

Discovering Hidden Relationships in Physiological Signals For Prediction of Periventricular Leukomalacia

Ali Jalali¹, Daniel J. Licht² and C. Nataraj³

Abstract—This paper is concerned with predicting the occurrence of Periventricular Leukomalacia (PVL) using vital data which are collected over a period of twelve hours after neonatal cardiac surgery. The vital data contain heart rate (HR), mean arterial pressure (MAP), right atrium pressure (RAP), and oxygen saturation (SpO₂). Various features are extracted from the data and are then ranked so that an optimal subset of features that have the highest discriminative capabilities can be selected. A decision tree (DT) is then developed for the vital data in order to identify the most important vital measurements. The DT result shows that high amplitude 20 minutes variations and low sample entropy in the data is an important factor for prediction of PVL. Low sample entropy represents lack of variability in hemodynamic measurement, and constant blood pressure with small fluctuations is an important indicator of PVL occurrence. Finally, using the different time frames of the collected data, we show that the first six hours of data contain sufficient information for PVL occurrence prediction.

I. INTRODUCTION

Newborns with congenital heart disease are at high risk of brain injury and adverse neurodevelopmental outcomes [1], [2]. A study by Miller et al. [3] showed that full-term newborns with congenital heart defect (CHD) have widespread brain abnormalities before they undergo cardiac surgery. The imaging findings of their study are similar to those in premature newborns and may reflect abnormal brain development in utero. Licht et al. [4] showed that before surgery, term infants with hypoplastic left heart syndrome and transposition of the great arteries have brains that are smaller and structurally less mature than expected.

Periventricular leukomalacia (PVL) is a type of brain injury that affects infants. The condition involves the

death of small areas of the brain tissue around fluid-filled areas called ventricles. Research has shown a high incidence of PVL both before and after cardiac surgery in neonates [5], [6]. Recently, there has been a growing interest in clinical research to aim to understand the progression and pathology of PVL, to develop protocols for the prevention of PVL development and to examine the trends in outcomes of individuals with PVL [7]. The study of relationship between preoperative cerebral blood flow and preoperative neurologic conditions was carried out by Licht et al. [8].

Despite advancement in research in the field of PVL, there are no treatments currently prescribed for PVL; furthermore, clinical investigation of these patients almost always have low accuracy [9]. This is due to fact that the origin of PVL and its physiology still remain to be clearly understood.

A computer based decision making tool, also referred to as an “Intelligent Patient Monitoring Tool” or simply *The IPM tool*, will help the care-givers aggregate different types of physiological data and discover the hidden knowledge or patterns in the data to quickly make the correct decision.

In our previous paper [10], we have shown how decision tree (DT) will help us to highlight the role of blood dioxide carbon contents (HCO₃ and PaCO₂). In this study, we investigate how DT can help us to predict the occurrence of the PVL using the vital data. The aim is to identify the most important risk factors based on the classification rules to be extracted. These rules will enable better management of the patient targeting the reduction of events, as well as, reduction of the cost of therapy, due to the expected restriction of interventions to necessary cases only.

II. MATERIALS AND METHODS

A. Data Collection

Data from 44 neonates for a period of 12 hours after neonatal cardiac surgery were collected according to a pre-specified protocol at the Children’s Hospital of Philadelphia (CHOP). Subjects of this study are limited to two cases of congenital heart disease, hypoplastic

¹A. Jalali is with Department of Mechanical Engineering, Villanova University, 800 E. Lancaster Ave., Villanova, PA, USA email: ali.jalali@villanova.edu

²D.J. Licht is Director of Neurovascular Imaging Lab, Children’s Hospital of Philadelphia, Colket Bldg Philadelphia, PA, USA email: licht@email.chop.edu

³C. Nataraj is Mrs. and Mr. Mortiz, Sr. Endowed Professor in Engineered Systems and Chair of the Department of Mechanical Engineering, Villanova University, 800 E. Lancaster Ave., Villanova, PA, USA nataraj@villanova.edu

left heart syndrome (HLHS) and transposition of great arteries (TGA), accounting for the fact that these two diseases are considered to have the highest likelihood of PVL occurrence as their postoperative effect. Clinical and demographic characteristics of the study cohort are shown in Tab. I.

