



Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Fetal states identification in cardiotocographic tracings through discrete emissions multivariate hidden Markov models

Edoardo Spairani^{a,b,*}, Giulio Steyde^b, Salvatore Tagliaferri^c, Maria G. Signorini^b, Giovanni Magenes^a

^a Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Pavia, Italy

^b Department of Electronics, Information and Bioengineering (DEIB), Politecnico di Milano, Milano, Italy

^c Department of Obstetrical–Gynaecological and Urological Science and Reproductive Medicine, Federico II University, Naples, Italy

ARTICLE INFO

Keywords:

Cardiotocography
Clustering
Fetal states
Hidden Markov models

ABSTRACT

Background and objectives: Computerized Cardiotocography (cCTG) allows to analyze the Fetal Heart Rate (FHR) objectively and thoroughly, providing valuable insights on fetal condition. A challenging but crucial task in this context is the automatic identification of fetal activity and quiet periods within the tracings. Different neural mechanisms are involved in the regulation of the fetal heart, depending on the behavioral states. Thereby, their correct identification has the potential to increase the interpretability and diagnostic capabilities of FHR quantitative analysis. Moreover, the most common pathologies in pregnancy have been associated with variations in the alternation between quiet and activity states.

Methods: We address the problem of fetal states clustering by means of an unsupervised approach, resorting to the use of a multivariate Hidden Markov Models (HMM) with discrete emissions. A fixed length sliding window is shifted on the CTG traces and a small set of features is extracted at each slide. After an encoding procedure, these features become the emissions of a multivariate HMM in which *quiet* and *activity* are the hidden states. After an unsupervised training procedure, the model is used to automatically segment signals.

Results: The achieved results indicate that our developed model exhibits a high degree of reliability in identifying quiet and activity states within FHR signals. A set of 35 CTG signals belonging to different pregnancies were independently annotated by an expert gynecologist and segmented using the proposed HMM. To avoid any bias, the physician was blinded to the results provided by the algorithm. The overall agreement between the HMM's predictions and the clinician's interpretations was 90%.

Conclusions: The proposed method reliably identified fetal behavioral states, the alternance of which is an important factor in the fetal development. One key strength of our approach lies in the ease of interpreting the obtained results. By utilizing a small set of parameters that are already used in cCTG and possess clear intrinsic meanings, our method provides a high level of explainability. Another significant advantage of our approach is its fully unsupervised learning process. The states identified by our model using the Baum-Welch algorithm are associated with the "Active" and "Quiet" states only after the clustering process, removing the reliance on expert annotations. By autonomously identifying the clusters based solely on the intrinsic characteristics of the signal, our method achieves a more objective evaluation that overcomes the limitations of subjective interpretations. Indeed, we believe it could be integrated in cCTG systems to obtain a more complete signal analysis.

1. Introduction

Computerized cardiotocography (cCTG) is increasingly assuming a central role in the scientific literature [1,2]. This diagnostic technique enables the assessment of fetal condition by quantifying variations in the Fetal Heart Rate (FHR) through linear and non-linear methods, both in

time and frequency domains [3]. Although CTG tracings are typically analyzed in a qualitative way by visual inspection in clinical practice, different numerical algorithms have shown their potential in increasing the diagnostic accuracy, bettering the capacity to identify potential pathological conditions both during antenatal period and labor. The use of Information Technology in the analysis of CTG recordings can

* Corresponding author at: Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Via Ferrata 5, 27100, Pavia, Italy.

E-mail address: edoardo.spairani01@universitadipavia.it (E. Spairani).

<https://doi.org/10.1016/j.cmpb.2023.107736>

Received 23 March 2023; Received in revised form 11 July 2023; Accepted 26 July 2023

Available online 30 July 2023

0169-2607/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

improve the diagnostic value of CTG, both because it confers objectivity and reproducibility to the method [1] and because it allows investigating characteristics of the signal that cannot be properly identified by visual inspection [2,3].

For a proper understanding of cardiocardiographic tracings, it is essential to know the pathophysiological mechanisms underlying changes in FHR both ante- and intra-partum. In particular, a correct interpretation cannot help but consider how fetal behavioral states influence heart rate and fetal responsiveness [4].

The transition from one to the other behavioral state is a sign of integrity and good maturation of the fetal central nervous system (CNS) and autonomic nervous system (ANS). For example, a trace with no accelerations in 20 min of recording (“nonreactive”) could be due to a quiet behavioral state. The longer the period of time in which there are no accelerations, the greater the risk that the fetus may be in a hypoxic condition. Conversely, a highly responsive well-oxygenated fetus might have incoming accelerations for more than an hour and simulate fetal tachycardia, a condition associated with hypoxia [5].

For these reasons, the automatic identification of fetal behavioral states is a challenging but fundamental task in cCTG. Being able to spot the distinct fetal stages within the CTG tracings has the potential to increase the interpretability and reliability of this diagnostic methodology and is of fundamental clinical relevance. In fact, from the duration of the different phases, important information about the state of fetal well-being is derived. It has been observed that prolonged phases of inactivity are important indicators of pathological conditions, and alterations in the physiological alternation of fetal states have been associated with several conditions in pregnancy [6–10]. Moreover, CTG parameters have been shown to vary substantially according to the behavioral state, which suggests that they should be more correctly interpreted knowing the fetal state in which they were computed [11–14].

Indeed, as discussed in [15], the fetal behavioral state has been shown to affect the responsiveness of CNS.

Nijhuis [4] reported that for fetuses at term (i.e., from 36–38 weeks of gestation), 4 distinct behavioral states can be identified, which closely follow the behavioral states observed in neonates. These states are referred to as 1F (quiet sleep), 2F (active sleep), 3F (quite awake) and 4F (active awake). Each is characterized by a typical FHR pattern and is associated with a different motility and eye movements profile.

In fetuses not yet at term this distinction is less clear and is often difficult to identify these 4 behavioral states by definition.

In the context of the non-stress test, it is more common to consider only two states (i.e., “active” and “quiet”) which can be more consistently identified also at earlier gestational ages [16,17].

The identification of different fetal states within CTG tracings is, to date, still left to the clinician’s experience and can consequently lead to discordance in interpretations. In the wider context of automatic identification of fetal behavioral states in the FHR, only a few simple algorithms for their identification have been presented in [18–20].

