The 10th IEEE International Conference on E-Health and Bioengineering - EHB 2022
Grigore T. Popa University of Medicine and Pharmacy, Iasi, Romania, November 17-18, 2022

Compression of Different Time Series Representations in Asphyxia Detection

Bárbara Silva[1], Maria Ribeiro[2], Teresa S. Henriques[3]
Affiliation 1: Faculty of Engineering, University of Porto, FEUP, Porto, Portugal, up201906415@edu.fe.up.pt
Affiliation 2: Institute for Systems and Computer Engineering, Technology and Science (INESC-TEC), Porto, Portugal and also with Computer Science Department, Faculty of Sciences, University of Porto, Porto, Portugal, maria.r.ribeiro@inesctec.pt
Affiliation 3: Centre for Health Technology and Services Research, CINTESIS, Faculty of Medicine, University of Porto, Porto, Portugal and Department of Community Medicine, Information and Health Decision Sciences, MEDCIDS, Faculty of Medicine, University of Porto, Porto, Portugal teresasarhen@med.up.pt

***Abstract*— Physiological signals offer a vast amount of information about the well-being of the human system. Understanding the behavior and complexity of these signs is important for accurate assessments and diagnoses. This study focuses on fetal heart rate (FHR) analysis and its potential to detect perinatal asphyxia by analyzing how different representations of the FHR series could aid in asphyxia detection. Additionally, different compression schemes were applied to evaluate the potential of compression as a measure of complexity. For this purpose, text files containing data of the last hour of the FHR before birth were converted into different types of images (Time Series, Time Series with fixed axes, Recurrence Plot and Poincaré Plot). We then applied compression schemes for text (BZIP2 and GZIP) and images (Lempel-Ziv-Welch, DEFLATE, and JPG) in 5, 10, and 30-minute windows. Correlation analysis revealed that similar compressed formats, such as BZIP2/GZIP and TIFF LZW/TIFF DEFLATE/JPG LOSSY/JPG LOSSLESS, showed the highest values and the correlation between uncompressed and compressed formats became increasingly more negative for larger time windows. Mann-Whitney test between groups (with and without asphyxia) revealed that compressed patterned images, such as Recurrence Plots, showed the highest potential in detecting asphyxia. Moreover, we confirm that larger time windows allow for better detection, due to the presence of more detailed patterns. These findings confirmed the potential of time series image representation in detecting fetal conditions, as well as show that the compression of images leads to better results than the compression of text files.**

Keywords*— *data compression; time series; fetal heart rate; perinatal asphyxia; recurrence plot

I. INTRODUCTION

Physiological signals, such as blood pressure, heart rate, and respiratory rate offer important information on the overall well-being of the human body. These signals tend to fluctuate with uncertain patterns, that may be affected by external and internal factors, resulting in complex functions.

Fetal monitoring is possible through the noninvasive recording of signals such as the fetal heart rate (FHR), uterine contractile activity, and fetal movements. The FHR is associated with the highest diagnosis significance for the neonatal outcome. However, for a correct interpretation of this signal, its high complexity must be considered [1].

The pathology evaluated in this study is perinatal asphyxia, characterized by low oxygen intake by the baby during birth. According to recent studies, the incidence of this problem is two in 1000 in developed countries but up to 10 times higher in developing countries [2]. The assessment of FHR may help to identify patterns characteristic of this condition.

Entropy has been detailed as a measure of complexity, most associated with a state of disorder or randomness. Another approach to evaluating a system's complexity is the use of compression, which approximates the Kolmogorov complexity [3]. Compressors have been increasingly explored for their ability to assess complexity and its consequent application in different scientific areas. The advantage of compressors comes from the possibility of using them in different representations.

Data compression (DC) has numerous approaches, with studies on the subject being carried out in distinct research areas (from computer science to the medical field). The basis of these techniques is to find parts of data that can be eliminated without compromising its integrity. Therefore, a compression algorithm will look for repetitions that indicate the presence of redundancy and irrelevancy [4].

Different kinds of data characterization generate contrasting DC approaches. A survey on recent data compression techniques [5] classified them based on four main distinct aspects: data quality, coding schemes, data type, and applications. This work focuses mainly on the distinction of DC techniques based on quality (lossy and lossless) and their applications for different data types (such as text and image).

