ELSEVIER

Computers in Biology and Medicine 36 (2006) 21–40

Computers in Biology and Medicine

www.intl.elsevierhealth.com/journals/cobm

Hypoplastic left heart syndrome: knowledge discovery with a data mining approach

Andrew Kusiak[a,*], Christopher A. Caldarone[b], Michael D. Kelleher[c], Fred S. Lamb[d], Thomas J. Persoon[e], Alex Burns[a]

[a]*Intelligent Systems Laboratory, MIE 3131, Seamans Center, The University of Iowa, Iowa City, Iowa 52242 - 1527, USA*
[b]*Division of Cardiovascular Surgery, The Hospital for Sick Children, University of Toronto, 555 University Avenue, Toronto, Ontario, M5G 1X8, Canada*
[c]*Department of Pediatrics, Division of Critical Care, Children's Memorial Hospital, 2300 Children's Plaza, Box 73 Chicago, IL 60614, USA*
[d]*Department of Pediatrics, The University of Iowa Hospital and Clinics, The University of Iowa, Iowa City, IA 52242, USA*
[e]*Department of Pathology, The University of Iowa Hospital and Clinics, The University of Iowa, Iowa City, IA 52242, USA*

Received 16 March 2004; accepted 19 July 2004

Abstract

Hypoplastic left heart syndrome (HLHS) affects infants and is uniformly fatal without surgical palliation. Post-surgery mortality rates are highly variable and dependent on postoperative management. A data acquisition system was developed for collection of 73 physiologic, laboratory, and nurse-assessed parameters. The acquisition system was designed for the collection on numerous patients. Data records were created at 30 s intervals. An expert-validated wellness score was computed for each data record. To efficiently analyze the data, a new metric for assessment of data utility, the combined classification quality measure, was developed. This measure assesses the impact of a feature on classification accuracy without performing computationally expensive cross-validation. The proposed measure can be also used to derive new features that enhance classification accuracy. The knowledge discovery approach allows for instantaneous prediction of interventions for the patient in an intensive care unit. The discovered knowledge can improve care of complex to manage infants by the development of an intelligent bedside advisory system.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Hypoplastic left heart syndrome; Data mining; Medical knowledge discovery; Classification accuracy; Classification quality; Medical decision making

---

* Corresponding author. Tel.: +1 319 335 5934; fax: +1 319 335 5669.
E-mail address: andrew-kusiak@uiowa.edu (A. Kusiak).
URL: http://www.icaen.uiowa.edu/∼ankusiak.

0010-4825/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2004.07.007

1. Introduction

Hypoplastic left heart syndrome (HLHS) is a heart disease of newborn infants. The occurrence of HLHS is rare, effecting between 0.16 and 0.36 in every 1000 births, but it is inevitably fatal without surgical intervention [1]. The Norwood procedure has emerged in the last decade as the most common treatment [2] and consists of a three stage surgical intervention [3]. Stage I of the Norwood procedure includes three main components: an atrial septectomy, an anastomosis of the proximal pulmonary artery to the aorta with homograft augmentation of the aortic arch, and an aortopulmanary shunt. As a result of this procedure the patient's right ventricle is connected to the aorta so that it can force the delivery of oxygenated blood through the braches of the aorta.

Although the procedure is lifesaving, a consequence of this reconstruction is the creation of a "balanced circulation" which implies a precarious metastable balance between the pulmonary and systemic circulations. The most critical time for the neonate is the surgery itself and the time immediately following surgery spent in the pediatric intensive care unit (PICU) [3]. Typically, complications are attributed to the unstable balance between the pulmonary and systemic circulation. There are rapid and massive shifts in the cardiac output, pulmonary resistance, and systemic resistance for the first 3–4 days after surgery. An experienced team of physicians, nurses, and therapists is required to successfully navigate the changes in this period. However, even the most experienced teams report significant mortality due to the extremely complex relationships among physiologic parameters in a given patient. The mortality rate for the three-stage procedure is highest following the first stage [4] and can reach as high as 42% [2].