In this paper, we present an unsupervised method based on Hidden Markov Models (HMMs) for clustering FHR signals’ points as belonging to “active” and “quiet” states. We believe that HMMs could be particularly well-suited for this task for several reasons. Unlike other clustering techniques, they correctly capture the inherent temporality of the analyzed signal and they naturally exploit the clinical observation that FHR signals change their characteristics in time, alternating among different phases that present consistent similarities and are generated by variations in the fetal state, which cannot be directly observed. Moreover, they are fully data-driven and present the great advantage to be very well interpretable.

The proposed approach is based on shifting a fixed-length sliding window on the FHR tracing and extracting a small set of features at each slide. The parameters are considered to be the set of emissions/observations of the HMM while the “active” and “quiet” phases are the hidden states. After the unsupervised training phase, the model can

automatically assign each point of the shifting window to the state under which it is more likely to have observed the emitted set of observations [21].

An illustration of the procedure is shown in Fig. 1.

2. Methods

2.1. Overview of the proposed method

The core idea behind this approach is to exploit the a priori knowledge of the existence of two distinct fetal stages (A and Q) to develop an unsupervised classification model for their identification within an FHR signal. The need for adopting an unsupervised learning approach, which aims at discovering patterns and relationships in data without any predefined guidance (no labelled data as in the supervised case) derives from the absence of an objective external source of information that can be used to identify the labels (i.e., the ground truth).

In particular, in our work, we set up a discrete emissions multivariate HMM with two possible states (A and Q), one for each possible fetal phase. The HMM states are not directly visible, but they are observable through a discrete set of emissions; the latter consists of the discretized values of a set of parameters computed on a moving window running on the FHR signal. Hence, at each shift of the sliding window, the computation of a set of quantitative features, describing different aspects of the FHR signal, is performed. Each variable in the set is then codified so that the parameters’ set is in the form of a binary array, that is passed to the trained HMM. The latter gives back the state (A or Q) most likely to have given rise to the observed set of emissions.

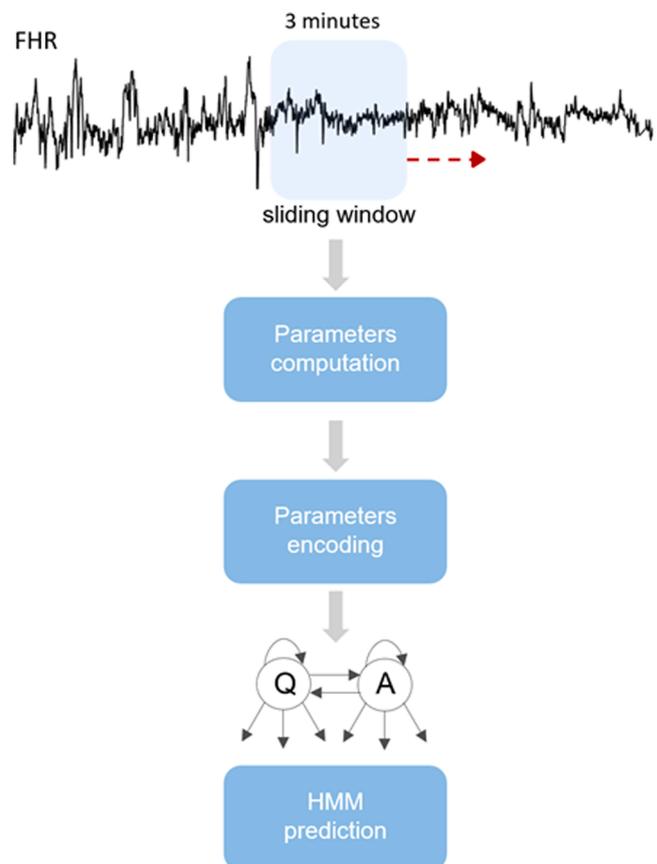


Fig. 1. Illustration of the whole described procedure. At each shift of a 3-min sliding window, a set of selected parameters, describing different aspects of the FHR chunk within the 3-min window, is obtained. The computed set is then properly encoded and passed in input to the HMM. The latter predicts the state which is more likely to have generated the observed sequence of parameters.

The remaining part of the section has the following structure: a brief introduction on HMMs is provided in Section 2; a description of the data employed to train and test the developed model is furnished in Section 3; Sections 4 and 5 respectively describe the computation of the parameters set and its categorization while Sections 6 and 7 expose the details of the presented HMM and of its training.

2.2. A brief introduction to hidden Markov models

Classical Markov models are random stochastic processes in which the probability that rules the transition from one state of the system to another one only depends on the immediately preceding one and not on the whole trajectory of states that have brought to the actual one (memory-lessness property). In standard Markov models, the states of the system are directly observable.

HMMs, on the other hand, are Markov models in which the states are not directly observable, but are inferable through a set of random variables, called observations or emissions, that are probabilistically related to the unobservable states [21].

An HMM is defined through:

- A set S of N possible states $\{S_1, \dots, S_N\}$.
- A set A of M possible emissions/observations $\{a_1, \dots, a_M\}$.
- A $N \times N$ transition matrix called P such that $P(i, j) = p_{ij} = p(S_j|S_i)$.
- A $N \times M$ emission matrix called E such that $E(i, j) = e_{ij} = p(a_j|S_i)$.
- A $1 \times N$ vector π_0 of prior probabilities.

The P matrix regulates state changes, while the E matrix rules the probability of a given state to emit a certain symbol (observation). HMMs can present univariate or multivariate emissions. In the first case, a single observation is emitted at each time, while in the second one, a set of observations is issued.

The P and E matrices are learnt during the training phase through the Baum-Welch algorithm, a special case of the expectation-maximization (EM) algorithm [22]; the algorithm requires an initial estimate of π_0 , E and P , which are then updated to the current values during the training.

Once an estimate of matrices E and P has been obtained, the model can be exploited to make predictions. Specifically, given a particular sequence of observations, the model will use the learnt knowledge to predict the trajectory of states most likely to have given rise to the observed sequence. This task is addressed through the Viterbi algorithm [23]. The prototype of an HMM is depicted in Fig. 2.

2.3. The employed dataset

The dataset we used in our work is the one described in [24], which comprises 17,483 FHR tracings sampled at 2 Hz and lasting at least 20 min. From this wider set, we isolated records for gestational weeks between 30 and 40. A preprocessing step was also necessary to clean up the

traces, which are affected by different kinds of artifacts. The denoising step includes the linear interpolation of signal losses lasting less than 15 s.

2.4. Parameters computation

We decided to perform the computation of a set of parameters on 3-minute moving windows (360 points of the tracing), shifting along the FHR signal with a 5-second stride (10 signal points).

