Wei, Y.-S. Lu, and X.-L. Liu, "Fetal heart rate analysis using a nonlinear baseline and variability estimation method," in *Proc. 5th Int. Conf. Biomed. Eng. Inform.*, Oct. 2012, pp. 532–536.
[19] M. Cesarelli, M. Romano, G. D'Addio, M. Ruffo, P. Bifulco, G. Pasquariello, and A. Fratini, "Floatingline estimation in FHR signal analysis," in *5th Eur. Conf. Int. Fed. Med. Biol. Eng.* Springer, 2011, pp. 179–182.
[20] S. Das, K. Roy, and C. Saha, "A novel approach for extraction and analysis of variability of baseline," in *Proc. Int. Conf. Recent Trends Inf. Syst.*, Dec. 2011, pp. 336–339.
[21] M. Cesarelli, M. Romano, and P. Bifulco, "Comparison of short term variability indexes in cardiotocographic foetal monitoring," *Comput. Biol. Med.*, vol. 39, no. 2, pp. 106–118, Feb. 2009.
[22] P. Fuentealba, A. Illanes, and F. Ortmeier, "Analysis of the foetal heart rate in cardiotocographic recordings through a progressive characterization of decelerations," *Current Directions Biomed. Eng.*, vol. 3, no. 2, pp. 423–427, 2017.
[23] H. Cao, D. E. Lake, J. E. Ferguson, II, C. A. Chisholm, M. P. Griffin, and J. R. Moorman, "Toward quantitative fetal heart rate monitoring," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 111–118, Jan. 2006.
[24] J. Y. Kwon, I. Y. Park, J. C. Shin, J. Song, R. Tafreshi, and J. Lim, "Specific change in spectral power of fetal heart rate variability related to fetal acidemia during labor: Comparison between preterm and term fetuses," *Early Hum. Develop.*, vol. 88, no. 4, pp. 203–207, Apr. 2012.
[25] H. Gonçalves, A. P. Rocha, D. Ayres-de Campos, and J. Bernardes, "Internal versus external intrapartum foetal heart rate monitoring: The effect on linear and nonlinear parameters," *Physiolog. Meas.*, vol. 27, no. 3, p. 307, Feb. 2006.
[26] M. G. Signorini, G. Magenes, S. Cerutti, and D. Arduini, "Linear and nonlinear parameters for the analysisof fetal heart rate signal from cardiotocographic recordings," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 3, pp. 365–374, Mar. 2003.
[27] S. M. Siira, T. H. Ojala, T. J. Vahlberg, K. G. Rosén, and E. M. Ekholm, "Do spectral bands of fetal heart rate variability associate with concomitant fetal scalp pH," *Early Hum. Develop.*, vol. 89, no. 9, pp. 739–742, Sep. 2013.
[28] J. O. E. H. van Laar, C. H. L. Peters, S. Houterman, P. F. F. Wijn, A. Kwee, and S. G. Oei, "Normalized spectral power of fetal heart rate variability is associated with fetal scalp blood pH," *Early Hum. Develop.*, vol. 87, no. 4, pp. 259–263, Apr. 2011.
[29] J. Van Laar, C. Peters, R. Vullings, S. Houterman, J. Bergmans, and S. Oei, "Fetal autonomic response to severe acidaemia during labour," *BJOG, Int. J. Obstetrics Gynaecology*, vol. 117, no. 4, pp. 429–437, Mar. 2010.
[30] J. O. E. H. Van Laar, C. H. L. Peters, R. Vullings, S. Houterman, and S. G. Oei, "Power spectrum analysis of fetal heart rate variability at near term and post term gestation during active sleep and quiet sleep," *Early Hum. Develop.*, vol. 85, no. 12, pp. 795–798, Dec. 2009.
[31] J. O. E. H. Van Laar, M. M. Porath, C. H. L. Peters, and S. G. Oei, "Spectral analysis of fetal heart rate variability for fetal surveillance: Review of the literature," *Acta Obstetricia Et Gynecologica Scandinavica*, vol. 87, no. 3, pp. 300–306, Mar. 2008.
[32] M. Cesarelli, M. Romano, P. Bifulco, F. Fedele, and M. Bracale, "An algorithm for the recovery of fetal heart rate series from CTG data," *Comput. Biol. Med.*, vol. 37, no. 5, pp. 663–669, 2007.
[33] S. M. Siira, T. H. Ojala, T. J. Vahlberg, J. O. Jalonen, I. A. Välimäki, K. G. Rosén, and E. M. Ekholm, "Marked fetal acidosis and specific changes in power spectrum analysis of fetal heart rate variability recorded during the last hour of labour," *BJOG, Int. J. Obstetrics Gynaecology*, vol. 112, no. 4, pp. 418–423, Apr. 2005.

