ORIGINAL PAPER



An exploratory approach to fetal heart rate-pH-based systems

George Georgoulas¹ · Petros Karvelis² · Vaclav Chudacek³ · Jiri Spilka³ · Miroslav Burša³ · Lenka Lhotska³ · Chrysostomos D. Stylios² · George Nikolakopoulos⁴

Received: 21 May 2019 / Revised: 4 May 2020 / Accepted: 9 June 2020 © Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

This paper presents an exploratory approach of the fetal heart rate (FHR) analysis, aiming to highlight potential limitations of the current predictive modeling attempts. To do so, a set of features that are usually encountered in FHR analysis as well as features extracted using a variant of symbolic aggregate approximation were projected onto a lower-dimensional space where patterns can easily be discerned. The results show, both in a qualitative and a quantitative manner, that there is high overlap between the classes that are formed using solely the umbilical cord pH information, irrespective of the selected dimensionality reduction method. These findings suggest that there is probably a limit to the performance expectation of the current pH-based systems and that alternative approaches should be also pursued to enhance the utility of computer-based decision support technologies.

Keywords Dimensionality reduction · Exploratory analysis · Feature extraction · Fetal heart rate

1 Introduction

Electronic fetal monitoring (EFM) was introduced into clinical practice 40 years ago, and it has progressively become an indispensable means for fetal surveillance during delivery [1]. EFM mainly refers to the acquisition and interpretation of the cardiotocogram (CTG), which in turn depicts the fetal heart rate (FHR) and the uterine contractions (UC). CTG is believed to contain valuable information for fetal well-being, which an experienced clinician can retrieve by eye inspection and intervene in case they foresee an adverse outcome. However, the everyday use of CTG did not reduce the inter- and intra-observer variability [2], while it was also blamed for an increase in the number of unnecessary

Petros Karvelis pkarvelis@gmail.com

- ¹ DataWise Data Engineering LLC, 1938 Volberg St, Atlanta, GA 30318, USA
- ² Laboratory of Knowledge and Intelligent Computing, Department of Informatics and Telecommunications, University of Ioannina, 47100 Arta, Greece
- ³ CIIRC, Czech Technical, University in Prague, Prague, Czech Republic
- ⁴ Control Engineering Group, SRT Department, Luleâ University of Technology, 97187 Luleå, Sweden

cesarean sections [3]. This fact, along with the firm belief that CTG and especially the FHR convey "hidden" information which might not been fully exploited by eye inspection [4], led to the development of computerized systems. The ultimate goal of such a system was to remove any personal bias and establish a consistent interpretation, diminishing inter- and intra-observer variability, while at the same time explore the recent advances in signal processing and artificial intelligence (AI), without, however, living up to those early expectations [5].

Most of the computerized systems encountered in the literature rely on a supervised classification approach: (i) a set of features is extracted from the signal (a way to battle the "curse of dimensionality" and potentially lead to better results) and (ii) those features are subsequently used to train a classifier [6–8]. Initially, for the feature extraction stage, methods for automatically replicating the guidelines issued by the International Federation of Obstetricians and Gynecologists (FIGO) [4] were employed. Soon after, new features brought in from other mature fields were adopted and tested, e.g., time-domain features, frequency-domain features, etc. [9–18].

All these approaches assume that FHR tracings form concrete classes that can be learned by a classifier. One of the most common ways to create those target classes is by using the pH value of the umbilical cord artery [19]. However, a review of a number of studies revealed that there seems to be an upper limit to the performance of this kind of systems [20].

On the other hand, very few studies followed an unsupervised/exploratory approach, before attempting to develop predictive models. In [21, 22] clustering methods, also known as unsupervised classification methods were tested. In [21] a clustering by compression approach was pursued, and three clusters were identified along with a cluster with some cases that deviated from the identified pattern. The method was applied to a small number of cases and the retrospective labeling of the clusters primarily relied on Apgar score. In [22], the k-means algorithm was used, and the resulting two clusters seemed to capture the dynamics controlling non-pathological and intrauterine growth-restricted (IUGR) fetuses. In these approaches, the explicit assumption made was that a clustering structure indeed existed in the FHR recordings, even though they did not impose a predefined labeling.