The parameters we decided to include in our study are those which, based on our prior knowledge, we expect to vary the most between Active and Quiet states.

In particular, the regressor set we considered comprehends two groups of features, the first of which includes parameters that are directly obtained from the FHR signal, and the second of which is formed by a single signal-independent feature, hence not straightly derived from the tracing.

The first group is formed by 5 parameters and includes the variance of the signal or total power (PWT), DELTA, Sample Entropy (SampEn), the power in the very-low frequency range (VLF) and the number of accelerations within the window (Accel.). All features are computed onto the whole three-minute windows except from DELTA, which is obtained only using the points belonging to the central minute of each window, for consistency with its definition [25].

PWT and DELTA are two linear measures of variability in the time domain, which are known to increase during fetal activity [11,26]. DELTA is a commonly used parameter in clinical practice and consists of the difference between the maximum and minimum value of the signal after the application of a low-pass and down-sampling procedure excluding accelerations and decelerations [25]. Wide variations in terms of this parameter among distinct behavioral states have been reported in the literature [11]. SampEn is a family of statistical indices that measure regularity, or predictability, by counting the presence of repetitive patterns. SampEn has been shown to increase during quiet states [11, 26]. VLF was computed as the percentage of the power of the detrended signal at frequencies below 0.03 Hz. This feature has been shown to vary substantially between active and quiet states in [11].

The accelerations are identified according to the definition provided in [27] as periods longer than 15 s during which the FHR remains at least 5 bpm above the baseline and that have a maximum amplitude of at least 10 bpm.

The second group of parameters coincides with the percentage of perceived fetal movements (FMP) inside the window. The FMP signal is the result of the pregnant woman pressing a button, integrated with the system, to indicate the perception of fetal movements during the CTG examination. Technically, the FMP is a signal that has the same duration as the FHR trace, where the perceived movements are indicated by the value 1 and the remaining points have a value of 0. In our study, we refer to FMP as the percentage of perceived movements within the 3-min window.

Accelerations and fetal movements are perhaps the most typical characteristic of activity phases in CTG traces. However, their presence does not automatically indicate that the fetus is in an active state. FMPs, for example, may be present also in the quiet phase 1F, although more sporadically, and could be the result of the mother's misperception [28].

The choice of integrating a set of indices derived from the FHR tracing (PWT, DELTA, SampEn, VLF, Accel.) together with a signal-independent feature (FMP) has the intent to allow objective and subjective information to co-participate and work in tandem to the process of identifying fetal stages.

A summary reporting the computed parameters is shown in Table 1. Parameters were computed using MATLAB R2022b (The Math Works, Inc.).

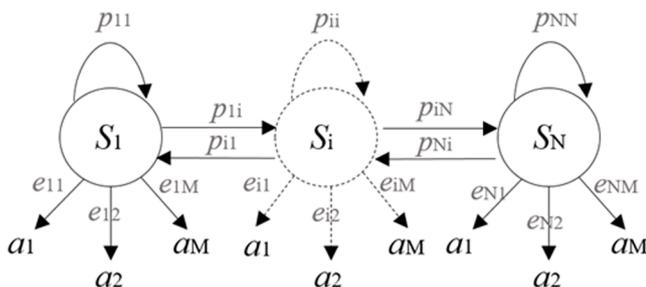


Fig. 2. Scheme of a generic Hidden Markov Model with N states and M possible emissions for each state. The probability to observe the j th emission in the i th state is ruled by the emission probability e_{ij} . On the other hand, the probability of shifting from state i to state j is governed by p_{ij} .

Table 1
Summary of Computed Parameters.

Method	Parameter	Sequence length	Hypothesis
Time domain	DELTA	1 min	Variability of FHR signal in the time domain
	PWT	3 min	
	Accel	3 min	
Frequency domain analysis	VLF	3 min	Quantification of the activity of the autonomic nervous system
Signal regularity and predictability	SampEn	3 min	Presence of recurrent patterns in a single scale
Signal Independent	FMP	3 min	Mother's perception of fetal movement

2.5. Parameters encoding: the observation set

Once the parameters' set is obtained at each shift of the 3 min sliding window, the latter needs to be properly encoded to be passed as input to the HMM.

As previously stated, the idea is to interpret the group of regressors computed at each sliding of the moving window as the set of observable random variables emitted by the actual state of the HMM (i.e., the emissions). These observations are probabilistically related to the system's state through the emission matrix E, which is estimated during the training phase.

Since each parameter is continuous, we decide to put ourselves in a simplified case, by converting each variable to categorical. PWT, VLF, SampEn, DELTA and FMP are categorized according to the 33.3rd and 66.6th percentile values of their respective PDFs, which have been computed considering all the signals in the dataset. Accel., on the other hand, is binarized with respect to the presence or absence of accelerations within the 3-minute window. Thus, each 3-minute excerpt of the FHR signal is converted into a binary vector of 17 symbols, formed by 5 triplets and a final tuple. The first 5 groups of 3 digits respectively indicate low, middle and high values of PWT, VLF, SampEn, DELTA and FMP, and the last group of 2 symbols stand for the presence or absence of accelerations.

For example, let's assume we consider a 3 min chunk of FHR signal, described by high PWT, middle VLF, low SampEn, high DELTA, high FMP and presence of accelerations; the latter will be coded as "0 0 1 0 1 0 1 0 0 0 1 0 0 1 1 0" (see Fig. 3).

2.6. The proposed HMM for the fetal state assessment

The intent of the present work is to demonstrate how HMMs can be used as a tool capable of enabling the unsupervised identification of fetal states within an FHR signal.

In our attempt to solve the problem, we make the modeling assumption that the system can be described by an HMM with two possible states ($N = 2$), i.e., activity (A) and quiet (Q). The evolutionary dynamics of this system are governed by its emission and transition probabilities, which are estimated during the training phase, as will be explained in Section 7. The emission set A is hence composed of 17 possible observations: $A = \{PWT \text{ low, PWT middle, PWT high, VLF low, VLF middle, VLF high, SampEn low, SampEn middle, SampEn high, DELTA low, DELTA middle, DELTA high, FMP low, FMP middle, FMP high, Accel. yes, Accel. no}\}$. This modeling choice hence makes the E matrix of size $N \times M$, and the P matrix of size $N \times N$ where $N = 2$ and $M = 17$. A representation of the developed HMM is depicted in Fig. 4. The HMM was implemented in Python, version 3.7.