Lossless compression reduces bits without losing information. It usually exploits statistical redundancy to

978-1-6654-8557-9/22/$31.00 ©2022 IEEE

represent data without losing any information, allowing it to reverse the problem. On the other hand, lossy compression reduces bits by removing unnecessary or less important information. There is a corresponding trade-off between preserving information and reducing size. Lossy data compression schemes are designed by research on how people perceive the data in question.

Knowing the potential of the FHR analysis in the diagnosis and the compressors' significant assessment of signal complexity, this study aimed to evaluate how different compression of multiple representations of the FHR time series behaved in the detection of the asphyxia state.

II. METHODS

A. Database and Methodology

The data used for this study is part of the CTU-CHB Intrapartum Cardiotocography Database, provided by Physionet [12,13]. From the 9164 recordings collected, 552 were carefully selected following specific criteria.

Of the 552 files, 246 were chosen for this study with less than 30% of signal loss. These 246 fetuses were divided into 2 groups based on their pH value. Babies with a pH value lower than 7.15 were considered to have asphyxia. The database also contains additional parameters such as maternal, delivery, fetal and fetal outcome data.

The original text file was in an uncompressed format (comma-separated values - CSV). Each CSV file was then compressed into a BZIP2 file and a GZIP file. The BZIP2 format takes advantage of the Burrows-Wheeler Transform algorithm while the GZIP compression scheme is based on the DEFLATE algorithm, a combination of Lempel-Ziv (specifically LZ77) and Huffman coding. All three formats were applied in time windows of 5, 10, and 30 minutes, resulting in 12, 6, or 2 files from each original file.

B. Time Series Representations

Additionally, four types of images were considered. The first one corresponds to the simple Time Series (TS) with an adjustable axis according to each data scale. The second one, referred to as Time Series Fixed (TSF), is the same representation but with a previously fixed x and y-axis. The third one is a Recurrence Plot (RP) and the fourth is a Poincaré Plot (PC), used as very different but still significant representations of time series data.

The RP is an advanced method of nonlinear data analysis, used as a visualization tool to examine the m-dimensional phase space trajectory. The matrix identifies the times at which a dynamical system's state recurs, i.e., where some trajectories change back to a former condition [7]. By simple visual analysis, the resulting image can provide information about the time series tendencies. When building the RP, three parameters can be adjusted: embedding dimension, delay time, and distance threshold [8].

To create the RP, the code was adapted from [14] containing an *embed* function and a *recurrence plot* function to create the plot itself. The parameters used for the RP were m=2 (embedding dimension), τ=1 (time delay), and ε= 20% of the standard deviation (threshold).

The Poincaré plot (PC), also referred to as a delay map, is a specific type of recurrence plot used to represent and quantify the correlation between two successive time-series data points. Therefore, the x-axis will represent the X_n points versus the X_{n+1} points in the y-axis. This technique has been widely used in the analysis of physiological signals for their specific fluctuation dynamic [9,10,11].

For the Poincaré Plot, the methods were adapted from [10], which provides the *MsPplots* function for the generation of Multiscale Poincaré plots. In this study, a simpler approach was taken by defining the scale parameter as 1.

Each of the image types mentioned was also processed in 5 different formats. The TIFF format was considered an uncompressed reference. However, it is also possible to save a file as TIFF with a compression scheme, such as LZW (Lempel-Ziv-Welch) or DEFLATE. Besides that, JPG was also used, with both lossy and lossless compression.

C. Statistical Analysis

Two results were considered for the analysis: the compression result, referring to the size of the compressed file, and the compression ratio, which corresponds to the size of the compressed file divided by the size of the uncompressed file. These results were then evaluated with two approaches.

Firstly, Spearman's correlation was calculated to determine the similarity between different formats and representations. Secondly, to determine if the different representations are reliable in the detection of asphyxia, the Mann-Whitney test was performed between variables of patients with and without asphyxia. These analyses were made for last time windows when there's a higher probability of detecting asphyxia, as it is closer to birth.