Specifically, the inability to directly measure crucial parameters in the postoperative infant results in the need for physicians to infer the value of crucial, but immeasurable, parameters from a group of obtainable parameters used to monitor the infant. Obtainable postoperative parameters include: pulse, heart rhythm, systemic blood pressure, common atrial filling pressure, urine output, physical exam, and systemic and mixed venous oxygen saturations. Based on these values, inferences are made as to the value of crucial life-saving parameters (e.g., pulmonary and systemic blood flow). These parameters change rapidly in the postoperative period, and subtle constellations of changes in the obtainable parameters are often unnoticed by the inexperienced caregiver but lead to a "sudden" postoperative death. Closer analysis of the medical record often reveals clusters of changes that should have signaled a modification of the direction of postoperative therapy.

An experienced team can assimilate the numerous measurable parameters and, using experience-based intuition, infer the value of crucial control parameters, thereby managing a neonate with more success during the critical postoperative period.

There are two categories of problems involved in caring for the infants after surgery. The first issue involves the nature of the decision-making required for the care. Decisions made even by the most experienced physician are far from being ideal as the relationships between the measurable and inferred parameters are highly complex, nonlinear, and frequently not known.

The second problem involves the difficulty in communicating correct response patterns. These response patterns are commonly referred to as "wisdom". In this setting, wisdom is the ability to successfully interpret the multiple, complex, and unknown relationships between the parameters available and to generate a successful therapeutic plan. Because these relationships are complex, the transfer of this wisdom from experienced to inexperienced personnel is extremely difficult, even within a single institution. Furthermore, the transfer of wisdom from a high-volume health center to a low-volume center is even more difficult. Consequently, infants treated in low-volume centers are denied the benefit of the wisdom available in high-volume centers. Furthermore, even within a high-volume institution, the lack of continuous (24 h a day) supervision by experienced personnel at the bedside can deny a critically-ill infant the benefit of available wisdom.

An essential step to understanding the complex relationships between parameters is being able to define and predict the health of a patient. In this paper, a data mining approach is proposed to capture the complex interactions among physiological variables and therapeutic interventions. The approach yields a set of rules that are both easily interpretable and highly accurate. These rules will be used to predict the health status of a postoperative neonate, interventions, and other user defined outcomes. The data mining approach is aided by a new metric for the selection of features (parameters) to be transformed that in turn leads to higher accuracy in predicating the "wellness score" of a patient. Every percent increase in prediction accuracy is significant because it improves the overall understanding of postoperative management.

The next section details the collection of data and the development of the "wellness score". The score is essential to the data mining algorithms and it is used as an indicator of the patient's health.

2. Data collection

The patients who were subjects in this study had been diagnosed with HLHS and had undergone the first stage Norwood procedure for palliation of HLHS. There were no other selection criteria.

Data collection began upon admission of the patient to the PICU immediately after Norwood surgery and lasted between 18 and 36 h. Three categories of data were collected: continuously monitored physiologic parameters, intermittently monitored physiologic parameters, and interventions.

The continuously monitored physiologic parameters are listed in Table 1. Transducers to measure these parameters were attached or implanted in the patient and were connected to a bedside monitoring device Individual patient monitors transmitted the data to a patient data server (PDS), which collected the data from all patients in the PICU. Collected data was transmitted via an Ethernet to a data collection server. Software developed by the research team pulled parameters of interest from the data stream and stored them in a database. The system collected and added data to the data set in 30 s intervals. Two data sets, one with 5134 records and the other with 14,000 records were collected. The first data set was composed of data collected from six patients and the later was generated from eleven patients. Though the results reported in this paper are based on the first data set, the computational results generated from the second data set confirmed the former. All of the parameters that were collected were determined by clinical experts.

