



Deep Gaussian Mixture Model on Multiple Interpretable Features of Fetal Heart Rate for Pregnancy Wellness

Yan Kong¹, Bin Xu^{1,2}(✉) , Bowen Zhao², and Ji Qi²

¹ Global Innovation Exchange Institute, Tsinghua University, Beijing, China
xubin@tsinghua.edu.cn

² Department of Computer Science and Technology, Tsinghua University,
Beijing, China

Abstract. About 1% of the newborns have disorders of cardiac rhythm and conduction that result in heart failure, brain damage and even sudden newborn death. To ensure the safety of pregnancy, the electronic fetal monitoring (EFM) technique is widely used in obstetrics. However, the existing automatic diagnosis methods suffer from two main problems: insufficient features and low interpretability. In order to improve the interpretability and effect of the method, we propose a novel fully interpretable method. We first propose an iterative local linear regression (ILLR) method of linear complexity, which calculates over all local ranges of the fetal heart rate (FHR) and generates local gradients and coefficients of determination, that are used as indicators of intensity and typicality of fetal heart activity. Then, we elaborate the methodology of extraction of dozens of features by interpretable methods. Finally, we propose an interpretable deep Gaussian mixture model that can automatically select multiple features, which is composed of a mixture model based on Gaussian model weighted by features and a regression model. We conduct cross validation experiments on the full benchmark intrapartum database CTU-UHB, which shows that our method obtains significant improvements of 5.61% accuracy over state-of-the-art baselines.

Keywords: Electronic fetal monitoring · Cardiotocography · Iterative local linear regression · Deep Gaussian mixture model

1 Introduction

Electronic fetal monitoring (EFM), also known as cardiotocogram (CTG) is a technology of detecting whether a fetus is appearing signs of distress during the labor. This technique is widely used in clinics to ensure the safety of pregnancy. According to statistics, about 1% of the newborns are found to have disorders of cardiac rhythm and conduction that shown to result in heart failure, brain damage and even sudden newborn death [20]. As shown in Fig. 1, electronic fetal monitoring technology includes several steps: obtaining monitoring record data,

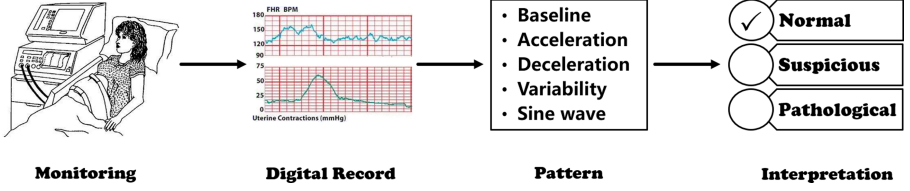


Fig. 1. The overview of the Electronic Fetal Monitoring technology.

identifying FHR patterns and making EFM interpretation. The interpretation standard of EFM is different between different countries and regions of the world, like FIGO [2], ACOG [1], and NICE [15]. At the moment, the main problem about EFM is that it is often difficult for doctors to agree on a result according to the same guideline [7]. As a result, the low-level hospitals are still using the non-stress testing (NST) or contraction stress testing (CST) as evaluation methods, whose specificities are shown no more than 50% [10, 11]. However, this paper proposes the view that the use of automatic diagnosis system can better solve the problem of inconsistent diagnosis results, and furthermore improve the overall medical level of developing areas as well as regional medical equity.

There are some problems of the automatic analysis technology with EFM. The existing feature studies on EFM are far from enough [14]. Due to this limitation, current feature based methods are not effective in actual. And the existing uninterpretable methods can not reach good effect, either.

The study of this paper focuses on an effective analytical method of EFM records by using interpretable methods, including data preprocessing and feature extraction of various scopes, and an interpretable model with the ability of multiple feature learning and high sensitivity on small data. In order to verify the effect of our method, we carry out a verification of our complete classification method on the open-access intrapartum CTG database named CTU-UHB [4].

The contributions of this paper are as follows.