TABLE I

DEMOGRAPHIC CHARACTERISTICS OF THE COLLECTED DATA. DHCA IS DEEP HYPOTHERMIC CARDIAC ARREST TIME, CPB IS CARDIOPULMONARY BYPASS TIME, CCC IS CROSS CLAMPED DURATION TIME.

Male, %	59
Diagnosis, %HLHS	55
DHCA time, mean \pm SD	27 \pm 26
CPB time, mean \pm SD	102 \pm 31
CCD time, mean \pm SD	61 \pm 19
PVL, %	45
Extent, mean \pm SD	94 \pm 251

For each patient, we collected vital data as well as blood gas measurements. The sampling time for the vital data varies both inter and intra-patient between 4-17 seconds.

B. Developed Algorithm

Next, we discuss the steps involved with the designed algorithm for the task of data classification and role extraction. The patient data is collected at the hospital which will be used to form the pool of features. A mutual information based algorithm will then be applied to rank the features in the feature pool. After forming the ordered feature set the optimal feature subset that encapsulates the most critical features will be selected. Compared to the original feature set, the optimal feature subset is reduced in size; however, by maximizing the class separability measure, this subset will result in higher accuracy in the final prediction. The selected features will then feed to the decision tree (DT). In Fig. (1) we provide a schematic overview of the proposed algorithm.

C. Feature Extraction

Next, the feature pool will be developed based on the collected set of physiological measurements. The feature pool contains features generated from the vital data and represents the characteristics of this data. Currently, the derived features of vital measurements include: min, max, mean, variance, skewness, kurtosis, trend, energy of wavelet coefficients in different time frames and multi scale entropy (based on sample entropy measure). Minimum and Maximum of the data are important

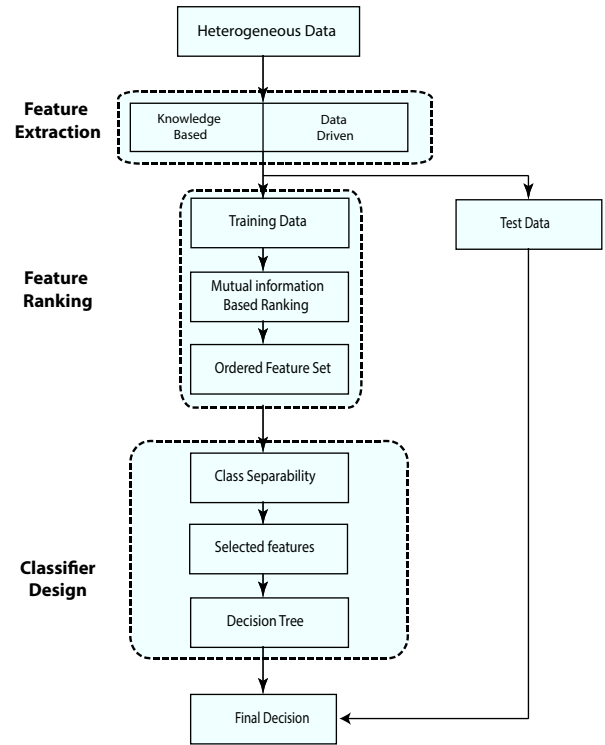


Fig. 1. Schematic of the proposed algorithm.

because they could potentially be representative of some mechanisms triggered when a measurement passes critical values. Skewness and kurtosis are third and fourth order statistical moment of a random variable defined by (1).

$$m_n(x) = E\{(x - \mu)^n\} \quad (1)$$

where, n is the order, μ is the mean value of the data and E is the expected value.

