Unsupervised Detection of Anomalies in Fetal Heart Rate Tracings using Phase Space Reconstruction

Liu Yang,[◇] Marzieh Ajirak,[◇] Cassandra Heiselman,[*] J. Gerald Quirk,[*] Petar M. Djurić,[◇]
[◇]Department of Electrical and Computer Engineering, Stony Brook University
[*]Department of Obstetrics, Gynecology and Reproductive Medicine, Stony Brook University
Stony Brook, NY 11794, USA

***Abstract*—Detection of anomalies in time series is still a challenging problem. In this paper, we provide a new approach to unsupervised detection of anomalies in time series based on the concept of phase space reconstruction and manifolds. We propose a rotation-insensitive metric for quantifying the similarity of manifolds and a method that uses it for estimating the probability of an outlier. The proposed method does not rely on any features and can be used for signals with variable lengths. We tested it on both synthetic signals and real fetal heart rate tracings. The method has promising performance and can be used for interpreting the severity of fetal asphyxia.**

I. Introduction

Anomaly detection in time series has attracted attention in many disciplines over the past decades. In a recent review, it has been reported that in the past, much effort has been put on various kinds of problems related to outlier detection in time series including work with different data categories, outlier types, and nature of tasks [1]. In practical problems, the types of outliers can represent a couple of unexpected points in a data stream, several rare events embedded in time series, or an entire anomalous observation sequence. Although there are plenty of works for detecting odd points in time sequences [1], the detection of consecutive abnormal samples or an overall anomalous sequence has not been explored well. However, such problems are not uncommon in many applications.

This paper focuses on the detection of anomalies in a set of unlabeled time series data. The intended application is improved interpretation of cardiotocography (CTG) recordings, and in particular fetal heart rate (FHR) tracings. CTG monitoring has been widely used in delivery rooms for alerting physicians of inadequate level of oxygen transported to a fetus through the umbilical cord in the process of delivery. The traditional approach based on inspecting FHR patterns visually is unsatisfactory because of the high inter- and intra-variability of the obstetricians' decision making [2]. In the past decade, there has been much effort to explain FHR tracings by appealing to machine learning techniques. They include feature-based, contraction-dependent and neural network-based approaches [3]. Almost all of the reported methods require a set of known "healthy" or "non-healthy" labels of recordings, which are usually determined according to the umbilical cord pH value acquired after birth.

The authors thank the support of NIH under Award RO1HD097188-01.

Notwithstanding the extensive studies of FHR assessment aided by labels, experts have not reached an agreement on the thresholds that classify the fetuses, and there is even no settlement about the rationality of using the pH value for deciding the labels of the fetal status. Clinical data, such as pH values, indeed contain information, but they are all of newborns and in principle, one might argue, they should not be used for evaluating the fetal health during the last two hours before delivery. In other words, labels that represent the real-time status of fetuses do not exist. With this in mind, unsupervised learning from FHR tracings is particularly important for fetal monitoring. Even so, there are only a few related papers on this subject. For example, [4] explored unsupervised clustering of FHR by using deep Gaussian processes (GPs). In the paper, it was shown that after five-layers of space projections, the error of classification was reduced, and FHR tracings of unhealthy fetuses could be separated from the normal ones. However, in this work, the authors used only 10 recordings of the last 30 minutes of tracings.

In this paper, we develop a new method for detecting dissimilar time series from a pool of series. The method is different from previous works by the way how we use mapping and clustering. A common strategy for detecting outliers is extracting meaningful features and mapping them to a lower dimensional space, e.g., by using principal component analysis (see [5] for details). The method proposed in [6] requires features and is based on measuring the deviations within or among the cluster centroids obtained from features. Similarly, [7] reformulated the task of outlier detection as optimization of clustering, where entropy and dynamic warping of the time series were used. The authors in [8], extracted representative shapes of the normal class from datasets with inferred labels and the anomalies were learned at the same time. In our work, we map the time series into a Euclidean space. But instead of extracting features or inferring labels, we reconstruct the manifolds of time series in a 3D phase space, and design a rotation-insensitive metric of manifold-to-manifold distances for quantifying their similarity. Then we employ a distance-based probability estimation of anomaly detection. First, we test the proposed method on synthetic data and then we discuss its application on discovering anomalies among 85 FHR tracings.