- We propose a framework of new interpretable methods to acquire an automated analysis of EFM diagnosis with high efficiency. Our first innovation is the ILLR method that can calculate the least squares linear regression and coefficient of determination over N local ranges of signal of length N with $O(N)$ complexity both in time and space. The second is the methodology of feature extraction of FHR records, which can obtain 77 FHR features in various scopes that are closer to the clinical application than existing methods. Third one is, we propose an deep Gaussian mixture model (DGMM) that learn multiple features comprehensively. The model is completely interpretable, which is very in line with the preference of health and medical application.
- We evaluated our framework by cross validation on CTU-UHB dataset, and the result shows that our methods obtain significant improvements of 5.61% accuracy over the state-of-the-art baseline.

The rest of this paper is organized as follows. Section 2 introduces the related research work on automated analysis of FHR. Section 3 describes our methodology of EFM analysis including preprocessing, feature extraction as well as the interpretation and structure of DGMM. Section 4 presents the evaluation of different methods, the controlled experiments on different types of features, and the interpretation of the knowledge of the DGMM. Section 5 concludes the researches of FHR analysis methods in this paper.

2 Related Work

There are many researches on automatic analysis of FHR. [8,9] researched the approaches to determining the baseline, accelerations and decelerations. There are also some works on feature analysis [12,16,18]. [13] researched the analysis methods of FHR and put forward that computer-based EFM will be improved by better signal quality, more additional information, and more clinical data.

There have been efforts on the public intrapartum dataset CTU-UHB [4], as this can be the standard benchmark data that include discrete signal data, available biochemical marks as well as some more general clinical features. The CTG dataset contains 552 discrete records in which 46 ones are cesarean, and there is no data on inherited congenital diseases. Researches on this dataset can be divided into two paradigm.

In the first paradigm, researchers are exploiting the features in the frequency domain using wave transform. [6] proposed short time Fourier transform (STFT) to obtain the visual representation of time-frequency information. In addition, [3] and [22] proposed continuous wavelet transform (CWT) for the same purpose. However, they both have the test leakage problem since the training and test data come from similar images generated by the wavelet of different scales. [21] proposed image-based time-frequency (IBTF) analysis and LS-SVM. Besides, [19] used fast Fourier transform (FFT) to check the energy in different frequency ranges but did not get good enough result.

As for researches beyond frequency domain methods, [5] researched on feature based artificial neural network (ANN). They labeled the data by specialists for supervised learning, but only used those of certain results. Therefore, the effect is not guaranteed in general cases. [14] have used 54 features to do classification experiments by least squares support vector machines (LS-SVMs). But they came to the conclusion that using only 3 features can work best. It owes to the low comprehensiveness of their model to learn complicated features.

3 Approach

As shown in Fig. 2, our interpretable framework includes 3 steps: data preprocessing, feature extraction and classification by the DGMM model.

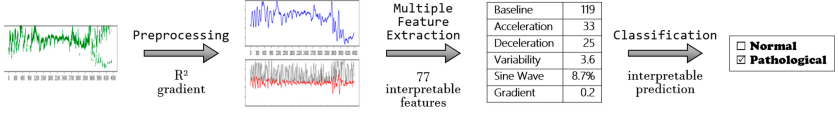


Fig. 2. The outline of our analysis framework of EFM

3.1 Data Preprocessing

The current EFM technique produces two types of discrete digital signals, namely fetal heart rate (FHR) and uterine contraction (UC). Our method only requires FHR traces for adaptability. In the preprocessing stage, we first reset the abnormal data of a period less than 15 s. Then, the ILLR method is used to perform local linear regression and data fitting.

Iterative Local Linear Regression. The ILLR method is based on least squares linear regression. This paper will give the steps and needed equations for the method. It can be proved this method can perform linear regression and linear fitting both of N times on an FHR trace of length N by $O(N)$ complexity.