The dataset was divided into two groups, one of those with asphyxia (n=35) and the other without (n=211). Mann-Whitney test was used to compare values from babies with and without asphyxia. FHR segments of 5, 10, and 30 minutes were considered. For text, there were 5 results to evaluate (three formats - CSV, BZIP2, and GZIP - and two ratios) in three different time windows. For images, 20 image types (4 image types in 5 formats) and 16 compression ratios were analyzed in the three time windows.

III. RESULTS

A. Data Characterisation

Table I presents relevant clinical values regarding the patients with and without asphyxia. The results include the median of variables such as pH (total median = 7.26), Apgar index (total median = 9), gestation weeks (total median = 40), weight (total median = 3370), maternal age (total median = 29) and the percentage distribution of the baby's sex (in the full group, 121 (49%) are male). The Apgar and pH values are slightly higher in the group without asphyxia.

TABLE I. DATA CHARACTERISTICS

Variable	Asphyxia (n=35) Median [Q1, Q3]	No Asphyxia (n=211) Median [Q1, Q3]
pH	7.12 [7.08,7.13]	7.27 [7.23, 7.32]
Apgar	8 [8, 9]	9 [9, 10]
Gestation Weeks	41 [40,41]	40 [39, 41]
Weight	3390 [3250, 3650]	3370 [3050, 3630]
Maternal Age	29 [26, 31]	29 [27, 33]
Sex (male)	18 (51.4%)	103 (48.8%)

The data itself was evaluated in its simple text form and four types of images (TS, TSF, RP, and PC). An example of the above-mentioned image representations is presented in Fig. 1, regarding the last 10 minutes before birth.

B. Correlation Results

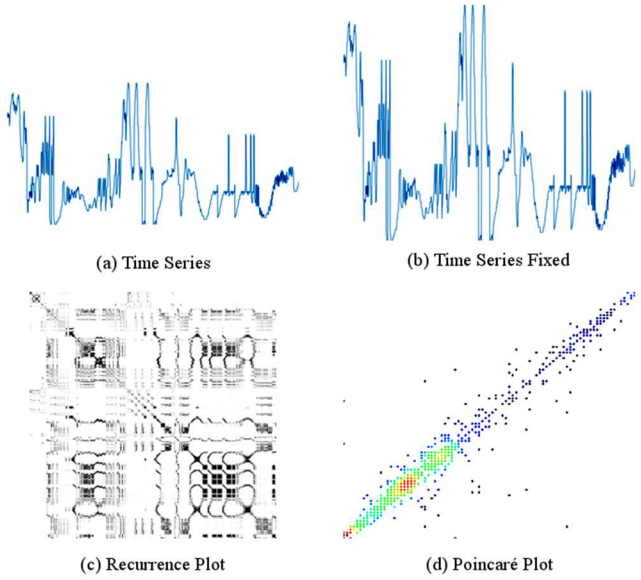

Fig. 1. Example of different image' representations for the last 10 min.

For the text results, a high correlation between BZIP2/GZIP was found throughout all time intervals, while the uncompressed format (CSV) shows a low correlation with both compressed formats, with a tendency for a more negative correlation as the time interval increases. In the last 5 minutes, the correlation ranges from -0.05 between CSV/BZIP2 to 0.95 between BZIP2/GZIP. For the same pairs, in the last 10 minutes, the correlation ranges from -0.20 to 0.94, and for the last 30 minutes, from -0.37 to 0.95.

For images, four correlation matrices were created for each time interval, one for each image type (Tables II and III).

TABLE II. CORRELATION MATRIX BETWEEN FIVE COMPRESSORS (TIFF, LZW - LEMPEL-ZIV-WELCH, DEF - DEFLATE, JPG LOSSY, AND JPG LOSSLESS) FOR TIME SERIES (GREY) AND TIME SERIES FIXED (BLUE)