II. Background

A. GP-based Phase Space Reconstruction

The phase space of a dynamic system contains infinite number of potential states forming a smooth manifold. An attractor of a system in phase space carries its all topological information and features. The manifold of a dynamic system in the real world is usually hidden. In order to recover manifolds from observed signals, Floris Takens proposed a powerful theorem, which shows that generically the hidden manifold can be reconstructed by a visible time series observation under a couple of conditions [9].

Theorem II.1 (**Takens' Theorem**). *Let M be a compact manifold of (integer) dimension d and suppose the generic pairs (ϕ, y) satisfy*

- *$\phi : M \to M$ is a C^2-diffeomorphism of M in itself.*
- *$y : M \to \mathbb{R}$ is a C^2-differentiable function.*

Then, the map $\Phi_{(\phi,y)}$: $M \to \mathbb{R}^{2d+1}$ given by

$$\Phi_{(\phi,y)} := [y(t), y(\phi(t)), y(\phi^2(t)), \ldots, y(\phi^{2d}(t))]$$

is an embedding of M in $\mathbb{R}^{2d+1}$.

The most popular generic function ϕ has a time-delay form $\Phi = [y(t), y(t-\tau), y(t-2\tau), \ldots, y(t-2d\tau)]$, which shows that a shadow version of the original manifold can be rebuilt simply by using one of the time series projections and its delays. Although Takens theoretically proves that for recovering a d-dimensional manifold, an $E = 2d+1$-dimensional delay embedding is sufficient, the actual values of d or E are unknown for the desired but latent M. Moreover, the choice of time delay τ also affects the performance of the phase space reconstruction (PSR). The traditional strategy for determining the optimal E and τ is by grid searching [10]. A newly proposed GP-based method for PSR has been shown to have greater tolerance for noise and to be more efficient for learning manifolds from highly correlated initial delay embedding [11]. The method also does not rely on searching for the optimal parameters for embedding.

We pursue the idea that manifolds rebuilt by using signals from the same dynamic system should have similar topology. We apply the GP-based method for PSR to infer the latent manifolds of signals and then, we cast the problem of detecting unusual time series as one of detecting irregular manifolds.

B. Local Outlier Probabilities (LoOP)

Among the many unsupervised outlier detection techniques, the one published in [12] is one of the most popular methods. It produces a probability score for each object in a set based on their local density. This method depends on a reliable distance function between objects or a known distance matrix consisting of all pairwise distances. The n nearest neighbors of an object create a local scope for computing a probabilistic distance from the object to its neighbors. A reference distance of an object is the expectation of the probabilistic distances of its neighbors to their neighbors. Then, an outlier factor can be generated by comparing the local probabilistic distance and the reference distance. Once all the factors are obtained, the local outlier probabilities can be obtained through normalization.

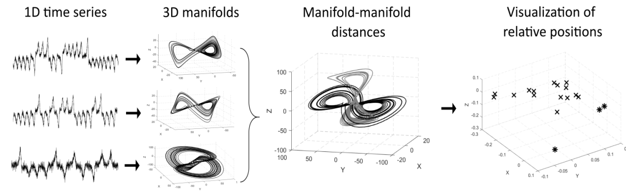

Fig. 1: A schematic diagram of the proposed method.

III. Problem Description

Since the fetuses with hypoxia are relatively rare compared to a large number of healthy fetuses, in this paper, we focus on the problem of anomalous time series detection that does not rely on a labeled training dataset. Suppose we are given a time series data set $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_N\}$ with N time series whose time lengths may be different, i.e., $\mathbf{s}_i \in \mathbb{R}^{t_i}$, $i = 1, 2, \ldots, N$. Our assumption is that most of the time series come from the same dynamic system, and we would like to detect the ones that show dissimilar dynamic characteristics (the outliers in our set).