For 2-dimensional data of length N , $p_i = (x_i, y_i)$, $i = 1, \dots, N$, given a range parameter r , we propose a function summation operator $S_i(f)$ to denote a summation operation for function f on the points close to p_i as

$$S_i(f) = \sum_j f(j), (|x_j - x_i| \leq r), \quad (1)$$

where f could be subscript operations, like $S_i(xy) = \sum_j x_j y_j$, $(|x_j - x_i| \leq r)$. And we specially define n as $S_i(1)$, which means the number of the points close to p_i . Consequently, the least squares solution of linear function $\hat{y} = \hat{a}_i x + \hat{b}_i$ of domain $\{p_j | |x_j - x_i| \leq r\}$ is equivalent to

$$\hat{a}_i = \frac{n \cdot S_i(xy) - S_i(x) \cdot S_i(y)}{n \cdot S_i(x^2) - S_i(x)^2}, \hat{b}_i = \bar{y} - \hat{a}_i \bar{x}, \quad (2)$$

where $\bar{x} = \frac{S_i(x)}{n}$, $\bar{y} = \frac{S_i(y)}{n}$. And the coefficient of decision is equivalent to $R_i^2 = 1 - \frac{SSE}{SST}$. It can be deduced $SSE = S_i((y - \hat{y})^2) = S_i((y - \bar{y} - (\hat{a}_i \cdot (x - \bar{x})))^2) = \frac{1}{n} (n \cdot S_i(y^2) - S_i(y)^2 + ((n \cdot S_i(x^2) - S_i(x)^2) \hat{a}_i - 2 \cdot (n \cdot S_i(xy) - S_i(x) \cdot S_i(y)) \cdot \hat{a}_i)$, and $SST = S_i((y - \bar{y})^2) = \frac{1}{n} \cdot (n \cdot S_i(y^2) - S_i(y)^2)$. Hence,

$$R_i^2 = 1 - \frac{SSE}{SST} = \frac{2(n \cdot S_i(xy) - S_i(x) \cdot S_i(y)) - \hat{a}_i (n \cdot S_i(x^2) - S_i(x)^2)}{n \cdot S_i(y^2) - S_i(y)^2} \hat{a}_i. \quad (3)$$

At last, we can refit the data with the following formula.

$$Y(x) = \frac{\sum_k R_k^2 \cdot (\hat{a}_k \cdot x_i + \hat{b}_k)}{\sum_k R_k^2}, k \in \{k | |x_k - x_i| \leq r\} \quad (4)$$

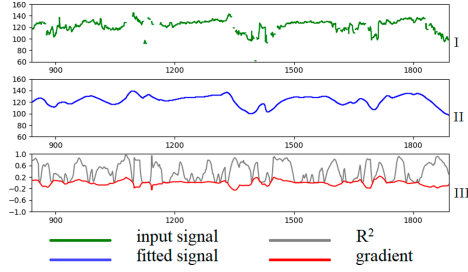


Fig. 3. Signal processing with ILLR method

According to the above formula, as the central point p_i changes, only such values need to be maintained to calculate the regression and refit the data: $S(1), S(x), S(y), S(x^2), S(xy), S(y^2)$. If we sort the points $p_i = (x_i, y_i), i = 1, \dots, N$ in order of x_i , and calculate iteratively over ordered central point p_i , the total complexity of time and space will be $O(N)$.

The ILLR method produce 2 kind of features. One refers to $\hat{a}_i$, which is called “gradient” in this paper. The other refers to R^2 , namely determination coefficient. We will introduce the significance of them later. And by the refit function, we can change the signal frequency, the missing data can also be completed.

3.2 Extraction of Multiple Interpretable Features

In our method, totally 77 features of diversified scopes are extracted from each FHR trace as shown in Table 2. And the feature extraction method is as follows.

Overall Features. In the first stage, we apply ILLR method on the whole FHR with the regression parameter $r = 15(s)$, and acquire the result as shown in Fig. 3. To the knowledge of this paper, the FHR is influenced by vagus nerve, atrioventricular status [17], blood oxygen saturation, and other factors. And we use the gradients in representation of their influence.¹ On the other hand, the R^2 , called coefficient of determination, is referred as the typicality of fetal heart activity. To the clinical knowledge, an overall reference to the FHR is needed, that is, whether the FHR activity is good and typical. For this reason, we abstracted 2 kinds of new features, which are the intensity and typicality of fetal heart activity as above.