In this current study, no assumption about the potential existence of classes/clusters was made. Instead, a number of dimensionality reduction techniques were applied to a set of features extracted from the largest, currently, freely available database [23] trying to understand the structure and the segregation of cases in the higher features space by having a look on lower projected spaces. The engaged features come from various domains as it is described in brief in Sect. 2 and are commonly used in computerized analysis of FHR. Moreover, a novel set of features, using symbolic aggregate approximation (SAX) [24], was also tested. For the dimensionality reduction stage, a linear and three non-linear techniques were selected in order to investigate quite different/representative techniques.

The main goal of this study was to investigate potential limitations of FHR classification when it is based solely on pH class formation, and the results presented in Sect. 3 and further discussed in Sect. 4 seem to corroborate that.

2 Methods

The whole study focuses on the reduction of the original high *N*-dimensional raw FHR time series, x[n], n = 1, 2, ..., N into just two dimensions that can be perceived by the human brain.¹ The basic procedure was performed in two stages. In the first stage, the FHR signal was transformed into a



Fig. 1 Overall procedure: **a** preprocessing and artifact removal were followed by feature ranking and selection; the feature selection stage was split into three branches with each one using a different classifier for the evaluation of the quality of the various feature sets. The best candidates for each one of the three branches were projected onto 2D using the selected dimensionality reduction method and the one with the higher inter/intra-class distance was selected, **b** preprocessing was followed by normalization, PAA and symbolization, leading to the "bag of patterns" representation. After that, the procedure is identical to the one followed in **a**

feature vector with usually much lower dimension, while in the second stage the feature vector was embedded into an even lower dimension space.

Before the application of the dimensionality reduction, the FHR signal was preprocessed to remove as much as possible of the contaminating artifacts. Then, two different feature extraction methods were tested. The first one computed a set of "conventional" features followed by feature selection. The second feature extraction approach transformed the original FHR signal into a symbolic representation that was then treated using a technique borrowed from the field of information retrieval.

The methods involved in the second stage belong to the family of dimensionality reduction methods. This family of methods is quite large and quite diverse. The scope of this work was not to exhaustively search for the "best" method to transform the data, rather than to check if different methods can lead to similar results/conclusions. Therefore, this work employed four dimensionality reduction methods: (a) the most used one of linear transformations, (b) one of its nonlinear counterparts, (c) a popular nonlinear method from the manifold learning family and (d) a method that is based on a neural network approach. The overall procedure for the two different feature sets is depicted in Fig. 1. As it can be

¹ In two and three dimensions, the human eye is an excellent pattern recognition tool and easily perceives the complexity of a potential classification task and can also discern the existence of structures/patterns in the data. However, the use of three-dimensional (3D) scatter plots was not exploited, to avoid the extra complexity imposed by the selection of the best viewpoint.

seen, the paths are similar and only the processing of the FHR was different.

2.1 Preprocessing-"conventional" feature extraction

FHR is a very noisy signal with the noise appearing as spiky artifacts and with a lot of missing values due to the displacement of the transducer. In this work, an artifact rejection scheme based on thresholding and interpolation was used [13].

As it was mentioned in the Introduction, numerous features have been used for the analysis of FHR having their origins in different domains. In this study, a total of 54 "conventional" features were extracted (21 unique methods with different parameter settings). For a detailed description of the involved features, the interested reader can refer to our previous works [11, 13, 19].

2.2 Feature selection

Sometimes, there is a tendency to extract more features than necessary or extract features that have little discriminative power. In such cases, a feature selection stage can be beneficial, increasing the generalization capability of the prediction system. Feature selection methods can be grouped basically into three categories: filters, wrappers, embedded methods and their combinations. Our previous work [13] showed that the classification accuracy of an FHR signal analysis system can be improved by the selection of a subset of the original feature set. Moreover, an initial set of experiments revealed that including all the extracted features leads to more overlapping between the classes in the lower-dimensional space. Therefore, a feature ranking method (filter approach) was combined with a nested subset method (wrapper approach) to derive a suitable candidate set of features before the dimensionality reduction stage. Such methods are fast and simple to implement and at the same time have been proven to work well in practice.