IV. The Proposed Method

In order to select the anomalous signals from a data set, the intuitive approach is to rely on a metric of similarity. However, when the lengths of the time series are different, especially when the observations are not synchronous, the similarity simply calculated based on the time-domain variability is very challenging. A common solution is extracting meaningful features adapted to the addressed problem and reducing the dimensionality from the feature space to a latent space. In this paper, we consider inferring the latent manifolds of the dynamic systems from the signals generated by the systems. The topology of the manifold would imply the characteristics of the dynamic system. Therefore, we propose to first map the 1D time series to a 3D phase space to reconstruct the corresponding manifold. Then we use a newly defined metric for quantifying the similarity between the manifolds. In the last step, we estimate the probability of the observed time series being an outlier. A visualization of the relative positions of the data in the Euclidean space is also provided. Figure 1 shows a flow of the proposed method.

The steps of the method are intertwined. The performance of the previous steps affect directly the performance of the subsequent steps. For the PSR, we use the GP-based inference method proposed in [11] because it is within the Bayesian framework, is more robust and is more efficient than traditional delay embedding. It is worth pointing out that the phase spaces reconstructed by different times of training are not necessary the same. This is because the machine is blind to the real latent space [13]. Specifically, the latent variable learned from a high-dimensional observation might be a product of its real value and a rotation matrix, which means that there may be an unknown rotation angle between two manifolds. For this reason, some existing metrics for measuring manifold-to-manifold distances do not work well in our study, for example [14]. In this paper, we propose a new approach to quantifying the similarity between manifolds that is not sensitive to their rotation. The details of the metric are explained in subsection IV-A. After computing the distance, we apply LoOP [12] to estimate the probability that a time series is anomalous. In subsection IV-B, we provide a position visualization of a time series.

A. Quantifying a Manifold-to-Manifold Distance

The manifolds of all the time series in the set $\mathcal{S}$ are reconstructed and stored in a set $\mathcal{M} = \{\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_N\}$. Each element of the set, $\mathbf{M}_i \in \mathbb{R}^{t_i \times 3}$, $i = 1, 2, \ldots, N$ represents a manifold which contains the 3D coordinates of the latent states corresponding to every moment of the time series. As discussed above, the manifolds in $\mathcal{M}$ do not necessarily share the same space. Therefore, one has to carefully process two manifolds to determine similarities between them. Again, the latent states in a manifold are one-to-one mapping of the samples of a time series. Thus, a manifold can be regarded as a "trajectory" of a signal in the phase space for a period of time. One can imagine that, if the time series coming from the same dynamic system are observed long enough, their corresponding trajectories in the phase space should be close to each other, even though the observation time and/or the observation duration are different. In other words, each manifold consists of a subset of all latent states of a dynamic system, and the state subsets corresponding to the homologous signals should have a larger intersection. This prompts us to seek a metric to quantify the similarity between manifolds by using probability distributions. To avoid the effect of manifold rotation, we introduce the centroid of the ith manifold, denoted by $\mathbf{O}_i \in \mathbb{R}^3$ and computed according to

$$\mathbf{O}_i = \frac{1}{t_i} \sum_{k=1}^{t_i} \mathbf{M}_i(k, :), \tag{1}$$

where the row vector $\mathbf{M}_i(k, :)$ indicates the kth latent state of the ith manifold. As illustrated in Fig. 2, for any manifold we can obtain a set of geometric vectors $\mathbf{V}_i = [\mathbf{v}_{i,1}, \mathbf{v}_{i,2}, \ldots, \mathbf{v}_{i,t_i}]^T \in \mathbb{R}^{t_i \times 3}$ from its centroid to its samples:

$$\mathbf{v}_{i,k}^T = \mathbf{M}_i(k, :) - \mathbf{O}_i, \tag{2}$$

where $k = 1, 2, \ldots, t_i$. Thus, the set of geometric vectors actually suggests the topology of the manifold. We denote the magnitude of the vectors by $\mathbf{v}_i = [|\mathbf{v}_{i,1}|, |\mathbf{v}_{i,2}|, \ldots, |\mathbf{v}_{i,t_i}|]^T$. Suppose that $\mathbf{A}_i$ and $\mathbf{B}_i$ are any two neighboring states of the manifold. The adjacent angle of vectors $\overrightarrow{O_i A_i}$ and $\overrightarrow{O_i B_i}$ is

$$\theta_{i,k} = \cos^{-1} \frac{\mathbf{v}_{i,k}^T \mathbf{v}_{i,k+1}}{|\mathbf{v}_{i,k}||\mathbf{v}_{i,k+1}|}, \tag{3}$$

where $k = 1, 2, \ldots, t_i - 1$. We collected the angles in the vector $\boldsymbol{\theta}_i = [\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,t_i-1}]^T$. Similar processing can be carried out for every manifold. Clearly, these two types of

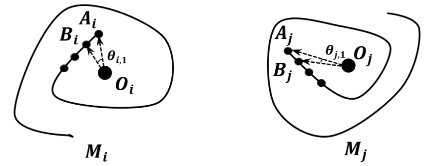

Fig. 2: Measurements of manifold topology.

measurements are no longer affected by space rotation (Fig. 2). Next, to compare the manifolds, the histograms of the elements of $\mathbf{v}_i$ and $\boldsymbol{\theta}_i$ are computed, respectively, to approximate the distributions of the magnitudes of vectors and their adjoining angles. Then, the manifold-to-manifold distance is defined as

$$d(\mathbf{M}_i, \mathbf{M}_j) = \frac{1}{2}\sqrt{J(p_{\mathbf{v}_i}(v), p_{\mathbf{v}_j}(v))} + \frac{1}{2}\sqrt{J(p_{\boldsymbol{\theta}_i}(\theta), p_{\boldsymbol{\theta}_j}(\theta))}, \tag{4}$$

where the function $J(\cdot, \cdot)$ computes the Jensen–Shannon divergence between two distributions [15]. The $p_{\mathbf{v}}(v)$ and $p_{\boldsymbol{\theta}}(\theta)$ represent the probability mass functions of $\mathbf{v}$ and $\boldsymbol{\theta}$, respectively, obtained from the histograms. In the above definition, the magnitudes of vectors are a proxy of the manifold shape, and the adjacent angles imply the rate of changing of the states.

B. Relative Positions in Euclidean Space

Once we have the pairwise distances between manifolds, the distance-based unsupervised anomaly detection technique, such as LoOP, can be applied directly. Given the distance matrix $\mathbf{D} \in \mathbb{R}^{N \times N}$, we can also visually build up a relative position relationship between the time series data in a Euclidean space, which is popular in many application problems. Assuming that the first time series corresponds to the origin of the coordinate, the Gram matrix $\mathbf{G}$ of a set of coordinate vectors in space can be computed element-by-element via

$$G_{i,j} = \frac{1}{2}(D_{1,j}^2 + D_{i,1}^2 - D_{i,j}^2) = \mathbf{x}_i^T \mathbf{x}_j, \tag{5}$$

where $D_{i,j}$ and $G_{i,j}$ are the (i, j)th elements of the matrices $\mathbf{D}$ and $\mathbf{G}$, respectively. The symbol $\mathbf{x}_i \in \mathbb{R}^3$ denotes the coordinates of the ith data point. Thus, we have $\mathbf{G} = \mathbf{X}\mathbf{X}^T$ where the matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^T \in \mathbb{R}^{N \times 3}$ is a collection of all the desired coordinates. Next, the matrix $\mathbf{G}$ can be factorized by singular value decomposition, or

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{U}^T, \tag{6}$$

where $\mathbf{S}$ is a diagonal matrix containing the eigenvalues of $\mathbf{G}$ in decreasing order. For visualization, the dimension in which the time series data can be embedded should be less or equal to three. Therefore, we trimmed the matrix $\mathbf{S}$ to $\tilde{\mathbf{S}} \in \mathbb{R}^{3 \times 3}$ that only holds the first three rows and the first three columns of the matrix $\mathbf{S}$. Then, the position matrix is calculated by

$$\mathbf{X} = \mathbf{U}\tilde{\mathbf{S}}^{\frac{1}{2}}. \tag{7}$$

The details of the method can be found in [16].