Baseline. In the second stage, we determine the baseline of FHR, which is variable over time. We pick out the moderate signal parts or the low point. Then,

¹ According to the principle of signal processing, the relationship between the factors and heart rate is like control signal and output signal. The 2-norm of the gradient divided by the length is the ratio of energy to time, namely the power, which reflect the intensity of fetal heart activity. But this paper is not in favor of obtaining the so-called control signal from the first order difference of the FHR.

we enumerate the combinations of them and use scoring method to select the best one. The score items include prior probability, part length, value distribution of the parts, and the difference between each part and the surrounding signals. Finally, we smooth the selected parts and join them with quintic S-curve.

Accelerations and Decelerations. For the third stage, we detect accelerations (acc) and decelerations (dec). We developed a complex boundary decision algorithm because the real cases are very complicated, like shape overlapping and length change from seconds to minutes. For a specific acceleration/deceleration, we define the following features. “Time” is the duration time. “Peak” is the maximum difference between the FHR and baseline in beats per minute (bpm). “Area” is the integral area with respect to baseline. “Rank” is defined as the average rank in Table 1. “Reference” is defined as $rank/10 * (1 - miss_rate)$. These indices are closer to the clinical application than original features, as the area index can reflect the blood oxygen compensation ability and distress degree of the fetus by experience.

Table 1. Ranks of accelerations and decelerations

Rank			1	2	3	4	5	6	7	8	9	10	12	15	18
Acc	area	(bpm·s)	150	200	300	450	600	750	900	1050	1200	1400	1800	-	-
	peak	(bpm)	8	10	15	17.5	20	23.3	26.7	30	35	40	50	-	-
	time	(s)	10	12	15	30	60	90	120	160	200	240	360	-	-
Dec	area	(bpm·s)	100	200	250	375	500	550	600	650	700	750	850	1000	-
	peak	(bpm)	-	10	15	17.5	20	21.7	23.3	25	26.2	27.5	30	32.5	35
	time	(s)	7.5	15	20	25	30	36	42	48	54	60	72	90	-

Variability. We determine the long-term variability (LTV) and short-term variability (STV) by range analysis on data of 60 s and 2.5 s. We also calculate baseline variability on the data without accelerations and decelerations and determine the type by the proportion of variation range.

Sine Wave. To detect sine waves, we first split FHR into 60-s segments with the overlap of 30 s. Then, we estimate a sinusoidal function for each segment and calculate the normalized square loss of the function to the FHR. Once there is a sine wave detected on a segment with the loss not more than 0.2 times the amplitude, we combine the adjacent wave segments of little change in the amplitude and wavelength. If a sine wave with 10 continuous periods exists, the FHR is of sine wave pattern.

3.3 Interpretable Deep Gaussian Mixture Model

The mixture model, also known as the hybrid model, is composed of several different models, which stands for the existence of submodels in a large model. As

Table 2. Features with indices of amounts. acc: acceleration, dec: deceleration

Type	Feature scope	Concrete content
Acc&Dec ₃₅	reference ₂	total reference coefficient of all acc/dec
	area ₂	total area of all acc/dec (<i>bpm·s</i>)
	time ₂	total time of all acc/dec (<i>s</i>)
	peak ₂	the maximum peak of all acc/dec (<i>bpm</i>)
	rank ₇	the merged rank and average rank of all acc/dec, and the minimum, maximum and median of the merged rank of accelerations for every 20 min
	number ₈	the numbers of acc/dec in 3 different scales, and the minimum and maximum numbers of accelerations for every 20 min
	subtype ₁₂	the quite real and total numbers of specific types: abrupt acceleration, prolonged acceleration, variable deceleration (VD), late or early deceleration (LD or ED), prolonged deceleration and severe deceleration
Top ₁₆	acceleration ₈	area (<i>bpm·s</i>), duration (<i>s</i>), peak (<i>bpm</i>) and rank of the first and second top acc/dec
	deceleration ₈	
ILLR ₆	gradient ₃	the 1-norm and 2-norm both divided by data length, the standard deviation
	R^2 ₃	
Other ₂₀	raw signal ₃	
	full time ₁	the full time of FHR record (<i>s</i>)
	quality ₁	1 - the miss rate of geometric length of the data
	baseline ₃	whether the global baseline is decided, the amount of 10-minute baseline periods, the average baseline (<i>bpm</i>)
	tachycardia ₃	the specific area (<i>bpm·s</i>), time (<i>s</i>), and the average value obtained by dividing area by time (<i>bpm</i>)
	bradycardia ₃	
	variability ₄	long-term variability (<i>bpm</i>), short-term variability (<i>bpm</i>), baseline variability (<i>bpm</i>), the type of baseline variability (1 4: absent, minimal, moderate, marked)
	sine wave ₂	whether the FHR is a sine wave pattern, the time of sine wave (<i>s</i>)