[34] P. A. Warrick and E. F. Hamilton, "Fetal heart-rate variability response to uterine contractions during labour and delivery," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2012, pp. 417–420.

[35] S. Cazares, M. Moulden, C. W. G. Redman, and L. Tarassenko, "Tracking poles with an autoregressive model: A confidence index for the analysis of the intrapartum cardiotocogram," *Med. Eng. Phys.*, vol. 23, no. 9, pp. 603–614, Nov. 2001.

[36] A. Illanes and M. Haritopoulos, "Fetal heart rate feature extraction from cardiotocographic recordings through autoregressive model's power spectral- and pole-based analysis," in *Proc. 37th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Aug. 2015, pp. 5842–5845.

[37] C. Granero-Belinchon, S. G. Roux, P. Abry, M. Doret, and N. B. Garnier, "Information theory to probe intrapartum fetal heart rate dynamics," *Entropy*, vol. 19, no. 12, p. 640, Nov. 2017.

[38] P. A. Warrick and E. F. Hamilton, "Mutual information estimates of CTG synchronization," in *Proc. Comput. Cardiol. Conf. (CinC)*, Sep. 2015, pp. 137–139.

[39] M. D. Costa, W. T. Schnettler, C. Amorim-Costa, J. Bernardes, A. Costa, A. L. Goldberger, and D. Ayres-de Campos, "Complexity-loss in fetal heart rate dynamics during labor as a potential biomarker of acidemia," *Early Hum. Develop.*, vol. 90, no. 1, pp. 67–71, Jan. 2014.

[40] M. Ferrario, M. G. Signorini, and G. Magenes, "Complexity analysis of the fetal heart rate variability: Early identification of severe intrauterine growth-restricted fetuses," *Med. Biol. Eng. Comput.*, vol. 47, no. 9, pp. 911–919, Sep. 2009.

[41] M. Ferrario, M. G. Signorini, G. Magenes, and S. Cerutti, "Comparison of entropy-based regularity estimators: Application to the fetal heart rate signal for the identification of fetal distress," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 1, pp. 119–125, Jan. 2006.

[42] G. Magenes, R. Bellazzi, A. Fanelli, and M. G. Signorini, "Multivariate analysis based on linear and non-linear FHR parameters for the identification of IUGR fetuses," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 1868–1871.

[43] M. Romano, M. Bracale, M. Cesarelli, M. Campanile, P. Bifulco, M. De Falco, M. Sansone, and A. Di Lieto, "Antepartum cardiotocography: A study of fetal reactivity in frequency domain," *Comput. Biol. Med.*, vol. 36, no. 6, pp. 619–633, Jun. 2006.

[44] S. Dong, G. Azemi, and B. Boashash, "Improved characterization of HRV signals based on instantaneous frequency features estimated from quadratic time–frequency distributions with data-adapted kernels," *Biomed. Signal Process. Control*, vol. 10, pp. 153–165, Mar. 2014.

[45] M. Romano, P. Bifulco, M. Cesarelli, M. Sansone, and M. Bracale, "Foetal heart rate power spectrum response to uterine contraction," *Med. Biol. Eng. Comput.*, vol. 44, no. 3, pp. 188–201, Mar. 2006.

[46] P. Fuentealba, A. Illanes, and F. Ortmeier, "Foetal heart rate signal spectral analysis by using time-varying autoregressive modelling," *Current Directions Biomed. Eng.*, vol. 4, no. 1, pp. 579–582, 2018.

[47] C. Cattani, O. Doubrovina, S. Rogosin, S. L. Voskresensky, and E. Zelianko, "On the creation of a new diagnostic model for fetal well-being on the base of wavelet analysis of cardiotocograms," *J. Med. Syst.*, vol. 30, no. 6, pp. 489–494, Dec. 2006.

[48] M. Bursa and L. Lhotska, "The use of convolutional neural networks in biomedical data processing," in *Proc. Int. Conf. Inf. Technol. Bio.- Med. Informat.* New York, NY, USA: Springer, 2017, pp. 100–119.