For ranking, a composite approach was considered based on the area under the receiver operating characteristic (ROC) curve (AUC), taking also into account correlations among selected features [25]. The advantage of AUC is that it is immune to class imbalance [26]. The employed approach penalizes features that are correlated with the ones already ranked higher before them. This way it forces selected features not only to have a high individual AUC value but also to be as less correlated as possible with each other. After ranking, the simplest wrapper based approach for the selection of the subset of features was used: nested subsets of features were formed, $S_1 = \{f_1\}, S_2 = \{f_1, f_2\}, \dots, S_{54} = \{f_1, f_2, \dots, f_{54}\}$, where f_j denotes the *j*th selected feature. Therefore, 54 such subsets were created, ranging from one feature to all 54 features and the performance of each subset was evaluated using a simple classification algorithm. The subset with the best performance was selected to undergo further dimensionality reduction.

2.3 Features based on symbolic manipulation of FHR

The use of symbolic time series analysis tools and representations is quite competitive to conventional methods for signal processing [27, 28] and very popular in the time series data mining community. One of the reasons for its popularity is that after the transformation of the real-valued time series into a sequence of symbols, tools from the mature field of text mining and information retrieval can be easily adopted. Among the various techniques for transforming a real-valued time series into a sequence (string) of symbols, SAX [24] is probably one of the most popular ones.

The SAX algorithm creates an approximation of a realvalued times series $x = \{x[1], x[2], ..., x[N]\}$ of an arbitrary length N by converting it to a string of symbols of arbitrary length w, where w < N. The entire process consists of the following steps:

2.3.1 Normalization

Before any further processing, the original time series is normalized to have zero mean and unit variance. This (for reallife data) usually leads to distributions of the transformed time series that are very close to the Gaussian distribution, which makes it easier for the discretization process [29].

2.3.2 Piecewise aggregate approximation (PAA)

PAA, independently introduced by Keogh et al. [30] and by Yi and Faloutsos [31], offers a means to produce a low(er) dimensionality representation of a time series. More specifically, a time series x of length N is transformed into a new time series $x_{PAA} = {\bar{x}[1], \bar{x}[2], ..., \bar{x}[w]}$ of length w, with:

$$\bar{x}[i] = \frac{w}{N} \sum_{\substack{j=\frac{N}{w}(i-1)+1}}^{\frac{N}{w}i} x[j], \quad i = 1, 2, \dots, w.$$
(1)

2.3.3 Discretization/symbolization

After the PAA step, a symbolic representation is created by partitioning the original continuous space. Since the original time series has been normalized, the new time series will approximately follow a Gaussian distribution and the "breakpoints" can be easily determined so as to produce equal-sized areas under a Gaussian curve.

2.3.4 Feature extraction from a SAX representation

The discretization stage creates a string, which needs to be transformed into a representation that can be handled easier by a conventional classification algorithm. The "bag of patterns" representation borrows ideas from the well-established field of text processing and the "bag of words" representation, treating the created string as a text document. The method counts the appearance of individual "words" in a string and creates a histogram of appearances for each time series. More specifically, the approach uses a sliding window of length n within which it applies SAX to derive a word of length m. With these individual words, it creates a feature vector of length A^w where A is the length of the alphabet that holds the frequency of appearance of each word. This representation can be coupled with any conventional classification algorithm.

2.4 Dimensionality reduction methods

Since the introduction of principal component analysis (PCA) [32], many different linear and nonlinear techniques have been proposed. In the rest of this section, the four methods employed in this study are briefly presented: (a) classical multidimensional scaling (MDS), a linear technique which has been proven to be highly competitive for real-life data [33, 34], (b) Sammon "mapping," one of the oldest nonlinear techniques, member also of the general MDS family, with successful applications as part of a classification scheme [35], (c) Isomap (isometric mapping), a nonlinear method coming from the family of manifold learning methods, developed to improve MDS, which has been successfully applied to gait analysis [36], and (d) curvilinear component analysis (CCA), which comes from the field of artificial neural network approaches aiming to be an improvement of the self-organizing maps (SOMs).

2.4.1 Multidimensional scaling

MDS is practically a set of techniques [32], both of linear and nonlinear nature with classical MDS being of the first variety. All members of the MDS family are based on the same principle/goal: retain the pairwise distances between the data points as much as possible. The quality of the mapping, which at the same time drives the algorithm for finding the most appropriate transformation, is called the stress function. In the case of classical MDS, the stress function is:

$$Str_{MDS}(Y) = \sum_{i} \sum_{j} \left(\left\| x_{i} - x_{j} \right\| - \left\| y_{i} - y_{j} \right\| \right)^{2},$$
(2)

where $x_i \in \Re^m$, $1 \le i \le N$ are the original data points in the higher dimensional space and $y_i \in \Re^d$, $1 \le i \le N$ are the reconstructed data points in the lower-dimensional space (m > d). The optimization (minimization) of the above stress function can be easily done using linear algebra operations.