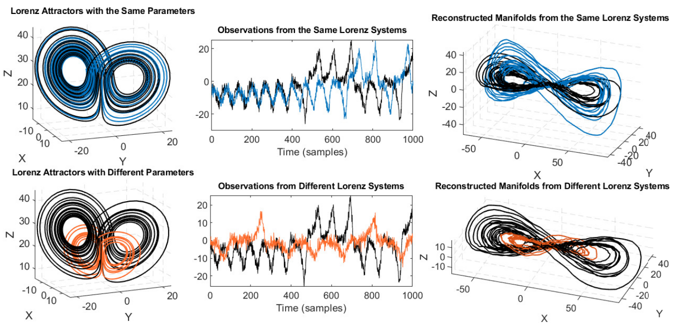

Fig. 3: Synthetic data and manifold reconstruction.

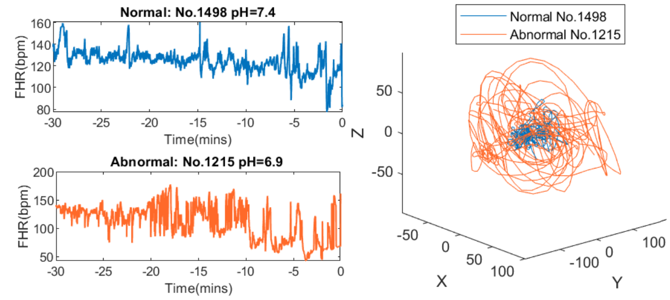

Fig. 4: Preprocessed FHR recordings and their manifolds.

V. Experiments

A. Synthetic Dataset from Lorenz System

In this section, we test our method by using a synthetic dataset generated from the well-known Lorenz systems. This system was developed for modeling atmospheric convection by Edward Lorenz and received lots of attention for having chaotic dynamics for certain parameters and initial conditions. The Lorenz differential equations are given by

$$\begin{aligned} dx_t/dt &= \lambda(y_t - x_t)\,, \\ dy_t/dt &= x_t(\rho - z_t) - y_t\,, \\ dz_t/dt &= x_t y_t - \beta z_t\,. \end{aligned} \tag{8}$$

Assuming that only y_t can be measured from the system, we generated 100 time series $\mathbf{y}$ of length L with a standard setting: $\lambda = 10$, $\beta = 8/3$ and $\rho = 28$. They are regarded as normal observations. The other five signals with the same time length are from the systems having parameters $[\beta, \rho] = [4, 50]$, $[5, 45]$, $[3, 65]$, $[7, 70]$ and $[1, 20]$, respectively. The parameter λ keeps the same value of 10. The five signals represent anomalies in this experiment. We picked the initial conditions of different observations uniformly from $[0, 1]$. To the signals we added i.i.d. white Gaussian noises with variance σ^2. Figure 3 compares the original Lorenz attractors, time series observations and reconstructed manifolds that are in the same dynamic systems and those that are not.

The performance of the proposed method is shown in Fig. 5, where we plotted the ROC curves of anomaly detection for $L = [100, 500, 1000, 2000]$ and $\sigma = [1, 2, 3, 4]$, respectively. The results match our intuition in that the accuracy improves as the observation time length increases, whereas the accuracy decreases with the increase of the noise variance. With a longer observation, the method has better capacity to capture more dynamic information of the system and explore more latent states in the phase space. As a result, a machine learning method can differentiate anomalous time series from normal ones more easily. The influence of the noise is mainly in the manifold reconstruction. We can see that, with long enough observations ($L = 2000$), the performance is maintained even with large noise ($\sigma = 4$). Figure 6 shows a result of position visualization when $L = 2000$ and $\sigma = 2$, where the scores next to the markers represent the probabilities of the corresponding time series of being anomalous.