[49] M. Romano, L. Iuppariello, A. M. Ponsiglione, G. Improta, P. Bifulco, and M. Cesarelli, "Frequency and time domain analysis of foetal heart rate variability with traditional indexes: A critical survey," *Comput. Math. Methods Med.*, vol. 2016, Mar. 2016, Art. no. 9585431.

[50] M. A. Colominas, G. Schlotthauer, and M. E. Torres, "Improved complete ensemble EMD: A suitable tool for biomedical signal processing," *Biomed. Signal Process. Control*, vol. 14, pp. 19–29, Nov. 2014.

[51] S. M. Kay and S. L. Marple, "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.

[52] N. Krupa, M. Ali, E. Zahedi, S. Ahmed, and F. M. Hassan, "Antepartum fetal heart rate feature extraction and classification using empirical mode decomposition and support vector machine," *Biomed. Eng. Online*, vol. 10, p. 6, Jan. 2011.

[53] B. Krupa, M. M. Ali, and E. Zahedi, "The application of empirical mode decomposition for the enhancement of cardiotocograph signals," *Physiolog. Meas.*, vol. 30, no. 8, p. 729, Jun. 2009.

[54] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. R. Soc. London. A, Math., Phys. Eng. Sci.*, vol. 460, no. 2046, pp. 1597–1611, 2004.

[55] J. Huang, "Study of autoregressive (AR) spectrum estimation algorithm for vibration signals of industrial steam turbines," *Int. J. Control Autom.*, vol. 7, no. 8, pp. 349–362, 2014.

[56] V. Chudáček, J. Spilka, M. Burša, P. Janků, L. Hruban, M. Huptych, and L. Lhotská, "Open access intrapartum CTG database," *BMC Pregnancy Childbirth*, vol. 14, no. 1, p. 16, Jan. 2014.

[57] J. Spilka, G. Georgoulas, P. Karvelis, V. P. Oikonomou, V. Chudáček, C. Stylios, L. Lhotská, and P. Janků, "Automatic evaluation of FHR recordings from CTU-UHB CTG database," in *Information Technology in Bio- and Medical Informatics*. Cham, Switzerland: Springer, 2013, pp. 47–61.

[58] T. Kupka, J. Wrobel, J. Jezewski, A. Gacek, and M. Jezewski, "Evaluation of fetal heart rate baseline estimation method using testing signals based on a statistical model," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Sep. 2006, pp. 3728–3731.

[59] R. Sameni, M. B. Shamsollahi, and C. Jutten, "Model-based Bayesian filtering of cardiac contaminants from biomedical recordings," *Physiolog. Meas.*, vol. 29, no. 5, p. 595, May 2008.

[60] P. Fuentealba, A. Illanes, and F. Ortmeier, "Progressive fetal distress estimation by characterization of fetal heart rate decelerations response based on signal variability in cardiotocographic recordings," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2017, pp. 1–4.

[61] H. M. Al-Angari, Y. Kimura, L. J. Hadjileontiadis, and A. H. Khandoker, "A hybrid EMD-kurtosis method for estimating fetal heart rate from continuous Doppler signals," *Frontiers Physiol.*, vol. 8, p. 641, Aug. 2017.

[62] M. Romano, G. Faiella, F. Clemente, L. Iuppariello, P. Bifulco, and M. Cesarelli, "Analysis of foetal heart rate variability components by means of empirical mode decomposition," in *Proc. 14th Medit. Conf. Med. Biol. Eng. Comput.* Paphos, Cyprus: Springer, 2016, pp. 71–74.

[63] Y. Lu, X. Li, S. Wei, and X. Liu, "Fetal heart rate baseline estimation with analysis of fetal movement signal," *Bio-Med. Mater. Eng.*, vol. 24, no. 6, pp. 3763–3769, 2014.

[64] M. R. Ortiz, E. R. Bojorges, S. D. Aguilar, J. C. Echeverria, R. Gonzalez-Camarena, S. Carrasco, M. J. Gaitan, and A. Martinez, "Analysis of high frequency fetal heart rate variability using empirical mode decomposition," in *Proc. Comput. Cardiol.*, Sep. 2005, pp. 675–678.

[65] P. Fuentealba, A. Illanes, and F. Ortmeier, "Cardiotocograph data classification improvement by using empirical mode decomposition," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5646–5649.