2.4.2 Sammon

Sammon "mapping" [37] is one of the nonlinear variants of the MDS family where the stress function is given by:

$$\operatorname{Str}_{\operatorname{Sammon}}(Y) = \frac{1}{\sum_{i} \sum_{j} \left\| x_{i} - x_{j} \right\|} \cdot \frac{\sum_{i} \sum_{j} \left(\left\| x_{i} - x_{j} \right\| - \left\| y_{i} - y_{j} \right\| \right)^{2}}{\left\| x_{i} - x_{j} \right\|}.$$
(3)

The minimization of the stress function is usually performed by using a pseudo-Newton method.

2.4.3 Isomap

Isomap belongs to the class of manifold learning techniques, which assume that the data lie on a submanifold of the original space [38]. Isomap attempts to preserve pairwise geodesic distance between data points—the distance between two points measured over the manifold—using graph distances to approximate it and then applies classical MDS to the created distance matrix.

2.4.4 Curvilinear component analysis

CCA is a kind of SOM that performs vector quantization in the input space and then nonlinear projection in the low dimensional space [39]. However, for small data sets, the quantization phase can be skipped (as in our case). CCA combines concepts both from SOM and MDS (and especially the Sammon variant) and in some cases, can provide the true structure of the manifold, while SOM and MDS might fail.

As in the case of MDS, CCA minimizes a stress function:

$$\operatorname{Str}_{\operatorname{CCA}}(Y) = \sum_{i} \sum_{j} \left(\left\| x_{i} - x_{j} \right\| - \left\| y_{i} - y_{j} \right\| \right)^{2} \cdot F\left(\left\| x_{i} - x_{j} \right\|, \lambda \right).$$
(4)

In Eq. 4, the Euclidian distance is explicitly used, but other distance metrics can be used. The weighting function $F(\cdot, \cdot)$ is typically chosen as a monotonically decreasing function of its arguments, such as the decreasing exponential:

$$F\left(\left\|x_{i}-x_{j}\right\|,\lambda\right) = \exp\left(-\frac{\left\|x_{i}-x_{j}\right\|}{\lambda}\right)$$
(5)

where λ is a parameter similar to the neighborhood radius in SOM. The minimization of the stress function is performed using a variant of the gradient descent method.

3 Experimental results

Both "conventional" as well as features based on the "bag of patterns" approach were extracted from FHR recordings. In both cases, the goal was the investigation of the structure of the feature space, by looking into a projection of it in a lower two-dimensional setting, in relation to the measured pH values. The intention was to explore whether there are limitations in the use of pH values for the class formation. Therefore, "good" if not "optimal" configurations of the original high dimensional space were sought using separability and classification measures in the intermediate stages, before projecting the selected representations into the lowerdimensional ones.

3.1 Data set

All the experiments were carried out using the CTU-UHB database [23]. The database consists of 552 records acquired using STAN and Avalon devices between years 2009 and 2012 at the obstetrics ward of the University Hospital in Brno, Czech Republic. The features were systematically extracted on 30 min long FHR signals at the end of the first stage of labor. Next, a pH threshold value of 7.05 [13, 19] was used to divide the recordings into "normal" and "pathological" ones. This value is commonly selected in pH-based classification systems.

3.2 Conventional features

After the extraction of the 54 features, the feature selection procedure was applied. For the nested subset selection part, a classification algorithm is needed. Three simple classification algorithms were used to select a discriminative feature set (a discriminative feature set usually leads to less overlapping between the classes in the lower-dimensional space), with respect to the classes formed using the pH value. The selection of simple classifiers instead of a more powerful one was done to avoid overfitting, since the whole data set was used both for training and performance assessment. (Note: this study was not interested in building an unbiased classification system rather than explore if the structure of feature space can actually support the development of such a system).