B. Open Access Intrapartum CTG database

In this section, we explore the performance of the new method on discovering fetuses with abnormal behaviors by processing their heart rate tracings. The dataset used in this study was collected at the University Hospital in Brno, Czech Republic [17]. We selected 80 healthy recordings based on the agreement of more than half of the experts (see [2]). The five adverse recordings are selected also under this criterion and they have the lowest pH values (the pH distribution is shown on the left of Fig. 8). All the FHR tracings were preprocessed according to [18]. We used only the last 30-min recordings (see Fig. 4).

We divided the 30-min time series into six non-overlapping 5-min segments and computed the statistics of the local outlier probabilities of FHR from normal (denoted by 0) to abnormal (denoted by 1) fetuses, shown in Fig. 7. This result suggests that with time getting closer to birth, the outlier probabilities of healthy and non-healthy fetuses have statistical separation except for the last 5 mins (Fig. 7). In the last 5 mins, due to intense pressure from the contractions, even the healthy babies are no longer tolerant. However, this intolerance is even more pronounced by unhealthy fetuses. Yet, the difference between the two classes is less clear than in the first 25 minutes. Finally, we showed the relative positions of the FHR tracings in Fig. 8, which correspond to the time period 10-15 mins before delivery.

VI. Conclusion

In this paper, we proposed a new approach for detecting anomalous time series in an unsupervised manner. The variable-length time series undergo phase space reconstruction to identify their manifolds. A rotation-insensitive metric of distances between two manifolds was proposed. We used this metric to detect outliers and to visually display them. This work was tested by using both synthetic data and real-world data. The synthetic test results show good performance for various lengths and noise levels. When applied to FHR tracings, the results suggest that the method is promising in revealing fetal asphyxia during the second stage of labour.

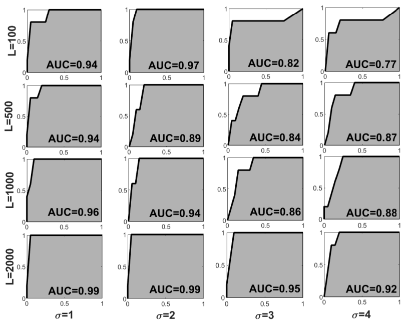

Fig. 5: ROC curves of synthetic testing.

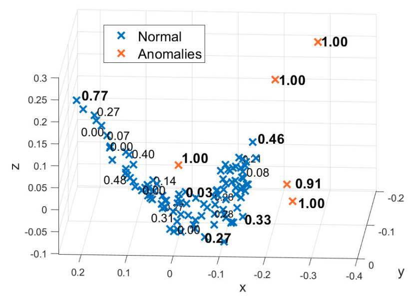

Fig. 6: A result of synthetic testing.

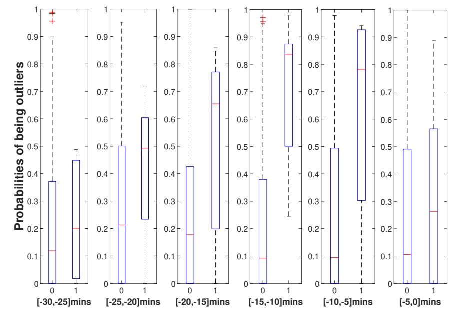

Fig. 7: Statistics of FHR outlier scores.

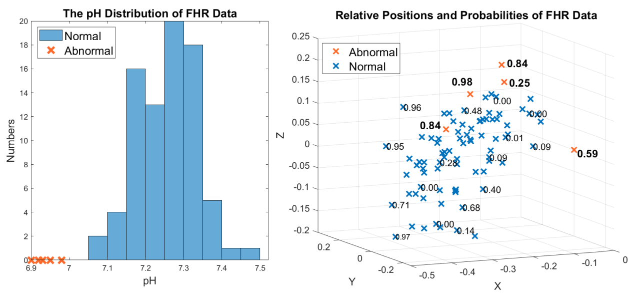

Fig. 8: FHR information and a testing result.