2.7. HMM training: Estimating the model's parameters and the most likely trajectory of states

After defining the structure of the model and coding each 3 min chunk of FHR, as explained in Sections 5 and 6, we proceed to the estimation of matrices E and P through the Baum-Welch's algorithm. This requires an initial estimate of π_0 , E and P, which are then updated to the current values during the training. To avoid any kind of bias due to particular a priori modeling choices, we chose the equiprobability condition for the initialization of π_0 , E and P.

For model training, we selected a subset of 9 signals, each containing at least one quiet and one activity stage. The training set thus results in 3273 3-min excerpts.

To identify the top-performing observation set, we trained several

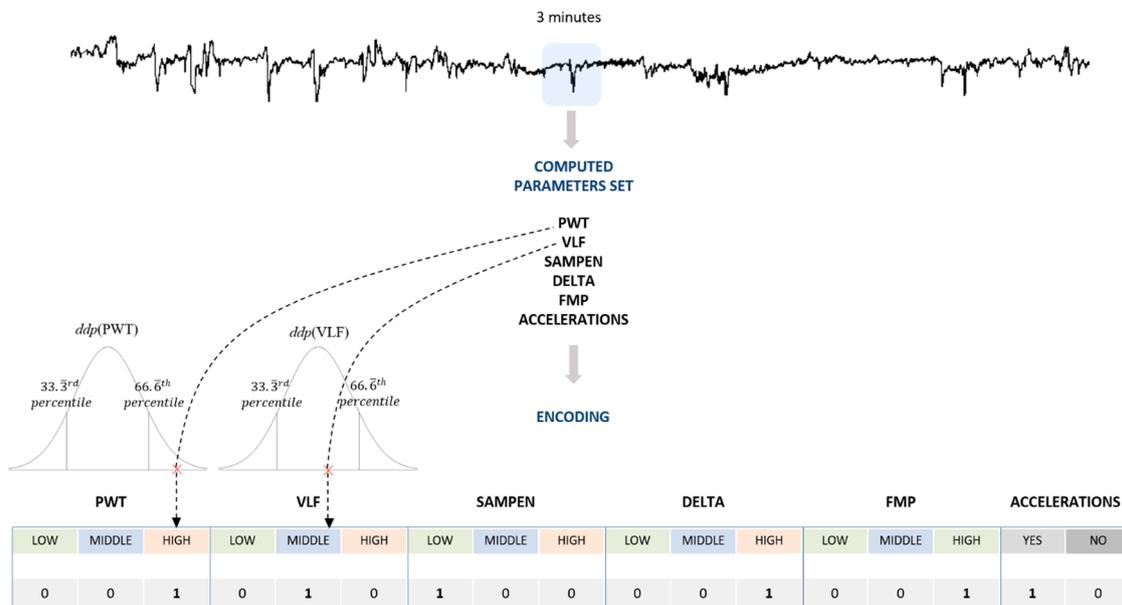


Fig. 3. Illustration of the encoding procedure. At each shift of the 3 min sliding window, a set of quantitative parameters is computed. PWT, VLF, SampEn, DELTA and FMP are categorized in low, middle and high according to the values of their pdf's percentiles 33.3 and 66.6. Accelerations, instead, are binarized with respect to the presence or absence of accelerations within the 3-minutes window. This way each 3-minutes excerpt of the FHR signal is encoded in the form of a 17 elements binary array of 1 and 0 composed of 5 triplets and a final couple.

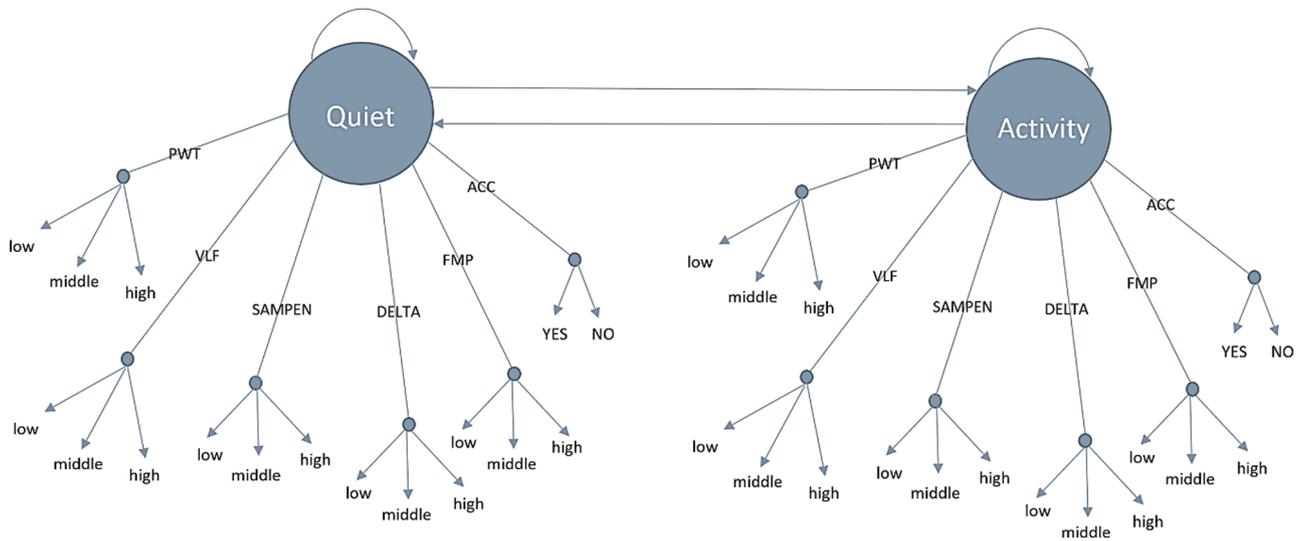


Fig. 4. Illustration of the HMM structure in the complete case. The developed HMM consists of 2 possible states (Quiet and Activity), each of which can generate the discrete set of observations.

HMMs, one for each possible combination of the parameters set. Excluding the case where no feature is included, since each parameter may or may not be included in the emission set, the number of possible combinations amounts to $2^n - 1 = 63$, where $n = 6$ equals to the number of features.

2.8. HMM testing: Finding the most likely trajectory of states

Once an estimate of matrices E and P has been obtained through the Baum-Welch’s algorithm, the model can be exploited to make predictions. Specifically, given a particular sequence of observations, the model will use the learnt knowledge to predict the trajectory of states most likely to have given rise to the observed sequence, through the Viterbi algorithm.