For this study we focus on the variation in the data in 1 min, 20 min and 2 hour periods and we would like to find the most important time scales in the different waveforms. The energy of the continuous wavelet transform (CWT) coefficients of the vital data at these time scales is a measure of variation in the different time frames.

Sample Entropy (SampEn) is measure of signal complexity and is the negative natural logarithm of the conditional probability of having signal window with length N , having repeated itself within a tolerance r for m points, will also repeat itself for $m + 1$ points, without allowing self-matches [11]. SampEn has been used in the literature to evaluate the cyclic behavior of heart rate variability (HRV) and blood pressure variability (BPV) [12], [13]

D. Feature Ranking

In this paper we apply the concept of mutual information to rank the features. Mutual information of random variables is a quantity that measures the mutual dependence of those random variables. Let x_i be the i th feature and $p(x_i)$ be its corresponding probability density function. The mutual information is then defined as:

$$I(x_i; c_k) = \sum_{i \in X} \sum_{k \in \{\pm 1\}} p(x_i, c_k) \log \frac{p(x_i, c_k)}{p(x_i)p(c_k)} \quad (2)$$

where, c_k represents the classes and $p(x_i, c_k)$ is the joint probability distribution of x_i and c_k . We will approach the feature selection process using mutual information as an optimization problem which seeks to find a subset S_{opt} from the whole feature set S by maximizing the information content $I(S; c_k)$ between the feature set and the output. In this proposal we use the algorithm proposed by Kappaganthu and Nataraj [14]. They maximize the mutual information of subset S_i using (4):

$$I(x_i; c_k) = \frac{1}{|S|} \sum_{x_i \in S} I(x_i; c_k) - \frac{1}{|S-1|^2} \sum_{x_i, x_j \in S} I(x_i; x_j) \quad (3)$$

$$x = \{x : x \in S_i \subset S\} \quad (4)$$

E. Class Separability Measure

Now that the feature vector has been formed, the next step is to measure the discriminative capacity of the feature vectors. The class separability measure is defined as a divergence between classes using the feature vector x ,

$$d = \int_{-\infty}^{\infty} (p(x|w_1) - p(x|w_2)) \ln \frac{p(x|w_1)}{p(x|w_2)} dx \quad (5)$$

where, $p(x|w_i)$ is the conditional probability of x with respect to w_i . An optimal feature subset is the feature subset that maximizes the class separability.

III. RESULTS

Now, the primary feature pool has been formed; the features have been ranked using the mutual information technique and the optimal feature set has been selected using the class separability measure. Table (II) shows 12 features that form the optimal feature set. The table shows that the energy of wavelet coefficients, sample entropy and kurtosis are the most important features for PVL occurrence prediction.

In this step we develop our classifier in order to classify PVL patients for healthy subjects. The DT constructed based on the optimal feature set to predict the occurrence of the PVL is represented in Fig. (3). Investigating this DT shows that high amplitude 20

TABLE II
OPTIMAL SUBSET OF THE FEATURES WHICH WILL BE USED FOR
DESIGNING THE CLASSIFIER.

Rank	Feature
1	2 hour variations in HR
2	MAP sample entropy
3	20 min variations in HR
4	Kurtosis HR
5	1 min variations in SpO ₂
6	20 min variations in SpO ₂
7	Kurtosis MAP
8	1 min variations in MAP
9	Min SpO ₂
10	HR sample entropy
11	2 hour variations in MAP
12	RAP sample entropy

minute variations and low sample entropy in the data is an important factor for prediction of PVL. Low sample entropy represents lack of variability in hemodynamic measurement, and constant blood pressure with small fluctuations is an important indicator of PVL occurrence.