Table 1
Continuously monitored parameters

Parameter	Abbreviation
Pulse	Pulse
Respiration	Resp
Systolic blood pressure	SBP
Diastolic blood pressure	DPB
Mean blood pressure	MBP
Central venous pressure	CVP
Oxygen saturation via pulse oximeter	SaO_2

Table 2
Intermittently monitored parameters

Parameter	Abbreviation
Venous oxygen saturation	SvO_2
Hemoglobin concentration	HgB
Blood sodium concentration	Na
Blood potassium concentration	K
Blood chloride concentration	Cl
Blood carbon dioxide concentration	CO_2
Blood ionized calcium concentration	Ca++
Blood glucose concentration	Glu
Blood pH	pH
Blood partial pressure of CO_2	pCO_2
Blood partial pressure of oxygen	pO_2
Blood carbon dioxide concentration	HCO_3
Blood base excess	Base XS
Blood lactic acid concentration	Lactic
Radial arterial blood pressure	Rad ABP
Femoral arterial blood pressure	FABP
Color	Skin color
Presence or absence of bowel sound	Bowel sound
Condition of patient’s abdomen	Abdomen
Patient’s breath sounds	Breath sound
Peripheral pulse, left upper extremity	PPLUE
Peripheral pulse, right upper extremity	PPRUE
Peripheral pulse, left lower extremity	PPLLE
Peripheral pulse, right lower extremity	PPRLE
Capillary refill, seconds	Cap refill
Right pupil size	R pupil
Right pupil reaction to light	R react light
Left pupil size	L pupil
Left pupil reaction to light	L react light
Urine output volume	Urine vol
Chest tube drainage volume	CT vol
Mediastinal tube volume	MT vol
Total fluid output	Fluid sum

The intermittently monitored parameters consisted of laboratory data and physiologic parameters observed at the bedside by nurses. These parameters were obtained retrospectively from nursing assessment records (flow sheets) and were added to the database (see Table 2). These parameters were collected as the parameter values were changed and updated; there was no regular data collection interval.

Interventions involved actions taken by the caregivers to treat the patients, primarily administration of medications, and are listed in Table 3. Both continuously administered intravenous medications and intermittently administered medications were included; however substances administered primarily as vehicles for other medications (i.e., normal saline solution) were not included. These parameters were collected intermittingly as their respective values changed.

Table 3
Interventions

Intervention	Explanation
Administer epinephrine	Administration of ionotropic drug ephinephrine via a continuous intravenous drip
Administer norepinephrine	Administration of ionotropic drug norephinephrine via a continuous intravenous drip
Administer dopamine	Administration of ionotropic drug dopamine via a continuous intravenous drip
Administer milrinone	Administration of ionotropic drug milrinone via a continuous intravenous drip
Change ventilator	make a change in the ventilator settings
Administer HCO3	Administration of an intravenous bolus of sodium bicarbonate
Administer FFP	Intravenous administration of fresh frozen plasma, a blood product
Administer KCl	Administration of an intravenous bolus of potassium chloride
Administer CaCl	Administration of an intravenous bolus of calcium chloride
Administer PRBC	Intravenous administration of packed red blood cells, a blood product
Administer Lasix	Administration of an intravenous bolus of furosemide (Lasix), a diuretic drug
Administer albumin	Intravenous administration of albumin solution
Ice to heart	Apply ice chips to the exposed heart
Pacer on/off	Apply signal from an external heart pacemaker
Administer amiodarone	Administration of an intravenous bolus of amiodarone
Liver pressure	Apply manual pressure to the patient’s liver
Epi bolus	Administration of an intravenous bolus of epinephrine
Change fluid input	Change the flow rate of intravenously administered normal saline solution
Administer platelets	Intravenous administration of platelets, a blood product
Administer atropine	Administration of an intravenous bolus of atropine
CPR	Administer external cardiopulmonary resuscitation
Administer insulin	Administration of an intravenous bolus of insulin
Internal heart compression	Apply internal (open chest) heart compression
Defib	Administer signal from an external heart defibrillator
Administer Mg	Administration of a bolus of magnesium chloride solution
Nipride	Administration of a bolus of nipride

Table 4
Example of a data object

Data Object	Time	Patient ID	Resp	K	CVP	DSP	Wellness
1	12:00:00	2X	88	1.2	2	48	5

In total data on 73 parameters were collected and stored in the database. For each 30 s interval during data collection a data object was created. A data object is defined as a point in time as a vector of 73 parameter values. These parameters included all continuous and intermittently assessed variables as well as interventions. Data objects were organized in a spreadsheet, with the rows representing data objects and the columns representing attributes. The time value of each object was recorded as the time in minutes following postsurgical admission to the PICU. An example of a data object can be seen in Table 4. Note that the data object contains the time, the data object number, the parameter values, and the wellness score for that specific data object. The sample data object displays only a small selection of the parameter values, whereas a full object will contain the values of all 73 parameters.