Compressors		TIFF	LZW	DEF	JPG LOSSY	JPG LL
TIFF	**5 min**		0.08	0.11	0.09	0.09
	10 min		0.11	0.10	0.11	0.10
	30 min		0.05	0.05	0.11	0.06
LZW	**5 min**	0.3		0.93	0.97	1.00
	10 min	-		0.89	0.97	1.00
	30 min	0.00		0.96	0.97	1.00
DEF	**5 min**	0.27	0.99		0.95	0.92
	10 min	-	0.99		0.92	0.89
	30 min	0.00	0.98		95	0.95
JPG LOSSY	**5 min**	0.31	0.99	0.99		
	10 min	-	0.99	0.99		
	30 min	0.11	0.98	0.98		
JPG LL	**5 min**	0.31	1.00	0.98	0.99	
	10 min	-	1.00	0.98	0.99	
	30 min	0.00	1.00	0.97	0.99	

LL - lossless

TABLE III. CORRELATION MATRIX BETWEEN FIVE COMPRESSORS (TIFF, LZW - LEMPEL-ZIV-WELCH, DEF - DEFLATE, JPG LOSSY, AND JPG LOSSLESS) FOR POINCARÉ PLOT (GREY) AND RECURRENCE PLOT (BLUE)

Compressors		LZW	DEF	JPG LOSSY	JPG LOSSLESS (LL)
TIFF	**5 min**	-0.24	-0.24	-0.18	-0.24
	10 min	-0.27	-0.30	-0.04	-0.05
	30 min	-0.37	-0.38	-0.35	-0.37
LZW	**5 min**		0.99	0.86	0.86
	10 min		0.99	0.07	0.09
	30 min		0.99	0.83	0.91
DEF	**5 min**	0.85		0.87	0.83
	10 min	0.85		0.07	0.09
	30 min	0.97		0.81	0.88
JPG LOSSY	**5 min**	0.92	0.81		0.94
	10 min	0.95	0.81		0.95
	30 min	0.99	0.98		0.89
JPG LL	**5 min**	0.93	0.73	0.90	
	10 min	0.93	0.73	0.91	
	30 min	0.98	0.96	0.98	

In the last 5 minutes, for the TS images, the correlation ranges from 0.075 between TIFF/LZW to 0.995 between LZW/JPGLOSSLESS. For the TSF images, the correlation ranges from 0.307 between TIFF/LZW to 0.997 between LZW/JPGLOSSLESS. For the RP images, the correlation between compressed formats ranges from 0.728 between DEF/JPGLOSSLESS to 0.928 between LZW/JPGLOSSLESS. Finally, for the PC images, the results show negative values for

the TIFF format, ranging from -0.244 between TIFF/LZW to 0.987 between LZW/DEF.

In the last 10 minutes, similar to what was seen in the 5 minutes interval, the compressed formats show high values of correlation for TS, TSF, and RP. However, for this time interval, the PC image only shows a high correlation between LZW/DEF (0.990) and JPGLOSSY/JPGLOSSLESS (0.949) with a minimum of 0.066 for DEF/JPGLOSSLESS. For the TS images, the correlation ranges from 0.099 between TIFF/DEF to 0.995 between LZW/JPGLOSSLESS. For the TSF images, the correlation between compressed formats showed high values between 0.983 for DEF/JPGLOSSLESS to 0.997 for LZW/JPGLOSSLESS. The same can be seen for RP images, ranging from 0.713 for DEF/JPGLOSSLESS with a top value of 0.928 for LZW/JPGLOSSLESS.

In the last 30 minutes, for the TS images, the correlation ranges from 0.053 between TIFF/LZW to 0.995 between LZW/JPGLOSSLESS. For the TSF images, the results range from -0.008 between TIFF/LZW to 0.997 between LZW/JPGLOSSLESS. The matrix for RP images is consistent with the ones seen for the last 5 minutes and the last 10 minutes. The values are also consistently high, with the top value of 0.990 for LZW/JPGLOSSY. Finally, for the PC images, the results show negative values for the TIFF format, ranging from -0.384 between TIFF/DEF to 0.999 between LZW/DEF.

Regarding the correlation between text/image, in the last 10 minutes, the PC LZW image showed high values of correlation with BZIP2 and GZIP while the uncompressed format PC TIFF showed moderate correlation values with CSV.

C. *Asphyxia Detection*

For text, a significant difference was verified for CSV in the last 5, 10, and 30 minutes (p-values: 0.027, 0.019, and 0.003 respectively) and GZIP only in the last 5 minutes (p-value: 0.020).