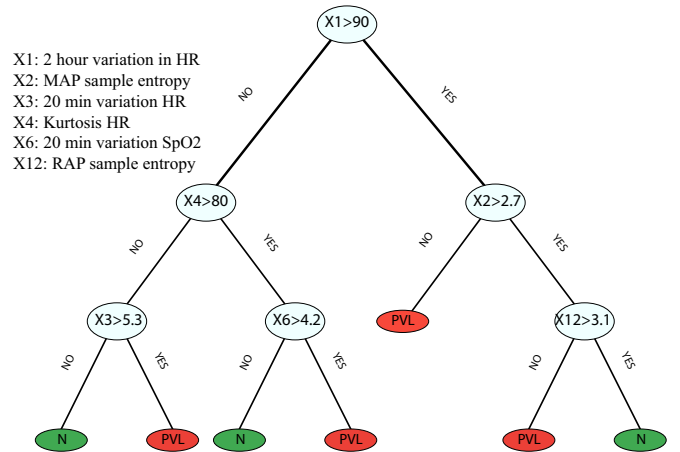


Fig. 2. Result of forming DT from the optimal feature set derived from vital measurements. The tree is pruned at two levels based on Fishers exact test (FET).

The classification accuracy of the designed classifier is shown using the receiver operating characteristic (ROC) curve. The higher area under the curve shows the higher classification accuracy. The plot shows high accuracy in classification although the relatively small data set makes it impossible to have a smooth curve.

In the final step of the work, we would like to know how long the data collection is sufficient for timely prediction of the PVL. To the best of our knowledge there is no study done so far to find the minimum length

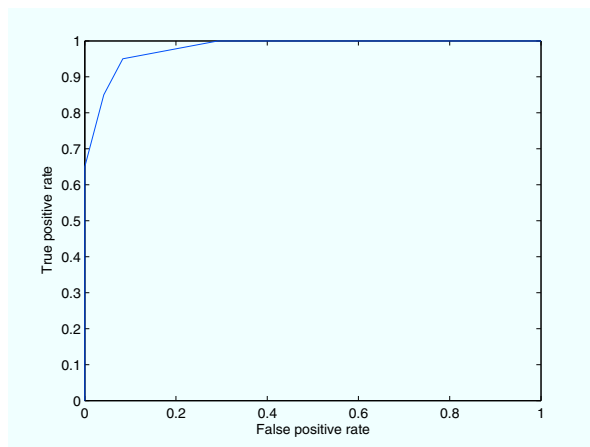


Fig. 3. Receiver operating characteristic (ROC) curve (plot of true positive rate vs. false positive rate).

of data needed after neonatal heart surgery for PVL prediction. To this end we trained and tested the DT with first 2, 4, 6 and 8 hours of data and we compared the prediction results with the complete data length. We used sensitivity, positive predictivity and their geometric mean, F-score, as criteria to evaluate the performance of the classifier. Table III shows that 6 hours of the data contains sufficient information for reliable PVL prediction. Another way to interpret this result is that after 6 hours the chance to prevent PVL will decrease significantly. If validated further, this would clearly be very important information for clinical practice.

TABLE III

PREDICTION ACCURACY FOR DIFFERENT LENGTH OF DATA. TP: TRUE POSITIVE, FP: FALSE POSITIVE, TN: TRUE NEGATIVE, FN: FALSE NEGATIVE.

DATA	TP	FP	FN	TN	Se	PP	F-Score
All	19	2	1	22	95%	90%	0.93
8Hours	19	2	1	22	95%	90%	0.93
6Hours	18	1	2	23	90%	95%	0.93
4Hours	16	4	4	20	80%	80%	0.80
2Hours	15	4	5	20	75%	79%	0.76

ACKNOWLEDGMENT

The research reported in this paper is supported by a grant from National Institute of Health (No. 1 R01 NS 72338 01A1). Authors are very thankful of Dr. Erin Buckley, Mrs. Jennifer M. Lynch, Peter J. Schwab, from CHOP for their endless efforts in data collection.