Once each FHR point is classified as belonging to the Active or Quiet phase, a postprocessing step is carried out. Since Quiet and Active stages are known to have a duration in the order of the tens of minutes [29], the label associated with short sub-sequences lasting less than 4 min and totally included in longer excerpts of the opposite state, is reversed.

3. Results

As discussed so far, the goal of the present work is to describe a method based on HMMs for the unsupervised clustering of fetal

behavioral states within FHR signals. Since the developed method is completely data-driven, to evaluate the model’s capacity to spot activity and quiet stages, we compared the predictions provided by our models with the annotations of an expert clinician. To avoid any kind of bias on both sides, we provided the doctor with a set of 35 unlabeled signals and asked him to annotate them. Meantime we used our HMMs to classify each data point within the same signals supplied to the clinician, which represent our testing set. Then, to assess the degree of concordance we proceeded to compare the predictions of the HMMs with the annotations of the clinician, which we consider to be the ground truth. Table 2 shows the scores for the 5 best performing HMMs and the ones for the best-performing HMM with the exclusion of FMP, ordered by decreasing accuracy. The green and red dots in Table 2 respectively indicate included and excluded features. The scores reported in Table 2 include the overall accuracy (ACC), the True Active Rate (TAR), the True Quiet Rate (TQR), the False Active Rate (FAR) and the False Quiet Rate (FQR) as defined in equations from 1 to 5.

$$Acc = \frac{TA + TQ}{TA + TQ + FA + FQ} \tag{1}$$

$$TAR = \frac{TA}{TA + FQ} \tag{2}$$

Table 2
Top 5 performing HMMs + #1 without fmp.

	FEATURES						SCORES				
	PWT	VLF	SampEn	DELTA	FMP	Accel.	ACC	TAR	TQR	FAR	FQR
HMM #1	●	●	●	●	●	●	0.90	0.90	0.87	0.09	0.13
HMM #2	●	●	●	●	●	●	0.88	0.86	0.93	0.14	0.06
HMM #3	●	●	●	●	●	●	0.88	0.89	0.87	0.11	0.12
HMM #4	●	●	●	●	●	●	0.87	0.85	0.92	0.14	0.08
HMM #5	●	●	●	●	●	●	0.85	0.82	0.93	0.17	0.07
HMM #1 without FMP	●	●	●	●	●	●	0.84	0.8	0.86	0.2	0.14

$$TQR = \frac{TQ}{TQ + FA} \quad (3)$$

$$FAR = \frac{FA}{TA + FQ} \quad (4)$$

$$FQR = \frac{FQ}{TQ + FA} \quad (5)$$

The obtained results indicate that HMM #1, represented in the first row of Table 2, stands out with the highest overall accuracy of 0.90. It includes VLF, SampEn, DELTA, FMP, and Accel. Its outperforming ACC indicates a strong concordance between the predictions of HMM #1 and the annotations made by the clinician. HMM #1 also demonstrates a high TAR of 0.90, suggesting its effectiveness in correctly identifying active states.

Comparing HMM #1 to HMM #2, which includes the whole set of computed parameters, we observe that HMM #1 has a higher accuracy (0.90 vs. 0.88) and a higher TAR (0.90 vs. 0.86). This indicates that HMM #1 is more successful in accurately identifying active states. However, HMM #1 has a lower TQR compared to HMM #2, implying that it may struggle somewhat in accurately identifying quiet phases.

HMM #3, which excludes PWT and VLF, on the other hand, achieves a TAR of 0.89 and a TQR of 0.87. These values are comparable to those of HMM #1, indicating that HMM #3 is also capable of correctly identifying both active and quiet states with a high degree of accuracy, at the expense of FAR.

HMM #4, which doesn't comprise PWT and SampEn, exhibits slightly lower performance than HMM #1 and HMM #3, with an accuracy of 0.87 and a TAR of 0.85. However, it shows a higher TQR of 0.92, suggesting its ability to accurately identify quiet phases.

HMM #5, which doesn't include PWT and DELTA, achieves the lowest overall accuracy of 0.85. However, it still presents a TQR of 0.93 showing off a high discriminative power in identifying quiet stages.

The choice of the most suitable HMM depends on the specific requirements and priorities of the analysis, considering factors such as the desired balance between accurately identifying both active and quiet states.

In our specific case we are mostly interested in maximizing the classification accuracy, since our goal is assessing the HMMs' ability to correctly classify both active and quiet states, providing a reliable measure of the model's overall performance. ACC, in fact, provides an overall assessment of the concordance between the HMM predictions and the clinician's annotations, giving us a comprehensive understanding of the model's effectiveness.

The transition matrix P for the most accurate HMM (i.e., HMM #1) is reported in Table 3.

The analysis of P provides important insights into the system's dynamics since its values indicate the probability of the system to transit from one state to another.

The high values on the main diagonal of P, hence indicate the tendency of the system to remain in the actual state for long periods rather than rapidly shifting to the other one. In fact, the probabilities associated with remaining in a quiet or active state are respectively 0.985 and 0.982; on the other hand, the probability of shifting to an active state when the system is in a quiet phase is very low (0.015) and so is the probability of passing from an active stage to a quiet one (0.018).

This kind of behavior is consistent with what is generally observed in clinics, since both active and quiet phases are known to have a duration in the order of tens of minutes [16]. This means that once the fetus enters

either the quiet or active phase, it tends to persist in that state for a considerable amount of time before transitioning to the other state.

A visualization of the observation probabilities within the emission matrix E of HMM #1 is shown in Fig. 5, which provides insights into the relationship between the probabilities associated with different feature values and the corresponding fetal states.

From the observation of Fig. 5 we can ascertain how high probabilities associated with high values of VLF, DELTA, FMP, Accel., together with low values of SampEn are more likely to reflect a state of fetal activity, rather than a quiet one. On the other hand, high probabilities associated with low values of VLF, DELTA, FMP, Accel. and high values of SampEn are very likely to be related to a quiet phase.

The analysis of Table 2 combined with the findings depicted in Fig. 5, sheds light on the heavy impact of Accel and FMP, which appear to be the most important features in the clustering process.

To evaluate the contribution of FMP, which is the only signal-independent feature, we compared the scores obtained by HMM #1 with the ones obtained with HMM #1 with the exclusion of FMP.

The inclusion of FMP in HMM #1 resulted in an improvement in the classification accuracy by approximately 6%. This suggests that FMP contributes positively to the overall performance of the model in accurately classifying fetal states.

However, it is noteworthy that even when FMP was excluded from the feature set, the obtained scores remained acceptable, particularly in terms of True Positive Rate (TPR). This indicates that the other features included in the regressor set are still capable of capturing relevant information and effectively distinguishing between fetal activity and quiet phases.