[66] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2008.

[67] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 4144–4147.

[68] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing* . Boston, MA, USA: McGraw-Hill, 2000.

[69] P. Fuentealba, A. Illanes, and F. Ortmeier, "Spectral-based analysis of progressive dynamical changes in the fetal heart rate signal during labor by using empirical mode decomposition," in *Proc. Comput. Cardiol. (CinC)*, Sep. 2018, pp. 1–4.

[70] N. Kumar, A. Suman, and K. Sawant, "Relationship between immediate postpartum umbilical cord blood pH and fetal distress," *Int. J. Contemp. Pediatrics*, vol. 3, no. 1, pp. 113–119, 2016.

[71] J. D. Gibbons and S. Chakraborti, "Nonparametric statistical inference," in *International Encyclopedia of Statistical Science*. Cham, Switzerland: Springer, 2011, pp. 977–979.

[72] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol.-Heart Circulatory Physiol.*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[73] S. Pincus, "Approximate entropy (ApEn) as a complexity measure," *Chaos, Interdiscipl. J. Nonlinear Sci.*, vol. 5, no. 1, pp. 110–117, 1995.

[74] P. Fergus, M. Selvaraj, and C. Chalmers, "Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces," *Comput. Biol. Med.*, vol. 93, pp. 7–16, Feb. 2018.

[75] J. Spilka and V. Chudáček, M. Koucký, L. Lhotská, M. Huptych, P. Janku, G. Georgoulas, and C. Stylios, "Using nonlinear features for fetal heart rate classification," *Biomed. Signal Process. Control*, vol. 7, no. 4, pp. 350–357, Jul. 2012.

[76] S. M. Pincus and R. R. Viscarello, "Approximate entropy: A regularity measure for fetal heart rate analysis," *Obstet Gynecol*, vol. 79, no. 2, pp. 249–255, Feb. 1992.
[77] J. Spilka, V. Chudácek, M. Koucký, and L. Lhotská, "Assessment of non-linear features for intrapartal fetal heart rate classification," in *Proc. 9th Int. Conf. Inf. Technol. Appl. Biomed.*, Nov. 2009, pp. 1–4.
[78] J. Spilka, G. Georgoulas, P. Karvelis, and V. Chudáček, C. D. Stylios, and L. Lhotská, "Discriminating normal from 'Abnormal' pregnancy cases using an automated FHR evaluation method," in *Proc. 8th Hellenic Conf. Artif. Intell.* Ioannina, Greece: Springer, 2014, pp. 521–531.
[79] J. Spilka, R. Leonarduzzi, V. Chudácek, P. Abry, and M. Doret, "Fetal heart rate classification: First vs. Second stage of labor," in *Proc. 8th Int. Workshop Biosignal Interpretation*, Osaka, Japan, Nov. 2016, pp. 1–3.
[80] G. Georgoulas, P. Karvelis, J. Spilka, V. Chudáček, C. D. Stylios, and L. Lhotská, "Investigating pH based evaluation of fetal heart rate (FHR) recordings," *Health Technol.*, vol. 7, nos. 2–3, pp. 241–254, Nov. 2017.
[81] K. Yu, J. G. Quirk, and P. M. Djurić, "Dynamic classification of fetal heart rates by hierarchical Dirichlet process mixture models," *PloS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0185417.
[82] J. Spilka, J. Frecon, R. Leonarduzzi, N. Pustelnik, P. Abry, and M. Doret, "Sparse support vector machine for intrapartum fetal heart rate classification," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 3, pp. 664–671, May 2017.
[83] S. G. Kwak and S.-H. Park, "Normality test in clinical research," *J. Rheumatic Diseases*, vol. 26, no. 1, pp. 5–11, 2019.
[84] G. Georgoulas, C. D. Stylios, and P. Groumpos, "Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 5, pp. 875–884, May 2006.
[85] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intell.)*, Jun. 2008, pp. 1322–1328.
[86] M.-L. Huang and Y.-Y. Hsu, "Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network," *J. Biomed. Sci. Eng.*, vol. 5, no. 9, p. 526, Sep. 2012.
[87] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.
[88] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Müllers, "Fisher discriminant analysis with kernels," in *Proc. Neural Netw. Signal Process. 9th Proc. IEEE Signal Process. Soc. Workshop*, Aug. 1999, pp. 41–48.
[89] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.
[90] A. Georgieva, S. J. Payne, M. Moulden, and C. W. G. Redman, "Artificial neural networks applied to fetal monitoring in labour," *Neural Comput. Appl.*, vol. 22, no. 1, pp. 85–93, Jan. 2013.
[91] L. Xu, C. W. G. Redman, S. J. Payne, and A. Georgieva, "Feature selection using genetic algorithms for fetal heart rate analysis," *Physiolog. Meas.*, vol. 35, no. 7, p. 1357, May 2014.
[92] G. Warmerdam, R. Vullings, J. O. E. H. Van Laar, J. W. M. Bergmans, L. Schmitt, and S. G. Oei, "Detection rate of fetal distress using contraction-dependent fetal heart rate variability analysis," *Physiolog. Meas.*, vol. 39, no. 2, 2018, Art. no. 025008.
[93] C. Rotariu, A. Pasarica, H. Costin, and D. Nemescu, "Spectral analysis of fetal heart rate variability associated with fetal acidosis and base deficit values," in *Proc. Int. Conf. Develop. Appl. Syst. (DAS)*, May 2014, pp. 210–213.
[94] C. Rotariu, A. Pasarica, G. Andruseac, H. Costin, and D. Nemescu, "Automatic analysis of the fetal heart rate variability and uterine contractions," in *Proc. Int. Conf. Expo. Elect. Power Eng. (EPE)*, Oct. 2014, pp. 553–556.
[95] M. N. Zarmehri, L. Castro, J. Santos, J. Bernardes, A. Costa, and C. C. Santos, "On the prediction of foetal acidaemia: A spectral analysis-based approach," *Comput. Biol. Med.*, vol. 109, pp. 235–241, Jun. 2019.
[96] F. Perveen, A. Khan, T. Ali, and S. Rabia, "Umbilical cord blood pH in intrapartum hypoxia," *J. College Physicians Surgeons Pakistan*, vol. 25, no. 9, pp. 667–670, Sep. 2015.