the most suitable atomic model, the Gaussian model is robust enough to tackle with the complex realistic features that always conforms to the Gaussian distribution on the whole. In addition, the DGMM also learns the uniform weights of features to balance the information comprehensiveness, this further reduces over fitting and improves versatility.

Theoretical Basis of Mixture Model Based on Gaussian Model. For a data vector $\mathbf{d} = (v_1, v_2, \dots, v_P)$, supposing that each dimension is independent of the others, we define the relative probability function of the t^{th} Gaussian model, $g(\mathbf{d}, M^{(t)})$, as formula 5.

$$\prod_{i=1}^P \frac{1}{\sigma_i^{(t)}} \cdot e^{-\frac{(v_i - \mu_i^{(t)})^2}{2(\sigma_i^{(t)})^2}} = \exp \left(\sum_{i=1}^P \left(-\log(\sigma_i^{(t)}) - \frac{1}{2} \left(\frac{v_i - \mu_i^{(t)}}{\sigma_i^{(t)}} \right)^2 \right) \right) \quad (5)$$

In avoid of exponential explosion that make the power of the exponent become 0, the exponent should be extracted and applied to the *softmax* function, as shown in formula 6.

$$\begin{aligned} PD(\mathbf{d}) &= softmax \left(\log(g(\mathbf{d}, M^{(1)})), \dots, \log(g(\mathbf{d}, M^{(T)})) \right), \\ \text{where } softmax(e_1, \dots, e_n) &= \frac{1}{\sum_i \exp(e_i)} \exp(e_1, \dots, e_n) \end{aligned} \quad (6)$$

Weighting Features. The over fitting and homogenization problems are common for the network of ordinary GMM. As a solution, we apply a non-negative weight to each feature for relaxation, $\mathbf{w} = (w_1, \dots, w_P)$, $w_1, \dots, w_P \geq 0$. And we introduce the prior probability of the submodels $\exp(\mathbf{P}) = \exp(p^{(1)}, \dots, p^{(T)})$. Consequently, we derive the final formula 7 from $\log g(\mathbf{d}, M^{(t)})$ as follow.

$$\sum_{i=1}^P w_i \left(-\log(\sigma_i^{(t)}) - \frac{1}{2} \left(\frac{v_i - \mu_i^{(t)}}{\sigma_i^{(t)}} \right)^2 \right) + p^{(t)} \quad (7)$$

Topology of Deep Gaussian Mixture Model. As is shown in Fig. 4, the whole model consists of two parts. The first part is a mixture model involving feature learning that outputs $PD(\mathbf{d})$ as mentioned above. The rest is a linear regression model with one convolution layer with parameter $\mathbf{W}_1 = (u^{(1)}, u^{(2)}, \dots, u^{(T)})$, and outputs the result $R(\mathbf{d}) = \int PD(\mathbf{d}) * \mathbf{W}_1$. In addition, for initial start-up, MLE parameters can be used.