References

[1] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ArXiv*, vol. abs/2002.04236, 2020.

[2] L. Hruban, J. Spilka, V. Chudáček, P. Janků, M. Huptych, M. Burša, A. Hudec, M. Kacerovský, M. Koucký, M. Procházka, V. Korečko, J. Seget'a, O. Šimetka, A. Měchurová, and L. Lhotská, "Agreement on intrapartum cardiotocogram recordings between expert obstetricians," *Journal of Evaluation in Clinical Practice*, vol. 21, no. 4, pp. 694–702, 2015.

[3] G. Georgoulas, P. Karvelis, J. Spilka, V. Chudáček, C. D. Stylios, and L. Lhotska, "Investigating pH based evaluation of fetal heart rate (FHR) recordings," *Health and Technology*, vol. 7, no. 2, pp. 241–254, 2017.

[4] G. Feng, J. G. Quirk, and P. M. Djurić, "Supervised and unsupervised learning of fetal heart rate tracings with deep Gaussian processes," in *2018 14th Symposium on Neural Networks and Applications*, 2018, pp. 1–6.

[5] R. J. Hyndman, E. Wang, and N. Laptev, "Large-scale unusual time series detection," in *2015 IEEE International Conf. on Data Mining Workshop*, 2015, pp. 1616–1619.

[6] N. Laptev, S. Amizadeh, and I. Flint, "Generic and scalable framework for automated time-series anomaly detection," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2015, p. 1939–1947.

[7] S. Benkabou, K. Benabdeslem, and B. Canitia, "Unsupervised outlier detection for time series by entropy and dynamic time warping," *Knowledge and Information Systems*, vol. 54, 2018.

[8] L. Beggel, B. X. Kausler, M. Schiegg, M. Pfeiffer, and B. Bischl, "Time series anomaly detection based on shapelet learning," *Computational Statistics*, vol. 34, no. 3, pp. 945–976, 2019.

[9] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick 1980*, D. Rand and L.-S. Young, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 366–381.

[10] M. B. Kennel, R. Brown, and H. D. I. Abarbanel, "Determining embedding dimension for phase-space reconstruction using a geometrical construction," *Phys. Rev. A*, vol. 45, pp. 3403–3411, Mar 1992.

[11] G. Feng, K. Yu, Y. Wang, Y. Yuan, and P. M. Djurić, "Improving convergent cross mapping for causal discovery with Gaussian processes," in *IEEE Int. Conf. on Acoust., Speech and Signal Process.*, 2020, pp. 3692–3696.

[12] H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "LoOP: Local outlier probabilities," in *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, ser. CIKM '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 1649–1652.

[13] M. OpenCourseWare, "Hst. 582j/6.555 j/16.456 j biomedical signal and image processing," 2007.

[14] R. Wang, W. Gao, X. Chen, and S. Shan, "Manifold-manifold distance with application to face recognition based on image set," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, Jun 2008, pp. 1–8.

[15] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Transactions on Information Theory*, vol. 49, no. 7, pp. 1858–1860, 2003.

[16] G. M. Crippen and T. F. Havel, "Stable calculation of coordinates from distance information," *Acta Crystallographica Section A*, vol. 34, no. 2, pp. 282–284, 1978.

[17] V. Chudáček, J. Spilka, M. Bursa, P. Janku, L. Hruban, M. Huptych, and L. Lhotska, "Open access intrapartum ctg database," *BMC pregnancy and childbirth*, vol. 14, p. 16, Jan. 2014.

[18] G. Feng, J. G. Quirk, and P. M. Djurić, "Recovery of missing samples in fetal heart rate recordings with Gaussian processes," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 261–265.