PATRICIO FUENTEALBA (S'16) received the B.S. degree in electronics engineering from the Universidad Austral de Chile (UACH), in 2013. He is currently pursuing the Ph.D. degree in computer science with the Otto-von-Guericke University Magdeburg, Germany.

From 2013 to 2015, he was a Development Engineer in several Research and Development projects in the field of signal and image processing with UACH, where he is also an Assistant Professor with the Instituto de Electricidad y Electrónica, Facultad de Ciencias de la Ingeniería. His research interests include biomedical signal processing, time-varying spectral analysis, signal feature extraction, and computer-based classification. His awards and honors include the best score of the electronics engineering from the Austral University of Chile, in 2013 and a Fellowship for Ph.D. studies from the Chilean National Scholarship Program for Graduated Studies (Becas Chile - CONICYT), in 2015.

ALFREDO ILLANES was born in Santiago, Chile, in 1978. He received the Electronic Engineering degree from UTFSM, Valparaiso, Chile, in 2002, the M.Sc. degree in signal processing from the University of Nice Sophia Antipolis, France, in 2003, and the Ph.D. degree in signal processing from INRIA, Rennes, France, in 2008, in the area of biosignal processing and modeling.

From 2008 until 2015, he was an Assistant Professor with UACH, Valdivia, Chile, where his main research area was signal processing with applications in computer vision and vibratory processes. He is currently a Researcher with the Chair of Intelligent Catheters, Otto-von-Guericke University Magdeburg, Germany. His main research interests include time-variant signal processing and modeling in vibratory and biological processes.

FRANK ORTMEIER received the Ph.D. degree from the University of Augsburg, in 2005.

After three years, he employed as a Postdoctoral Researcher, Augsburg, he became an Associate Professor of Computer Systems in Engineering, Magdeburg, in 2008. Since 2013, he has been holding the Chair of software engineering with the Faculty of Computer Science, OvGU. He is currently a Full Professor and the Head of the Chair of Software Engineering (CSE) with Otto-von-Guericke University Magdeburg, Germany. He is also leading several research projects, coordinating the bachelor's degree program Computer Systems in Engineering and the master's degree program Digital Engineering. His research interests include driven by the idea of improving engineering tasks with methods from computer science – with a special focus on methods from Software Engineering, formal specification techniques, mobile assistance, and robotics. He is a founding member of the university's Center for Digital Engineering, Management and Operations (CeDEMO).