Figure 6 illustrates a visual comparison between the annotations made by the clinician and the predictions generated by HMM #1. Two exemplary signals are shown, where the green and black points represent Active and Quiet phases, respectively. Additionally, short red horizontal lines are used to indicate the segments in which the clinician's annotations differ from the model's predictions.

It is necessary to remark again that the labeling performed by the clinician was done on signals not annotated by our model, so that his judgment was not influenced by the predictions obtained from the HMM. Figure 6(a) illustrates a signal presenting a prolonged phase of fetal quiet within two active stages. By comparing the HMM predictions with the labels provided by the clinician we can assess a high degree of concordance. The only equivocal portions are concentrated in areas that straddle two distinct phases.

Figure 6(b), shows instead a signal which exhibits an initial quiet phase, lasting about 20 min, followed by an active stage approximately of the same duration. Even in this case, the model's predictions are aligned with the clinician's annotations, and the degree of concordance is still higher than in the preceding example.

From the observation of Fig. 6(a) and (b) it's possible to appreciate that portions labelled as quiet tend to exhibit lower variability and a lower tendency to accelerate as compared to activity phases. This tendency is consistent with what observed in Fig. 5, since parameters associated with signal's variability (i.e., DELTA) tend to assume higher values when the actual state is active. On the other hand, the more likely presence of accelerations within active phases determines a more predictable dynamic of the system. Again, this is consistent with what was observed from the inspection of E, since low values of SampEn are probable to reflect a condition of fetal activity.

4. Discussions and conclusions

In this work, we propose a method for unsupervised FHR signal clustering, based on Hidden Markov Models (HMM), to automatically identify fetal behavioral states of quietness and activity within CTG tracings. More specifically, the developed model is a Multivariate HMM with categorical emissions.

The obtained results seem to suggest that the developed model can

Table 3
Transition Matrix P For Hmm #1.

State	Quiet	Activity
Quiet	0.985	0.015
Activity	0.018	0.982

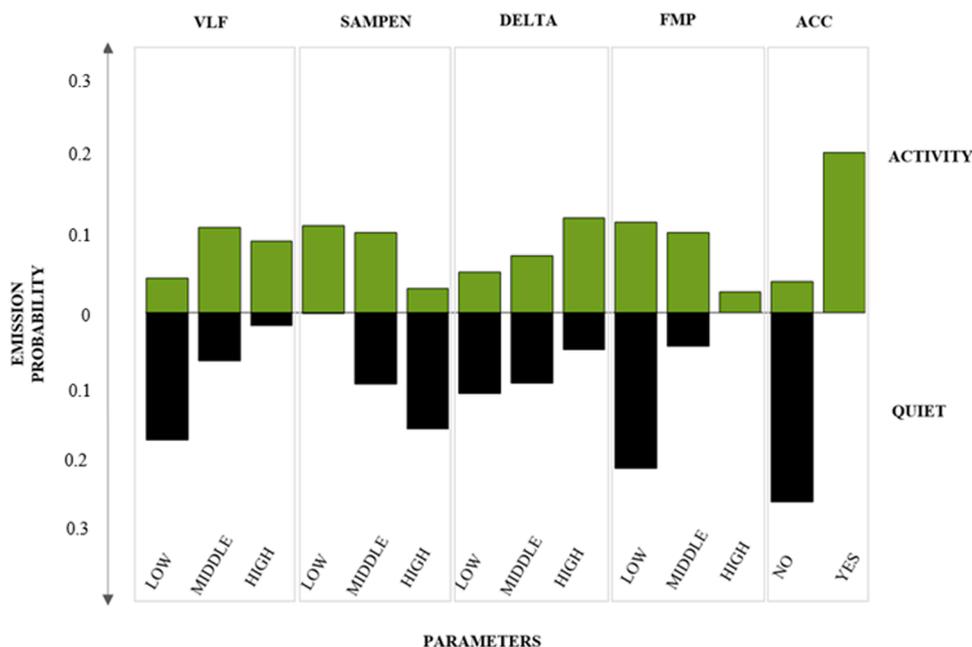


Fig. 5. Visual illustration of the emission probabilities for the best performing HMM (HMM #1).

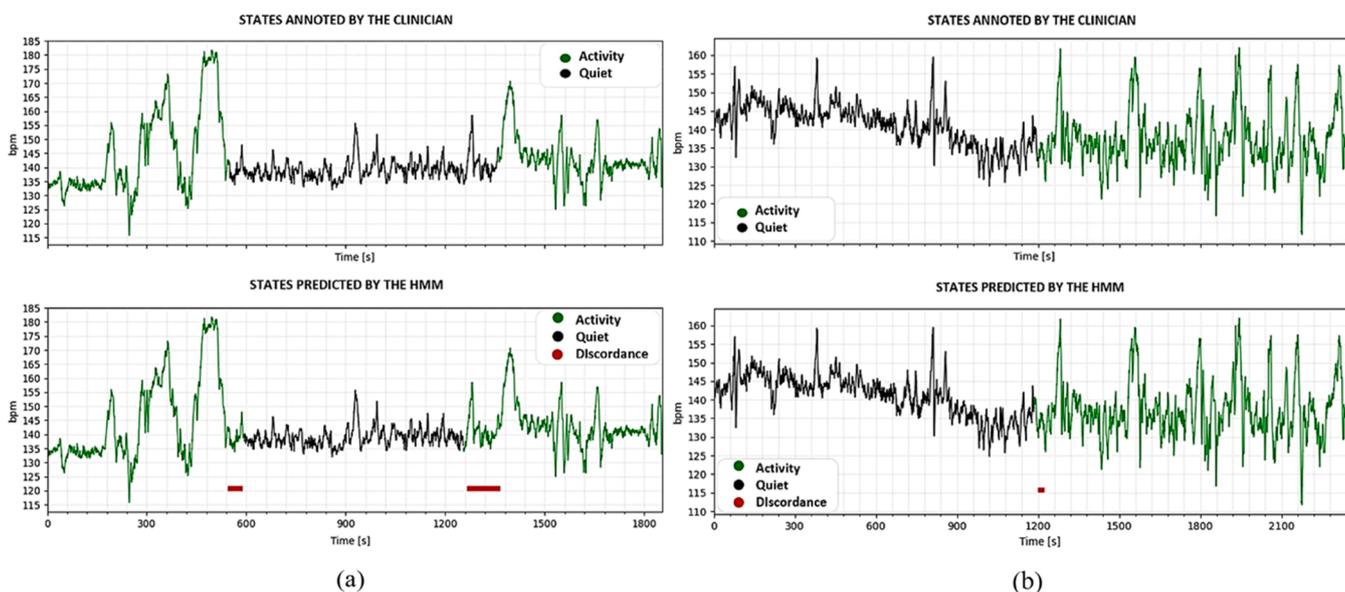


Fig. 6. comparison between the predictions of the best performing HMM (HMM #1) and the clinician’s annotations for two example signals (a) and (b). Green and black points respectively denote Active and Quiet phases. The red horizontal lines in predictions plots underline the portions in which the physician’s notes deviate from the model’s predictions.

identify quiet and activity states with a good degree of reliability. In fact, the predictions of the model have shown, for the best model, a degree of agreement of 90% with the interpretations of an expert clinician.