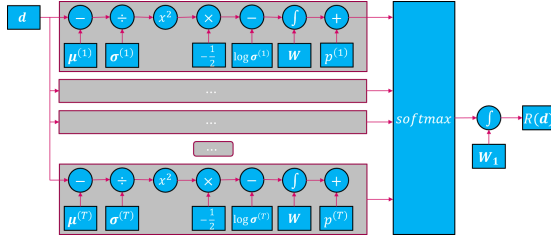


Fig. 4. The structure of deep Gaussian mixture model

4 Experiments

4.1 Dataset

We conduct experiments on the standard benchmark dataset CTU-UHB [4] to evaluate and compare our proposed method with different model parameter settings T . The database consists of 552 records of intrapartum data, each including the raw signal of FHR and UC for 30 to 90 min. The data are divided into 2 types: normal and pathological. We take the data with $\text{pH} < 7.15$ as pathological, and otherwise as normal. Eventually, 447 normal data and 105 pathological data are obtained.

4.2 Settings

We first evaluate our method by 10-fold cross-validation for a more reliable comparison with other methods. The feature dimension of DGMM is 108, including 77 original features, and normalized features that linearly related to time length. To compare with DGMM, DNN is used to test the same features, whose structure contains an input layer of size 77, three hidden full connection layers of size 128, 64 and 16, and an output layer of size 2. The result demonstrate the DGMM is much better than DNN on the same features, and reflect that the problem of overfitting will occur as the larger T is leading to better performance. The baseline methods compared with our method include ANN [5], LS-SVMs [14] and IBTF with LS-SVM [21]. Furthermore, an ablation study was conducted on all the features introduced above with different feature types, in which the model parameter was fixed with $T = 32$. In addition, we attempt to explain the DGMM's knowledge of key features learned from the data set in order of their uniform weight w of the model.

To measure the performances, the confusion matrix is used for each test. It has four prognostic indices: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). The positive and negative that represent the predicted result stand for pathological and normal respectively, and true or false represents the correctness of the prediction. The performance criteria is shown in Table 3.

Table 3. The performance criteria

Criterion	Accuracy (Acc)	Sensitivity (Se)	Specificity (Sp)	Quality index (QI)
Formulation	$\frac{TP+TN}{TP+TN+FP+FN}$	$\frac{TP}{TP+FN}$	$\frac{TN}{TN+FP}$	$\sqrt{Se \cdot Sp}$

Table 4. The results of our methods and previous methods

Method	ACC	Se	Sp	QI
DNN	67.57	42.86	73.38	56.08
DGMM (T = 16)	90.22	76.19	93.51	84.41
DGMM (T = 24)	90.22	79.05	92.84	85.67
DGMM (T = 32)	90.58	79.05	93.29	85.87
DGMM (T = 40)	89.86	79.05	92.39	85.46
ANN	84.97	74.46	89.61	81.68
LS-SVMs	76.94	68.48	77.68	72.94
IBTF	65.41	63.45	65.88	64.04

Table 5. The results of ablation study

Ablation type	ACC	Se	Sp	QI
None	90.58	79.05	93.29	85.87
Acc&Dec	71.01	54.29	74.94	63.78
Top	86.23	70.48	89.93	79.61
ILLR	83.51	58.10	89.49	72.10
Other	68.66	41.90	74.94	56.04
ILLR+Other	69.20	28.57	78.75	47.43

The results are obtained by 10-fold cross validation, based on DGMM (T = 32).

4.3 Results

Table 4 presents comparisons between the proposed model and baseline models. Note that the ANN method is tested on manually selected simple data in different stages, so we objectively calculate the average score from their report. In the comparison, our method shows a significant improvement in sensitivity, because DGMM can effectively learn feature relevance and global distribution regardless of the common error of true data, as if doctors are paying more attention to comprehensive interpretation of data. So the classification boundary is more general and human-like.

Table 5 further displays the results obtained by ablation experiments. The ablation group of *acc&dec* and of *other* both have a very low result, and the features in these groups are almost all features mentioned in obstetric guidelines. The ablation group of *top* shows a little change, because they have less chance to play a special role. And the controlled group of ILLR proposed in this paper shows a 7.07% decrease of accuracy, which shows the importance of the ILLR method.