Table 5
Elements of the wellness score

Parameter	Range for Wellness Score $= 1$	Range for Wellness Score $= 0$
Pulse	$150 \leqslant \text{Pulse} \leqslant 170$	Otherwise
CVP (Preload)	$8 \leqslant \text{CVP} \leqslant 41$	Otherwise
SaO_2	$65 \leqslant SaO_2 \leqslant 09$	Otherwise
SvO_2	$\text{SvO2} \geqslant 20$	$\text{SvO2} < 20$
Blood base excess	Blood base excess $\geqslant -5$	Blood base excess < -5
Hemoglobin (Hb)	$\text{Hb} \geqslant 12$	$\text{Hb} < 12$
pCO_2	$35 \leqslant pCO_2 \leqslant 50$	Otherwise
Urine output	Urine output > 0.5 cc/kg/h	Urine output < 0.5 cc/kg/h
Ionotropic support	None	See Table 5 for the formula

Table 6
Wellness score for ionotrope support (doses in mg/kg/h)

Ionotrope	Score $=0$	Score $=0.5$	Score $=1$
Epinephrine dose	0	< 0.1	$\geqslant 0.1$
Dopamine dose	0	< 5	$\geqslant 5$
Milrinone dose	0	< 0.75	$\geqslant 0.75$
Norepinephrine dose	0	< 0.1	$\geqslant 0.1$

For each data object, an expert-validated wellness score was determined. All the physiological parameters were selected by domain experts. Domain experts also determined acceptable ranges for the wellness score components based on clinical experience. The individual elements of the wellness score are listed in Table 5. If a wellness score attribute fell within an acceptable range the element was scored as 1; otherwise it was scored as 0. The clinical experts also determined that one intervention, administration of ionotropic drugs, should also be a constituent of the wellness score. If a patient required no ionotropic support, the ionotrope wellness score is 1. The ionotrope wellness score was diminished by addition of each iontrope according to the scheme shown in Table 6. The ionotrope portion of the wellness score could range from 1 (no ionotropes) to -3 (four ionotropes administered at maximum concentration). The total wellness score is the sum of the individual wellness score components.

The wellness score was computed to measure the health of a patient from parameter values that were easily obtainable. High wellness scores indicate that the patient is doing well and low scores indicate that the patient is doing poorly. An example of a computed wellness score can be seen in Table 7. The first row of values indicates the actual parameter values obtained from the data acquisition system and the second row displays what each parameter value contributes towards the overall wellness score. In this example the total of the parameters results in a wellness score of 6.

Preliminary results also indicate there is a positive correlation between the wellness score and the more accepted oxygen excess factor (O) expressed by the formula

$$O = SaO_2/(SaO_2 - SvO_2).$$

Table 7
Example wellness score

Physiological parameter								Ionotropic dose			
Pulse	CVP	SaO_2	SvO_2	Blood base excess	Hemo-globin	pCO_2	Urine Output	Epinep--hrine	Dopami-ne	Milrin-one	Norepin-ephrine
145	7.6	92.1	22.6	−1.8	9.9	33.4	0.87	0.12	4.2	0	0.68
1	0	0	1	1	0	0	1	1	0.5	0	0.5

The oxygen excess factor is used in place of the Qp/Qs (pulmonary to systemic flow) ratio. This ratio is used to determine the maximum oxygen delivery, but is difficult to calculate and often has to be estimated [5].

A comparison was made between the oxygen excess factor and the wellness score. The values for the SvO_2 were taken directly from the nurse's flow sheets and were stored in the database under the variable name SVO_2_man. By using the values from the flow sheets our sample was reduced to 126 points. The correlation between the oxygen excess factor and wellness score was then calculated to be 25.22%. While this value only indicates a slight correlation between the oxygen excess factor and wellness score, novel methods of data transformation could demonstrate a stronger relationship. The wellness score could become an alterative to gauging a patient's status condition as it is easy to compute in real time. This could provide an instantaneous score regarding the patients' current condition.