Two simple nonparametric classifiers, the minimum Mahalanobis distance classifier, a quadratic classifier and a simple parametric classifier, the *k*-nearest neighbor (*k*-nn) classifier [25] were selected.

Because the use of the 7.05 pH threshold creates two classes that are quite imbalanced (508/44) and the aforementioned classifiers cannot cope with that, the synthetic minority oversampling technique (SMOTE) was used, to alleviate the problem [40]. It must be noted that the synthetic samples were not used during the evaluation of the performance, which was done using the *g*-mean measure [13], which is also rather insensitive to class imbalance.

For each of the three simple classifiers, the "best" feature set, identified in terms of its performance, was subsequently transformed/reduced to two dimensions using one of the four dimensionality reduction algorithms. Therefore, three low dimensional projections (one for each one of the three simple classifiers) were created. Among the three, the one with the largest separability between the two classes was selected for further assessment. The measure of separability used is given below [25]:

$$J = \text{trace}\left(S_{w}^{-1}S_{b}\right) \tag{6}$$

where S_{w} is the within scatter matrix

$$S_{\rm w} = \frac{1}{M} \sum_{i=1}^{2} \sum_{m=1}^{M_i} \left(x_{i,m} - \mu_i \right) \left(x_{i,m} - \mu_i \right)^{\rm T}$$
(7)

 $S_{\rm b}$ is the between scatter matrix

$$S_{\rm b} = \frac{1}{M} \sum_{i=1}^{2} M_i (\mu - \mu_i) (\mu - \mu_i)^{\rm T}.$$
 (8)

 μ_i is the mean vector of class *i*, μ is the overall mean vector (both classes put together), and $x_{i,m}$ is the *m*th data point belonging to class *i*, M_i is the number of examples belonging in class *i* (*i*=1, 2) and $M_1 + M_2 = M$.

The results for each one of the four dimensionality reduction procedures are depicted in Fig. 2. As it can be seen, even though the four methods create slightly different representations, for all of them, a part of the space seems to belong primarily to the normal cases with very few pathological ones. The rest of the space is occupied both by pathological and normal cases, with the normal cases in this second part being, however, less dense compared to the former part. For example, for the case of the classic MDS (top left of Fig. 2), the left side of the scatter plot is occupied mainly be normal cases with few pathological ones, while the right part has relatively fewer normal cases and more pathological ones.

To have a quantitative measure of the projected feature space, the two-dimensional data sets were fed to a least squares support vector machine (LSSVM) [41] with radial basis function (RBF) Kernels, which was tuned using five-fold cross-validation. The produced separating boundaries are depicted in Fig. 3. From these figures, it is clear that the two classes are highly mixed and not well separated, as it is also confirmed by Table 1 which summarizes the performance measures for the LSSVM.

A better insight is given by Fig. 4, which depicts the data using as extra visualization information the actual



Fig.2 The two-dimensional projection of the conventional features for the four algorithms. Starting from the top left corner and moving clockwise: classical MDS, Sammon mapping, Isomap and CCA. Red circles correspond to pathological cases, while blue "xs" correspond to normal cases (pH > 7.05) (color figure online)



Fig.3 The decision boundaries drawn by an LSSVM classifier. **a** classical MDS, **b** Sammon mapping, **c** Isomap and **d** CCA (magenta regions corresponding to normal areas pH > 7.05) (color figure online)

pH value of each case. From this figure, it can be seen that even though there seems to be a region where the "blue" color prevails (normal cases), this is not done in a uniform manner. More importantly, the "red" colored pathological cases can lie very close to normal cases indicating

 Table 1
 Performance of the LSSVM on the two-dimensional projected space

	TP rate	TN rate	g-mean
MDS-conv	0.727	0.677	0.702
Sammon-conv	0.727	0.689	0.708
Isomap-conv	0.750	0.667	0.707
CCA-conv	0.682	0.691	0.686
MDS-SAX	0.659	0.656	0.657
Sammon-SAX	0.705	0.634	0.668
Isomap-SAX	0.636	0.697	0.666
CCA-SAX	0.705	0.644	0.673

Bold values indicate the highest score



Fig. 4 Scatter plots of the two-dimensional projection of the conventional features with the individual pH value depicted for each data point. Starting from the top left corner and moving clockwise: classical MDS, Sammon mapping, Isomap and CCA

that there will always be false alarms as well as some hit misses.