A desirable feature of our method is the ease of interpreting the obtained results. Indeed, the choice of a small set of parameters, already used in computerized Cardiotocography, with a clear intrinsic meaning, provides explainability to this approach.

Another appealing feature of our approach is that the learning process is fully unsupervised. In fact, the states identified by the model using the Baum-Welch algorithm are associated with the “Active” and “Quiet” states only a-posteriori. The choice of using a totally data-driven approach was suggested by the need to find a classification method able to go beyond the clinician’s interpretations. Letting the model

autonomously identify the clusters, rather than relying on expert annotations, removes the dependence on the annotator’s choice, thus hopefully reaching a more objective evaluation that is only based on the intrinsic characteristics of the signal. Indeed, in the absence of a more objective external source of information that can be used to identify the labels, e.g. ultrasound, the advantages of a supervised approach are limited.

Despite the achieved results seem to suggest that the approach can be effectively used as a tool to cluster active and quiet fetal stages, it should be taken into account that just a limited amount of data was available to test the performance of the proposed HMM. The limited quantity of testing data can be primarily attributed to the time constraints faced by the participating clinician. The time required to increase the testing set

of one order of magnitude would have caused an unacceptable delay in the publication of our work. Looking ahead, we are firmly committed to expanding our research efforts with the aim of gathering more extensive testing data. This could even benefit from the enrollment of a larger cohort of expert clinicians, boosting the truthfulness of testing ground truth. By doing so, we seek to enhance the robustness and generalizability of our results.

Future developments of the present work include defining different models tuned for gestational age and, for the last weeks of gestation, moving to a four-state model. Indeed, a model capable of distinguishing among all four behavioral states would be of great interest, although it has been shown in [4] that these states emerge with reasonable reliability only at the very end of the pregnancy.

We believe that the proposed method could represent a noticeable enhancement for the computerized analysis of the non-stress test. Indeed, the alternation of behavioral states is by itself of clinical interest and is an important pre-processing stage for the interpretation of CTG parameters that is often overlooked. Indeed, it has been clearly shown that CTG parameters widely vary between behavioral states. Reporting only their mean value without taking into account the states on which they were averaged could at least explain the large variability observed in CTG parameters, even within physiological pregnancies.

It is worth noting that including the FMP signal in the analysis only marginally increases the performance of the model, which can be successfully adopted by using the FHR signal alone.

Our method is applicable also when the FHR is extracted employing other methods rather than CTG, such as non-invasive electrophysiology. A compelling use case could be the analysis of behavioral states in long-term recordings, which could be very interesting for monitoring several pregnancy complications [6], from Intra Uterine Growth Restriction [7], diabetes [10,30], or hypertension [8], since all these conditions have been shown to have an impact on behavioral states.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is supported by the Italian Government-Progetti di interesse nazionale (PRIN) under the grant number 2017RR5EW3, 2019.