The wellness score was required due to the fact the data mining algorithms call for an outcome for each data object. The wellness score was used by the data mining algorithms to predict the sudden and often unpredictable swoons patients could suffer. During these swoons, teams of physicians and caregivers apply interventions (see Table 3) to correct the problems. For instance, if the patient's hemoglobin started to decrease to an undesirable level the team would commonly administer packed red blood cells, labeled "Administer PBRC" in Table 3, to rectify the situation. This is just one example of the many described in the database. Using the methods in the following sections the patients wellness score was predicted with a high degree of accuracy and intuitive rules were derived that could be used to facilitate care and management.

3. Methods

In this paper, a machine learning algorithm based on the rough set theory [6] is used. This algorithm represents a large class of algorithms generating decision rules from data. The use of learning algorithms for extraction of explicit knowledge from the HLHS data is novel. The literature on data analysis of HLHS is limited. Alonso–Betanzo [7] discussed the application of a neural network (NN) to predict fetal outcome based on the nonstress test (NST), which is used to evaluate intrauterine fetal condition and to identity risk factors. The investigators also used a Bayesian algorithm and a logistic regression approach. Their results indicated that NN was the best method for predicting the results of the NST. Tsien et al. [8] discussed the use of decision trees to detect patterns from physiological data signals in a neonatal ICU for the development of an alarm system. Several other examples of data mining applications include knowledge discovery in immunodeficiency virus patients, visual imaging, and disease diagnosis [9].

Table 8
Illustrative data set

Object No.	F1	F2	F3	F4	D
1	Yes	0	3	Medium	Good
2	No	4	1	Low	Bad
3	Maybe	0	3	Medium	Good
4	Yes	2	3	High	Good
5	No	2	1	Low	Bad
6	No	0	1	Low	Bad
7	Yes	3	1	High	Good
8	Maybe	1	1	Low	Bad

Table 9
Reducts of the data set in Table 8

Reduct No.	Reduct	Reduct Length
1	F1, F2	2
2	F1, F3	2
3	F2, F3	2
4	F4	1

The approach proposed in this paper extracts explicit decision rules that increase understanding the HLHS and can be used to predict outcomes. The decision rules have the following form

IF physiological parameters/interventions THEN Outcome (e.g., wellness score).

One of the basic terms used in rough set theory is reduct. It is defined as a minimal sufficient subset of features such that [10]:

(a) The reduct produces the same categorization of objects as the collection of all features.
(b) A reduct is a minimal subset of features with respect to the property (a).

Some of the above introduced terms and other concepts used in rough set theory are illustrated next based on the data in Table 8.

All reducts for the data set in Table 8 are listed in Table 9.

It can be seen in Table 10 that in reduct 1 all outcomes can accurately be described using only the two features (F1, F2).

A simple set of rules can be generalized from the reducts to accurately predict the outcome (Good, Bad).

Based on reduct 1 the following rules can be constructed:

Rule 1: IF (F1 = Yes) THEN (D = Good)

Rule 2: IF (F1 = No) THEN (D = Bad)

Rule 3: IF (F1 = Maybe and F2 = 0) THEN (D = Good)

Rule 4: IF (F1 = Maybe and F2 = 1) THEN (D = Bad)

Table 10
Reduct No. 1 from Table 9

Object No.	F1	F2	D	Rule
1	Yes	0	Good	
7	Yes	3	Good	Rule 1
4	Yes	2	Good	
2	No	4	Bad	
5	No	2	Bad	Rule 2
6	No	0	Bad	
3	Maybe	0	Good	Rule 3
8	Maybe	1	Bad	Rule 4

The rough set approach was used to extract rules from the data collected at the PICU. We have also developed a novel metric that was applied to select features for transformation. The transformed features are incorporated into the original data set to improve overall classification accuracy.

The proposed use of decision rule algorithms has several advantages over other machine learning methods and classical statistical approaches, namely:

- Generation of explicit knowledge in a form acceptable by a user. The user is able to understand the extracted knowledge, assess its usefulness, and learn new and interesting concepts.
- Controlled prediction accuracy. This characteristic is due to the nature of the decision-making approach considered in this paper.
- Evolution of knowledge. Most machine-learning algorithms are designed for extracting static knowledge. The feature transformation scheme used in this paper makes the knowledge time invariant.