3.3 "Bag of patterns" representation

For the SAX-derived features, a slightly different version from the original formulation was applied, since the absolute value of the FHR is considered an indicator of fetal well-being. Therefore, instead of locally applying SAX in a sliding window of length n, all signals were put together and globally normalized. Then, for each one of them, SAX was applied creating a long string of length w which was subsequently used to create a "bag of patterns" representation by using a sliding window of length m (the size of the word).

As in the case of the "conventional" features, different "bag of patterns" representations (size of the alphabet,



Fig. 5 The two-dimensional projection of the "bag of patterns" representation for the four algorithms. Starting from the top left corner and moving clockwise: classical MDS, Sammon mapping, Isomap and CCA. Red circles correspond to pathologic cases, while blue "xs" correspond to normal cases (pH > 7.05) (color figure online)

size of the produced string and length of the word) were tested. These different representations were fed to the same three simple classifiers (linear, quadratic and k-nn)—and SMOTE was also applied. For each classifier, the "best" feature set (in terms of classification performance) was projected onto two dimensions resulting into three representations (one for each classifier). Among the three representations, the one that corresponded to the larger separation between the two classes (inter/intra-class distance criterion) was selected.

The results are summarized in Figs. 5, 6 and 7. Again, the same pattern emerges (even though the resulting formations seem to be more complex compared to those created by the projection of the conventional features, with higher overlap): an area where the normal cases prevail with some pathological cases among them. Also, some less dense areas where the pathological cases make their presence more notable, however, with normal cases lying in close vicinity or even further away toward areas that someone would expect to be occupied by more severe cases (lower pH values). For example, in the case of Sammon mapping (Fig. 5), the area corresponding primarily to normal cases occupies the upper part of two-dimensional space, whereas the pathological cases are mainly concentrated in the lower part. This is further illustrated by the application of the LSSVM approach in Fig. 6. Finally, in Fig. 7 as in the case of the conventional features (Fig. 4), we do not have a "smooth" transition from "blue" to "red," which means that "perfect" separation within this setting is very difficult, if not impossible.



Fig. 6 The decision boundaries drawn by an LSSVM classifier. Starting from the top left corner and moving clockwise: classical MDS, Sammon mapping, Isomap and CCA, magenta regions corresponding to normal areas (color figure online)



Fig. 7 Scatter plots of the two-dimensional projection of the "bag of patterns" representation with the individual pH value depicted for each data point. Starting from the top left corner and moving clockwise: classical MDS, Sammon mapping, Isomap and CCA

4 Conclusions

In this study, an exploratory approach of the FHR-pH "correlation" was pursued in an attempt to investigate the viability of developing diagnostic systems for the period of delivery relying on features extracted from the FHR and calibrated solely using the umbilical cord pH. Since previous studies [13, 14, 19] have shown that we might be reaching a saturation limit of the performance of such systems, this study tried to shed light to the structure of

the problem. To do so, a large set of conventional features as well as features that are rarely used in FHR processing were tested. Then, techniques from the field of dimensionality reduction were applied to derive a representation of the problem in a space tractable by human perception.

Four different dimensionality reduction methods were tested. It turns out that even though the four methods returned a bit different "portraits" of the data in the low dimensional space they did share some common elements: (a) an area where the normal cases prevail is present in almost all embeddings, (b) as we move away from that area pathologic cases appear more frequently, however, scattered among them are normal cases and (c) there does not seem to be a smooth/gradual transition between the two areas.

From Figs. 4 and 7, it was made clear that no matter the threshold for normality, there will always be several misclassified cases. The results suggest that there is some structure in the way cases are scattered, but that structure is not enough for building an automatic system that can predict pH related classes with high classification performance since the overlap is quite high.

Regarding the two feature sets, the "conventional" ones seem to better reflect the pH information. This comes as no surprise, since these features are the result of years of research. However, the "bag of words" features also tell the same story, even with a bit higher overlap (lower performance as can be seen from Table 1).

This study does not intend to argue that in higher dimensions a better boundary could not be found. However, the observed limitation of the FHR–pH underlying structure indicates a few open issues for the research community; additional clinical information should be included as part of the feature set and/or alternative labeling process should be considered, keeping also in mind that it is not natural that there would be a simple separating line (pH based) between the normal and pathological FHRs group. Toward the latter, a latent class analysis (LCA) model for aggregating experts' opinion was proposed lately [42]. In the future work, a similar exploratory analysis, in light of the FHR-LCA framework, will be investigated.