For images, significant p-values were found in the last 10 minutes for PC DEF (0.023), PC JPG LOSSY (0.032), and PC JPG LOSSLESS (0.031). In the last 30 minutes, many significant differences between the two groups were found: for TSF LZW (0.033), TSF DEF (0.025), TSF JPG LOSSY (0.041), TSF JPG LOSSLESS (0.046), PC LZW (0.005), PC DEF (0.004), and all the RP compressed formats. For the TSF and PC images, the median compression result (size of compressed file) was lower in the group without asphyxia while in the RP images, the opposite occurred.

Finally, for image ratios, the statistically significant p-values were found in the last 10 minutes for PC DEF (0.023), PC JPG LOSSY (0.032), and PC JPG LOSSLESS (0.034) ratios. In the last 30 minutes, significant p-values were found: for TSF LZW (0.038), TSF DEF (0.029), TSF JPG LOSSY (0.05) ratios, PC LZW (0.005), PC DEF (0.004), and for the RP ratios. Table IV presents a schematic representation of the significant differences found between the groups.

TABLE IV. MANN-WHITNEY TESTE RESULTS

	Text	**Image**
Last 5 minutes	CSV*, GZIP*	-
Last 10 minutes	CSV*	PC* (DEFLATE, JPG LOSSY & JPG LOSSLESS) Ratios: PC* (DEFLATE, JPG LOSSY & JPG LOSSLESS)
Last 30 minutes	CSV**	TSF* (LZW, DEFLATE, JPG LOSSY & JPG LOSSLESS) PC** (LZW & DEFLATE) RP*** (LZW, DEFLATE, JPG LOSSY & JPG LOSSLESS) Ratios: TSF* (LZW, DEFLATE & JPG LOSSY) PC** (LZW & DEFLATE) RP***

Legend: * <0.05; ** <0.01 and *** < 0.001

IV. DISCUSSION

Starting with the text correlation, compressed formats (BZIP2 and GZIP) have high correlation values between themselves, reflecting the similar behavior of the compression techniques. Between compressed and uncompressed formats (CSV), the correlation is either null or negative, which could indicate that the bigger the original file size higher the compression will be.

For images, the results were mostly consistent. In the last 5 minutes, the pattern continues, with a high correlation between compressed formats and low values between uncompressed and compressed. An unusual result was obtained for the RP image once the RP TIFF results were nearly constant, which led the standard deviation value to 0, reflecting a non-valid value in correlation. In the last 10 minutes, the same phenomenon is present in both TSF and RP images. For the TS images, the result was as expected but in the PC images, high values of correlation were only observed in similar compression algorithms (LZW/DEF and JPGLOSSLESS/JPGLOSSY). Finally, for the last 30 minutes, the results are similar to those described for the last 5 minutes, following the same discussion.

In conclusion, the correlation results revealed that compressed formats show definitive higher values of correlation with other compressed formats for every data type. Besides that, the correlation between uncompressed and compressed formats tends to get more negative with larger time windows. In general, the 30 minutes window resulted in higher values of correlation between different representations.

Regarding asphyxia detection, in text formats, compression doesn't seem to have a clear impact on the difference between the two groups.

For images, the significant p-values were greater in larger time windows. The best results were found for the RP and PC

images, due to the extremely representative patterns these images generate. Accordingly, greater time windows allow for more patterns in the image, supporting the hypothesis that these are a good indicator of the discrepancy between the groups. One advantage of these methods is that they can be easily implemented in real-time analysis.

Furthermore, the relevant values were only found in compressed image formats, which indicates a clear improvement in the results with the aid of compression. Regarding the compression scheme, although all five formats achieved good results either for size or ratio, lossless algorithms seem to have more consistent results with the different image types (specifically Lempel-Ziv-Welch and DEFLATE).

V. Conclusions

The presented study aimed to explore how different representations of the FHR time series may reflect on the detection of fetal conditions, specifically, perinatal asphyxia. Besides that, an analysis of different compression algorithms was made for every representation to evaluate them as a measure of the complexity of physiological signals. The original 60 minutes FHR text files were converted to images and compressed in both formats.