4.4 Interpretation of DGMM

We present some of the features in weight order in Table 6. And the EFM type of each submodel can be explained according to the ACOG guideline, as follows. Model 4 reflects scarce LD and ED, but frequent prolonged decelerations, and straight baseline (very low variation power as shown by gradient), such is a very serious pathological pattern. Besides, it also implies frequent accelerations and long-term tachycardia. Model 5 shows low STV and long sinusoidal waveform (18% of 1.5 h). Model 6 means acceleration and deceleration scarcity, which is

Table 6. The feature expectation of submodel for DGMM (T=8)

Feature	$M^{(1)}$	$M^{(2)}$	$M^{(3)}$	$M^{(4)}$	$M^{(5)}$	$M^{(6)}$	$M^{(7)}$	$M^{(8)}$
LD or ED per 20 min	0.80	0.83	0.79	0.17	0.81	0.12	0.79	-0.01
Variability type *	2.45	2.97	2.15	3.04	2.47	2.27	2.18	2.24
Prolonged decelerations per 20 min	0.49	0.75	0.57	1.22	0.01	0.16	0.23	0.71
Minimum accelerations in 20 min	2.45	1.78	1.44	1.65	2.95	0.28	2.21	0.58
Accelerations per 20 min	2.18	2.13	1.24	2.33	1.49	1.05	1.38	1.07
Sine wave proportion	-0.44	0.08	0.05	-0.06	0.18	0.04	0.21	0.05
Normalized 2-norm of gradient	0.29	0.37	0.36	0.06	0.26	0.28	0.51	0.36
Tachycardia area per 20 min	-0.27	0.71	0.82	3.50	-0.39	0.48	0.94	1.83
Average tachycardia heart rate	3.16	5.43	7.46	4.59	7.09	6.13	1.64	6.80
EFM type **	1.83	1.55	1.42	-1.54	-1.82	-2.02	-2.07	-2.12

*: (1: absent, 2: minimal, 3: moderate, 4: marked)

** : (1: normal, -1: pathological)

also a bad pattern. Model 7 is long sinusoidal waveform and excessive variation power. Model 8 includes acceleration and deceleration scarcity as well as model 6, but also tachycardia that of both large area and high heart rate.

4.5 Discussion

Our method shows the best accuracy and quality index in contrast to other methods and is completely interpretable. In the methods of many researches, the researchers mostly cut off the partial data to improve the effect, because of the huge variability of features of FHR trace for an extended period. This also reflects the features we are dealing with are more complex and the classification work is more difficult. But our method can adapt to the changes of data length, and can also follow the baseline changes of FHR, so we solved the problem of processing long FHR traces with good effect.

5 Conclusion

In this paper, we propose a new function summation operator $S_i(f)$ that can be fast iterated, then we propose and prove the ILLR method that can calculate the least squares linear regression and coefficient of determination over N local ranges of signal of length N with the cost of $O(N)$. And we put forward the idea

of using the gradient and determination coefficient as the indicators of intensity and typicality of fetal heart activity.

We propose a completely interpretable methodology for preprocessing and extracting 77 different features from tens of scopes. And our experiments show the advantages of DGMM in multiple feature learning, as if doctors pay more attention to the comprehensive interpretation of data.

According to our controlled experiments, the features obtained by ILLR method work for at least 7% accuracy. And our experiments highlight the significance of the features related to accelerations and decelerations by showing them accounting for nearly 20% accuracy of the results. This paper also draws a rare conclusion that the top shape and second top shape of accelerations and decelerations have a little meaning in the analysis of FHR, too.