References

- Chen, H.Y., Chauhan, S.P., Ananth, C.V., Vintzileos, A.M., Abuhamad, A.Z.: Electronic fetal heart rate monitoring and its relationship to neonatal and infant mortality in the United States. Am. J. Obstet. Gynecol. 204, 491.e1–491.e10 (2011)
- Blackwell, S.C., Grobman, W.A., Antoniewicz, L., Hutchinson, M., Bannerman, C.G.: Interobserver and intraobserver reliability of the NICHD 3-tier fetal heart rate interpretation system. Am. J. Obstet. Gynecol. 205(4), 378.e1–378.e5 (2011)
- Alfirevic, Z., Devane, D., Gyte, G.M.: Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for

fetal assessment during labour. Cochrane Database Syst. Rev. 3, CD006066 (2006)

- FIGO: Guidelines for the use of fetal monitoring. Int. J. Gynaecol. Obstet. 25, 159–167 (1986)
- Balayla, J., Shrem, G.: Use of artificial intelligence (AI) in the interpretation of intrapartum fetal heart rate (FHR) tracings: a systematic review and meta-analysis. Arch. Gynecol. Obstet. 300, 7–14 (2019)
- Dash, S., Quirk, J.G., Djurić, P.M.: Fetal heart rate classification using generative models. IEEE Trans. Biomed. Eng. 61(11), 2796–2805 (2014)
- Petrozziello, A., Jordanov, I., Papageorghiou, A., Redman, W.G., Georgieva, A.: Deep learning for continuous electronic fetal monitoring in labor. In: Conference on Proceedings IEEE Engineering in Medical and Biology Society, pp. 5866–5869 (2018)
- Feng, G., Quirck, J, Djuric, P.: Supervised and unsupervised learning of fetal heart rate tracings with deep Gaussian processes. In: Presented at 14th Symposium on Neural Networks and Applications (NEUREL) (2018)
- Czabanski, R., Jezewski, M., Wrobel, J., Jezewski, J., Horoba, K.: Predicting the risk of low-fetal birth weight from cardiotocographic signals using ANBLIR system with deterministic annealing and epsilon-insensitive learning. IEEE Trans. Inf Technol. Biomed. 14(4), 1062–1074 (2010)
- Georgoulas, G., Stylios, C.D., Groumpos, P.P.: Predicting the risk of metabolic acidosis for newborns based on fetal heart rate signal classification using support vector machines. IEEE Trans. Biomed. Eng. 53(5), 875–884 (2006)
- Spilka, J., Chudáček, V., Koucký, M., Lhotská, L., Huptych, M., Janků, P., Georgoulas, G., Stylios, C.: Using nonlinear features for fetal heart rate classification. Biomed. Signal Process. Control 7(4), 350–357 (2012)
- Fanelli, A., Magenes, G., Campanile, M., Signorini, M.G.: Quantitative assessment of fetal well-being through CTG recordings: a new parameter based on phase-rectified signal average. IEEE J. Biomed. Health Inform. 17(5), 959–966 (2013)
- Georgoulas, G., Karvelis, P., Spilka, J., Chudáček, V., Stylios, C.D., Lhotská, L.: Investigating pH based evaluation of fetal heart rate (FHR) recordings. Health Technol. 7, 1–14 (2017)
- Spilka, J., Frecon, J., Leonarduzzi, R., Pustelnik, N., Abry, P., Doret, M.: Sparse support vector machine for intrapartum fetal heart rate classification. IEEE J. Biomed. Health Inform. 21(3), 664–671 (2017)
- Georgieva, A., Payne, S.J., Moulden, M., Redman, C.W.: Artificial neural networks applied to fetal monitoring in labour. Neural Comput. Appl. 22(1), 85–93 (2013)
- Czabański, R., Jeżewski, J., Horoba, K., Jeżewski, M.: Fetal state assessment using fuzzy analysis of fetal heart rate signals—agreement with the neonatal outcome. Biocybern. Biomed. Eng. 33(3), 145–155 (2013)
- Ocak, H., Ertunc, H.M.: Prediction of fetal state from the cardiotocogram recordings using adaptive neuro-fuzzy inference systems. Neural Comput. Appl. 23(6), 1583–1589 (2013)
- Dash, S., Quirk, J.G., Djuric, P.M.: Learning dependencies among fetal heart rate features using Bayesian networks. In: Engineering in Medicine and Biology Society (EMBC), pp. 6204–6207 (2012)
- Norén, H., Amer-Wåhlin, I., Hagberg, H., Herbst, A., Kjellmer, I.: Fetal electrocardiography in labor and neonatal outcome: data from the Swedish randomized controlled trial on intrapartum fetal monitoring. Am. J. Obstet. Gynecol. 188(1), 183–192 (2003)
- Spilka, J.: Complex approach to fetal heart rate analysis: a hierarchical classification model. Ph.D. Thesis, Czech Technical University in Prague Department of Cybernetics (2013)
- Costa Santos, C., Bernardes, J., Vitányi P.M., Antunes, L.: Clustering fetal heart rate tracings by compression. In: Computer-Based Medical Systems, CBMS, pp. 685–690 (2006)