A correlation analysis revealed high correlation values between compressed formats, which were higher for compression schemes with similar approaches. Between compressed and uncompressed formats, the correlation was low and increasingly negative with larger time windows.

For the detection of asphyxia, the results revealed that, for text, statistical differences were found in uncompressed formats. For images, the pattern representations (RP and PC) in their compressed formats showed the best results, for their characteristic way of representing the information of a time series. Besides that, larger time windows resulted in lower p-values, pointing to much more reliable detection of asphyxia.

There is a variety of work on time series image representation proving that this format provides information about the evolution and behavior of the series that a simple list of values does not. This is true, particularly in image representations that generate representative patterns such as recurrence plots. Therefore, future work should be focused on these representations.

Acknowledgment

This work was supported by National Funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., within CINTESIS, R&D Unit (reference UIDP/4255/2020). Maria Ribeiro was supported by FCT under the scholarship SFRH/BD/138302/2018.

References

[1] R. Czabánski, J. Je˙zewski, K. Horoba, and M. Je˙zewski, “Fetal state assessment using fuzzy analysis of fetal heart rate signals—agreement with the neonatal outcome,” Biocybernetics and Biomedical Engineering, vol. 33, no. 3, pp. 145–155, 2013.

[2] M. Gillam-Krakauer and C. W. Gowen, Birth Asphyxia. In StatPearls. StatPearls Publishing, 2021.

[3] T. Henriques, H. Gonçalves, L. Antunes, M. Matias, J. Bernardes, and C. Costa-Santos, “Entropy and compression: two measures of complexity,” Journal of Evaluation in Clinical Practice, vol. 19, no. 6, pp. 1101–1106, 2013.

[4] L. A. Fitriya, T. W. Purboyo, and A. L. Prasasti, “A review of data compression techniques,” International Journal of Applied Engineering Research, vol. 12, pp. 8956–8963, 2017.

[5] U. Jayasankar, V. Thirumal, and D. Ponnurangam, “A survey on data compression techniques: From the perspective of data quality, coding schemes, data type and applications,” Journal of King Saud University-Computer and Information Sciences, vol. 33, no. 2, pp. 119–140, 2021.

[6] L. Karamitopoulos and G. Evangelidis, “Current trends in time series representation,” in Proc. 11th Panhellenic Conference on Informatics, 2007, pp. 217–226.

[7] N. Hatami, Y. Gavet, and J. Debayle, “Bag of recurrence patterns representation for time-series classification,” Pattern Analysis and Applications, vol. 22, no. 3, pp. 877–887, 2019

[8] Z. Zhao, Y. Zhang, Z. Comert, and Y. Deng, “Computer-aided diagnosis system of fetal hypoxia incorporating recurrence plot with convolutional neural network,” Frontiers in physiology, vol. 10, p. 255, 2019.

[9] A. K. Golinska, “Poincaré plots in analysis of selected biomedical signals,” Studies in logic, grammar and rhetoric, vol. 35, no. 1, pp. 117–127, 2013.

[10] T. S. Henriques, S. Mariani, A. Burykin, F. Rodrigues, T. F. Silva, and A. L. Goldberger, “Multiscale Poincaré plots for visualizing the structure of heartbeat time series,” BMC medical informatics and decision making, vol. 16, no. 1, pp. 1–7, 2015.

[11] R. Satti, N.-U.-H. Abid, M. Bottaro, M. De Rui, M. Garrido, M. R. Raoufy, S. Montagnese, and A. R. Mani, “The application of the extended Poincaré plot in the analysis of physiological variabilities,” Frontiers in physiology, vol. 10, pp. 116, 2019.

[12] V. Chud´a˘cek, J. Spilka, M. Bur˘sa, P. Jank, L. Hruban, M. Huptych, and L. Lhotsk´a, “Open access intrapartum CTG database,” BMC pregnancy and childbirth, vol. 14, no. 1, pp. 1–12, 2014.

[13] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals,” Circulation, vol. 101, no. 23, pp. E215–20, Jun. 2000.

[14] N. Marwan, M. Romano, and M. Thiel, “Recurrence plots and cross recurrence plots.” [Online]. Available: http://www.recurrence-plot.tk