References

- [1] G.S. Dawes, M. Moulden, C.W.G. Redman, The advantages of computerized fetal heart rate analysis, *J. Perinat. Med.* (1991), <https://doi.org/10.1515/jpme.1991.19.1-2.39>.
- [2] A. Georgieva, P. Abry, I. Nunes, M. G.Frasch, Editorial: fetal-maternal monitoring in the age of artificial intelligence and computer-aided decision support: a multidisciplinary perspective", *Front. Pediatr.* (2022) 01–05, <https://doi.org/10.3389/fped.2022.1007799>.
- [3] M.G. Signorini, A. Fanelli, G. Magenes, Monitoring fetal heart rate during pregnancy: contributions from advanced signal processing and wearable technology, *Comput Math Methods Med* 2014 (2014), <https://doi.org/10.1155/2014/707581>.
- [4] J.G. Nijhuis, H.F.R. Prechtl, C.B. Martin, and R.S.G.M. Bots, "Are there behavioural states in the human fetus?", 1982.
- [5] FIGO Intrapartum Fetal Monitoring Expert Consensus Panel, FIGO consensus guidelines on intrapartum fetal monitoring, *Int. J. Gynecol. Obstet.* 131 (1) (2015), <https://doi.org/10.1016/j.ijgo.2015.06.020>.
- [6] C. Romanini, G. Rizzo, Fetal behaviour in normal and compromised fetuses. An overview, *Early Hum. Dev.* (1995), [https://doi.org/10.1016/0378-3782\(95\)01667-8](https://doi.org/10.1016/0378-3782(95)01667-8).
- [7] M.A.T. van Vliet, C.B. Jr Martin, J.G. Nijhuis, H.F. Prechtl, The relationship between fetal activity and behavioral states and fetal breathing movements in normal and growth-retarded fetuses, *Am. J. Obstet. Gynecol.* 153 (5) (1985) 582–588, [https://doi.org/10.1016/0002-9378\(85\)90483-1](https://doi.org/10.1016/0002-9378(85)90483-1).
- [8] H. Valensise, G. Ciotti, D. Giobbi, A.L. Tranquilli, D. Arduini, C. Romanini, Behaviour of fetuses from hypertensive mothers, *Hypertens Pregnancy B7* (1–2) (1988) 227–240, <https://doi.org/10.3109/10641958809023519>.
- [9] K. Sumiyoshi, Y. Kawagoe, M. Ohhashi, S. Furukawa, H. Sameshima, T. Ikenoue, Delayed rhythm formation of normal-structured, growth-restricted fetuses using fetal heart rate monitoring patterns, *J. Obstet. Gynaecol. Res.* 46 (8) (Aug. 2020) 1342–1348, <https://doi.org/10.1111/jog.14316>.
- [10] A. Kurjak, S. Panchal, S. Porovic, Fetal behavior in normal pregnancy and diabetic pregnancy, *Donald Sch. J. Ultrasound in Obstet. Gynecol.* 12 (2) (2018) 124–136, <https://doi.org/10.5005/jp-journals-10009-1562>. Jaypee Brothers Medical Publishers (P) LtdApr. 01.
- [11] H. Gonçalves, J. Bernardes, A.P. Rocha, D. Ayres-de-Campos, Linear and nonlinear analysis of heart rate patterns associated with fetal behavioral states in the antepartum period, *Early Hum. Dev.* 83 (9) (2007) 585–591, <https://doi.org/10.1016/j.earlhumdev.2006.12.006>.
- [12] L. Semeia, K. Sippel, J. Moser, H. Preissl, Evaluation of parameters for fetal behavioural state classification, *Sci. Rep.* 12 (1) (Dec. 2022), <https://doi.org/10.1038/s41598-022-07476-x>.
- [13] J. Brändle, H. Preissl, R. Draganova, E. Ortiz, K.O. Kagan, H. Abele, S.Y. Brucker, I. Kiefer-Schmidt, Heart rate variability parameters and fetal movement complement fetal behavioral states detection via magnetography to monitor neurovegetative development, *Front. Hum. Neurosci.* 9 (2015), <https://doi.org/10.3389/fnhum.2015.00147>.
- [14] S. Lange, P. Van Leeuwen, U. Schneider, B. Frank, D. Hoyer, D. Geue, D. Grönemeyer, Heart rate features in fetal behavioural states, *Early Hum. Dev.* 85 (2) (2009) 131–135, <https://doi.org/10.1016/j.earlhumdev.2008.07.004>.
- [15] I. Kiefer-Schmidt, J. Rauffer, J. Brändle, J. Münbinger, H. Abele, D. Wallwiener, H. Eswaran, H. Preissl, Is there a relationship between fetal brain function and the fetal behavioral state? A fetal MEG-study, *J. Perinat. Med.* 41 (5) (2013 Sep 1) 605–612, <https://doi.org/10.1515/jpm-2013-0022>.
- [16] D. Arduini, G. Rizzo, A. Vizzone, H. Valensise, C. Romanini, The fetal behavioural states: an ultrasonic study, *Prenat. Diagn.* 5 (1985) 269–276.
- [17] I.E. Timor-Tritsch, L.R.J. Dierker, R.H. Hertz, N.C. Deagan, M.G. Rosen, Studies of antepartum behavioral state in the human fetus at term, *Am. J. Obstet. Gynecol.* 132 (5) (Nov. 1978) 524–528, [https://doi.org/10.1016/0002-9378\(78\)90747-0](https://doi.org/10.1016/0002-9378(78)90747-0).
- [18] S. Vairavan, U.D. Ulusar, H. Eswaran, H. Preissl, J.D. Wilson, S.S. Mckelvey, C. L. Lowery, R.B. Govindan, A computer-aided approach to detect the fetal behavioral states using multi-sensor Magnetocardiographic recordings, *Comput. Biol. Med.* 69 (2016) 44–51, <https://doi.org/10.1016/j.combiomed.2015.11.017>.
- [19] A. Samjeed, M. Wahbah, A.H. Khandoker, L. Hadjileontiadis, Unsupervised fetal behavioral state classification using non-invasive electrocardiographic recordings, 2021 Computing in Cardiology (CinC), Brno Czech Republic (2021) 1–5, <https://doi.org/10.23919/CinC53138.2021.9662900>.
- [20] N. Pini, M. Lucchini, W.P. Fifer, R. Barbieri, Point process framework for the characterization of fetal sleep states, *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* (2020), <https://doi.org/10.1109/EMBC44109.2020.9176169>.
- [21] L.R. Rabiner, B.H. Juang, An introduction to hidden Markov models, *IEEE ASSP Mag.* 3 (1) (1986) 4–16, <https://doi.org/10.1109/MASSP.1986.1165342>.
- [22] L.E. Baum, T. Petrie, G. Soules, N. Weiss, A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *Ann. Math. Statist.* 41 (1) (1970) 164–171.
- [23] A.J. Viterbi, Error bounds for convolutional codes and an asymptotically optimum decoding algorithm, *IEEE Trans. Inf. Theory* 13 (2) (1967) 260–269, <https://doi.org/10.1109/TIT.1967.1054010>.
- [24] E. Spairani, B. Daniele, G. Magenes, M.G. Signorini, A novel large structured cardiotocographic database, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Jul. 2022, pp. 1375–1378, <https://doi.org/10.1109/EMBC48229.2022.9871340>.
- [25] D. Arduini, G. Rizzo, C. Romanini, Computerized analysis of fetal heart rate, *J. Perinat. Med.* 22 (Suppl. 1) (1994) 22–27, <https://doi.org/10.1515/jpme.1994.22.s1.22>.
- [26] M.G. Signorini, G. Magenes, S. Cerutti, D. Arduini, Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings, *IEEE Trans. Biomed. Eng.* 50 (3) (2003) 365–374, <https://doi.org/10.1109/TBME.2003.808824>.
- [27] R. Mantel, H.P. van Gelin, F.J.M. Caron, J.M. Swartjes, E.E. van Woerden, and H. W. Jongsmab, "Computer analysis of antepartum fetal heart rate: 2. Detection of accelerations and decelerations", 1990.
- [28] A.C. de Wit, J.G. Nijhuis, Validity of the Hewlett-Packard actograph in detecting fetal movements, *Ultrasound Obstet. Gynecol.* 22 (2) (Aug. 2003) 152–156, <https://doi.org/10.1002/uog.155>.
- [29] I.J.M. Nijhuis, J. ten Hof, J.G. Nijhuis, E.J. Mulder, H. Narayan, D.J. Taylor, G. H. Visser, Temporal organization of fetal behavior from 24-weeks gestation onwards in normal and complicated pregnancies, *Dev. Psychobiol.* 34 (4) (1999) 257–268, [https://doi.org/10.1002/\(SICI\)1098-2302\(199905\)34:2<257::AID-DEV2>3.0.CO;2-V](https://doi.org/10.1002/(SICI)1098-2302(199905)34:2<257::AID-DEV2>3.0.CO;2-V).
- [30] E.J.H. Mulder, G.H.A. Vissera, D.J. Bekedam, H.F.R. Prechtl, Emergence of behavioural states in fetuses of type-L-diabetic women, *Early Hum. Dev.* 15 (1987) 231.