- Ferrario, M., Signorini, M.G., Magenes, G.: Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses. Med. Biol. Eng. Comput. 47(9), 911–919 (2009)
- Chudacek, V., Spilka, J., Bursa, M., Janku, P., Hruban, L., Huptych, M., Lhotska, L.: Open access intrapartum CTG database. BMC Pregnancy Childbirth 14, 16 (2014)
- Lin, J., Keogh, E., Lonardi S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, pp. 2–11 (2003)
- 25. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 4th edn. Academic Press, San Diego (2009)
- Wasikowski, M., Chen, X.: Combating the small sample class imbalance problem using feature selection. IEEE Trans. Knowl. Data Eng. 22(10), 1388–1400 (2010)
- Lin, J., Keogh, E., Wei, L., Lonardi, S.: Experiencing SAX: a novel symbolic representation of time series. Data Min. Knowl. Discov. 15(2), 107–144 (2007)
- Yousofvand, L., Abdolhossein, F., Fardin, A.: Person identification using ECG signal's symbolic representation and dynamic time warping adaptation. Signal Image Video Process. 13(2), 245–251 (2019)
- Keogh, E., Lin J., Fu., A.: Hot sax: efficiently finding the most unusual time series subsequence. In: 5th IEEE International Conference on Data Mining, pp. 226–233 (2005)
- Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Dimensionality reduction for fast similarity search in large time series databases. J. Knowl. Inf. Syst. 3(3), 263–286 (2000)
- Yi, K., Faloutsos, C.: Fast time sequence indexing for arbitrary Lp norms. In: 26th International Conference on Very Large Databases, Cairo, Egypt, (2000)
- Cox, T.F., Cox, M.A.: Multidimensional Scaling. CRC Press, Boca Raton (2010)

- van der Maaten, L.J., Postma, E.O., van den Herik, H.J.: Dimensionality reduction: a comparative review. J. Mach. Learn. Res. 10, 66–71 (2009)
- Machado, J.A., Tenreiro, E., Dumitru, B.: Analysis of UV spectral bands using multidimensional scaling. Signal Image Video Process. 9(3), 573–580 (2015)
- Lerner, B., Guterman, H., Aladjem, M., Dinsteint, I., Romem, Y.: On pattern classification with Sammon's nonlinear mapping an experimental study. Pattern Recognit. 31(4), 371–381 (1998)
- Preben, F., Moeslund, T.: Invariant gait continuum based on the duty-factor. Signal Image Video Process. 3(4), 391 (2009)
- Sammon, J.: A nonlinear mapping for data structure analysis. IEEE Trans. Comput. 18(5), 401–409 (1969)
- Tenenbaum, J., De Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
- Demartines, P., Hérault, J.: Curvilinear component analysis: a selforganizing neural network for nonlinear mapping of data sets. IEEE Trans. Neural Netw. 8(1), 148–154 (1997)
- Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357 (2002)
- 41. Suykens, J., Vandewalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**, 293–300 (1999)
- Spilka, J., Chudáček, V., Janků, P., Hruban, L., Burša, M., Huptych, M., Zach, L., Lhotská, L.: Analysis of obstetricians' decision making on CTG recordings. J. Biomed. Inform. 51, 72–79 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.