For complete review of data mining techniques the reader may refer to [11].

4. Extension of the classification accuracy measure

Classification quality (CQ) is the measure of association between a feature and the outcome (e.g., wellness score). Its role in rough set theory can be compared to that of the correlation coefficient in statistics. For a given feature, CQ can be loosely defined as the ratio of objects with non-conflicting feature values to the total number of objects in the data set. Formal definitions of the classification quality and its extensions are presented in the Appendix. An example calculation of CQ is shown next using the artificial data of Table 11.

Consider the feature SaO_2 in Table 11. The two values $SaO_2 = 59$ are associated with the Wellness score $= 5$ and therefore it makes $2/5 = .4$ contribution to the CQ. The value of $SaO_2 = 62$ leads to two different outcomes, Wellness score $=3$ and Wellness score $=4$ and therefore does not contribute to the CQ. The classification quality of feature SaO_2 is $\mathrm{CQ}(SaO_2) = .4$. The classification quality of the Temperature is CQ(Temperature)$=5/5 = 1$. In both examples the value of 5 in the denominator is determined by the total number of objects in the data set.

Classification quality in its original form does not adequately characterize the features in terms of prediction accuracy, i.e., higher value of the CQ does not necessary imply higher classification accuracy.

Table 11
Example data set

Object	SaO_2	Temperature	Wellness Score
1	62	97	4
2	62	97	4
3	59	98	5
4	59	98	5
5	62	96	3

To alleviate the deficiency of the existing CQ, the following two new measures relating the properties of a data set with the classification accuracy are introduced:

- Reflective classification quality (RCQ),
- Combined classification quality (CCQ).

The reflective classification quality (RCQ) measure representing the ratio of the number of different classes and the number of feature values is presented next.

$$\mathrm{RCQ} = o_\mathrm{p}/n_\mathrm{f},$$

where

n_f = the number of distinct values of feature n,

o_p = the number of distinct values of the outcome.

Using the data in Table 11 the value for $n_\mathrm{temperature}$ equals 3 (96, 97, 98) and the value for o_wellness equals 2 (4, 5). The RCQ for the feature SaO_2 has a value 1 because the set for n_{SaO2} has only two members (59, 62) and the value for o_wellness still equals to 2. Note that if the values of o_p and n_f are equal, the value of $\mathrm{RCQ} = 1$. A data set with such properties is in fact a "perfect" set for machine learning as the number of rules equal to the number of classes (decision values).

The combined classification quality (CCQ) measure is defined as

$$\mathrm{CCQ}_P = \mathrm{CQ}_\mathrm{P} \cdot \mathrm{RCQ}_\mathrm{P}.$$

The CCQ is the result of the CQ and RCQ values for the same feature multiplied to each other.

The combined classification quality measure is of vital importance as it can be used for feature selection and construction of new features, thus increasing classification accuracy and limiting cross-validation, which is computationally expensive.

5. Feature transformation

In this section, the relationship between the combined classification quality measure and classification accuracy is discussed. In addition, the combined classification quality measure is used to derive new features increasing classification accuracy of the rules extracted from a transformed data set.

Confusion Matrix

	Good	Bad	None
Good	3	1	0
Bad	1	3	0

Classification Accuracy[%]

	Correct	Incorrect	None
Good	75	25	0
Bad	75	25	0
Average	75	25	0

Fig. 1. Cross-validation results based on the data scenario with feature F1.

Confusion Matrix

	Good	Bad	None
Good	0	1	3
Bad	4	0	0

Classification Accuracy [%]

	Correct	Incorrect	None
Good	0	25	75
Bad	0	100	0
Average	0	62.5	37.5

Fig. 2. Cross-validation results based on the data scenario with feature F2.

To establish the relationship between the combined classification quality measure and classification accuracy, four different data scenarios (each with one feature only) of the data set in Table 8 have been created. For example, the fist scenario is the data set of Table 8 with F1 only; the second is based on F2 data only, and so on. The cross-validation results with the one-out-of-*n* scheme for the four single feature scenarios are shown in Figs. 1–4.