REFERENCES

- [1] G. Wernovsky, A. J. Shillingford, and J. W. Gaynor, "Central nervous system outcomes in children with complex congenital heart disease." *Current Opinion in Cardiology*, vol. 20, no. 2, pp. 94–99, 2005.
- [2] S. Blackburn, "Central nervous system vulnerabilities in preterm infants, Part II." *The Journal of Perinatal & Neonatal Nursing*, vol. 23, no. 2, pp. 108–110, 2009.
- [3] S. P. Miller, P. S. McQuillen, S. Hamrick, D. Xu, D. V. Glidden, N. Charlton, T. Karl, A. Azakie, D. M. Ferriero, A. J. Barkovich, and D. B. Vigneron, "Abnormal brain development in newborns with congenital heart disease." *New England Journal of Medicine*, vol. 357, no. 19, pp. 1928–1938, 2007.
- [4] D. J. Licht, D. M. Shera, R. R. Clancy, G. Wernovsky, L. M. Montenegro, S. C. Nicolson, R. A. Zimmerman, T. L. Spray, J. W. Gaynor, and A. Vossough, "Brain maturation is delayed in infants with complex congenital heart defects." *The Journal of Thoracic and Cardiovascular Surgery*, vol. 137, no. 3, pp. 529–37, 2009.
- [5] J. W. Gaynor, "Periventricular leukomalacia following neonatal and infant cardiac surgery," *Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual*, vol. 7, pp. 133–140, 2004.
- [6] P. S. McQuillen and S. P. Miller, "Congenital heart disease and brain development," *Annals of the New York Academy of Sciences*, vol. 1184, pp. 68–86, 2010.
- [7] I. C. van Haastert, L. S. de Vries, M. J. C. Eijssermans, M. J. Jongmans, P. J. M. Helders, and J. W. Gorter, "Gross motor functional abilities in preterm-born children with cerebral palsy due to periventricular leukomalacia," *Developmental Medicine and Child Neurology*, vol. 50, no. 9, pp. 684–689, 2008.
- [8] D. J. Licht, J. Wang, D. W. Silvestre, S. C. Nicolson, L. M. Montenegro, G. Wernovsky, S. Tabbutt, S. M. Durning, D. M. Shera, J. W. Gaynor, T. L. Spray, R. R. Clancy, R. A. Zimmerman, and J. A. Detre, "Preoperative cerebral blood flow is diminished in neonates with severe congenital heart defects." *The Journal of Thoracic and Cardiovascular Surgery*, vol. 128, no. 6, pp. 841–849, 2004.
- [9] J. Chen, R. A. Zimmerman, G. P. Jarvik, A. S. Nord, R. R. Clancy, G. Wernovsky, L. M. Montenegro, D. M. Hartman, S. C. Nicolson, T. L. Spray, J. W. Gaynor, and R. Ichord, "Perioperative stroke in infants undergoing open heart operations for congenital heart disease." *Annals of Thoracic Surgery*, vol. 88, no. 3, pp. 823–829, 2009.
- [10] A. Jalali, D. J. Licht, and C. Nataraj, "Application of decision tree in the prediction of periventricular leukomalacia (pvl) occurrence in neonates after heart surgery." *Proceedings IEEE International Conference Engineering in Medicine and Biology Society (EMBC)*, vol. 2012, pp. 5931–5934, Aug 2012.
- [11] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy." *American Journal of Physiology: Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [12] J. R. Moorman, J. B. Delos, A. A. Flower, H. Cao, B. P. Kovatchev, J. S. Richman, and D. E. Lake, "Cardiovascular oscillations at the bedside: early diagnosis of neonatal sepsis using heart rate characteristics monitoring." *Physiological Measurements*, vol. 32, no. 11, pp. 1821–1832, 2011.
- [13] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample entropy analysis of neonatal heart rate variability." *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 283, no. 3, pp. R789–R797, 2002.
- [14] K. Kappaganthu, "An integrative approach for machinery prognostics," Ph.D. dissertation, Villanova University, 2011.