References

1. Arnold, K.C., Flint, C.J.: Acog practice bulletin no 106: intrapartum fetal heart rate monitoring: nomenclature, interpretation, and general management principles. *Obstetr. Gynecol.* **114**(1), 192–202 (2017)
2. Ayres-De-Campos, D., Spong, C.Y., Chandrachan, E.: Figo consensus guidelines on intrapartum fetal monitoring: cardiotocography. *Int. J. Gynecol. Obstetr.* (2015)
3. Bursa, M., Lhotska, L.: The use of convolutional neural networks in biomedical data processing. In: *International Conference on Information Technology in Bio- and Medical Informatics* (2017)
4. Chudáek, V., Spilka, J., Bura, M., Jank, P., Lhotská, L.: Open access intrapartum CTG database. *BMC Pregn. Childbirth* **14**(1), 16–16 (2014)
5. Comert, Z., Kocamaz, A.F.: Evaluation of fetal distress diagnosis during delivery stages based on linear and nonlinear features of fetal heart rate for neural network community. *Int. J. Comput. Appl.* **156**(4), 26–31 (2016)
6. Comert, Z., Kocamaz, A.F.: Fetal hypoxia detection based on deep convolutional neural network with transfer learning approach (2018)
7. Da, G., Luis, M., Costa-Santos, C., Ugwumadu, A., Schnettler, W.: Agreement and accuracy using the figo, acog and nice cardiotocography interpretation guidelines. *Acta Obstetrica et Gynecologica Scandinavica: Official Publication of the Nordisk Forening for Obstetrik och Gynekologi* (2017)
8. De Laulnoit, A.H., et al.: Automated fetal heart rate analysis for baseline determination and acceleration/deceleration detection: a comparison of 11 methods versus expert consensus. *Biomed. Sig. Process. Control* **49**, 113–123 (2019)
9. De Laulnoit, A.H., Boudet, S., Demailly, R., Peyrodie, L., Beuscart, R., De Laulnoit, D.H.: Baseline fetal heart rate analysis: Eleven automatic methods versus expert consensus 2016, pp. 3576–3581 (2016)
10. Devoe, L.D.: Nonstress testing and contraction stress testing. *Obstetr. Gynecol. Clin. North Am.* **26**(4), 535–556 (1999)
11. Devoe, L.D.: Nonstress and contraction stress testing. *The Global Library of Women's Medicine* (2009)
12. Doret, M., Spilka, J., Chudacek, V., Goncalves, P., Abry, P.: Fractal analysis and hurst parameter for intrapartum fetal heart rate variability analysis: a versatile alternative to frequency bands and lf/hf ratio. *PLOS ONE* **10**(8) (2015)

13. Georgieva, A., Abry, P., Chudacek, V., Djuric, P.M., Frasn, M.G., Kok, R., Lear, C.A., Lemmens, S.N., Nunes, I., Papageorgiou, A.T., et al.: Computer-based intrapartum fetal monitoring and beyond: a review of the 2nd workshop on signal processing and monitoring in labor (october 2017, oxford, uk). *Acta Obstetrica et Gynecologica Scandinavica* **98**(9), 1207–1217 (2019)
14. Georgoulas, G., Karvelis, P., Spilka, J., Chudáček, V., Stylios, C.D., Lhotská, L.: Investigating ph based evaluation of fetal heart rate (fhr) recordings. *Health and Technology* **7**(1) (2017)
15. Guideline, N.: Intrapartum care for healthy women and babies (2014)
16. Magenes, G., Signorini, M., Ferrario, M., Lunghi, F.: 2CTG2: a new system for the antepartum analysis of fetal heart rate, pp. 781–784 (2007). https://doi.org/10.1007/978-3-540-73044-6_203
17. Mohajer, M.P., Sahota, D.S., Reed, N.N., James, D.K.: Atrioventricular block during fetal heart rate decelerations. *Arch. Dis. Childhood-Fetal Neonatal Ed.* **72**(1), F51–F53 (1995)
18. Parer, J.T., Ikeda, T.: A framework for standardized management of intrapartum fetal heart rate patterns. *Am. J. Obstetr. Gynecol.* **197**(1), 26–e1 (2007)
19. Rotariu, C., Pasarică, A., Costin, H., Nemescu, D.: Spectral analysis of fetal heart rate variability associated with fetal acidosis and base deficit values. In: *International Conference on Development & Application Systems* (2014)
20. Southall, D., et al.: Prospective study of fetal heart rate and rhythm patterns. *Arch. Disease Childhood* **55**(7), 506–511 (1980)
21. Zafer, C., Fatih, K.A., Velappan, S.: Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment. *Comput. Biol. Med.* pp. S0010482518301458 (2018)
22. Zhao, Z., Deng, Y., Zhang, Y., Zhang, Y., Shao, L.: Deepfhr: intelligent prediction of fetal acidemia using fetal heart rate signals based on convolutional neural network. *BMC Med. Inform. Dec. Making* **19**(1) (2019)