The introduced terms and other concepts used in rough set theory are illustrated next based on the data in Table 8. The values of the three classification quality measures CQ, RCQ, and CCQ for the individual features F1–F4 are shown in Table 12.

The comparison of the combined classification quality (CCQ) measure of Table 12 with the average classification accuracy (CA) values of Figs. 1–4 indicate an interesting relationship between these two measures, namely, higher values of CCQ tend to produce higher values of CA. The latter observation does not hold for the classification quality (CQ) measure of rough set theory. The noted relationship between

Confusion Matrix

	Good	Bad	None
Good	3	1	0
Bad	0	4	0

Classification Accuracy [%]

	Correct	Incorrect	None
Good	75	25	0
Bad	100	0	0
Average	87.5	12.5	0

Fig. 3. Cross-validation results based on the data scenario with feature F3.

Confusion Matrix

	Good	Bad	None
Good	4	0	0
Bad	0	4	0

Classification Accuracy [%]

	Correct	Incorrect	None
Good	100	0	0
Bad	100	0	0
Average	100	0	0

Fig. 4. Cross-validation results based on the data scenario with feature F4.

Table 12
Classification quality, reflexive classification quality, and combined classification quality measures for the data set in Table 8

	F1	F2	F3	F4
CQ	0.75	0.375	0.375	1
RCQ	0.67	0.40	1	0.67
CCQ	0.5	0.15	0.375	0.67

the CCQ and CA is of interest to data mining for various reasons. First of all, the combined classification quality measure allows assessing the classification accuracy without performing the computationally expensive cross-validation [12]. Secondly, the combined classification quality measures of individual features will be used to derive new features, e.g., a joint feature (a set) of existing features. Again, the

Table 13
Data set with the derived joint features F1_F2, F1_F3, and F2_F3

Object No.	F1	F2	F3	F4	F1_F2	F1_F3	F2_F3	D
1	Yes	0	3	Medium	Yes_0	Yes_3	0_3	Good
2	No	4	1	Low	No_4	No_1	4_1	Bad
3	Maybe	0	3	Medium	Maybe_0	Maybe_3	0_3	Good
4	Yes	2	3	High	Yes_2	Yes_3	2_3	Good
5	No	2	1	Low	No_2	No_1	2_1	Bad
6	No	0	1	Low	No_0	No_1	0_1	Bad
7	Yes	3	1	High	Yes_3	Yes_1	3_1	Good
8	Maybe	1	1	Low	Maybe_1	Maybe_1	1_1	Bad

Confusion Matrix

	Good	Bad	None
Good	3	0	1
Bad	4	0	0

Average Classification Accuracy [%]

	Correct	Incorrect	None
Good	75	0	25
Bad	0	100	0
Average	37.5	50	12.5

Fig. 5. Cross-validation results based on the data scenario with derived feature F1_F2.

impact of the derived features on the classification accuracy can be evaluated a priori (for more details see [13]).

The concept of derived features is illustrated with the data in Table 13, where the joint features F1_F2, F1_F3, and F2_F3 have been derived based on the data in Table 8.

Each of the three data sets with one of the three joint features F1_F2, F1_F3, and F2_F3 was cross-validated with the one-out-of-*n* scheme. The cross-validation results are shown in Figs. 5–7.

The values of all three classification quality measures and the classification quality for all seven single-feature data sets are compiled in Table 14.

The following conclusions can be derived from the results in Table 14:

(a) Features F2 and F3 have the same value of classification quality CQ = 0.375, however, the values of the combined classification quality CCQ differ. The higher value of CCQ = 0.375 corresponds to the higher value of the average classification accuracy CA = 87.5%.
(b) The joint feature F1_ F3 was derived from the two individual features F1 and F3 with the highest value of the combined classification quality CCQ(F1) =0.5 and CCQ(F3) =0.375 among the features with the classification quality CQ $<$ 1. The resultant value of CQ for this derived feature is 1.

Confusion Matrix

	Good	Bad	None
Good	3	0	1
Bad	1	3	0

Average Classification Accuracy [%]

	Correct	Incorrect	None
Good	75	0	25
Bad	75	25	0
Average	75	12.5	12.5

Fig. 6. Cross-validation results based on the data scenario with derived feature F1_F3.

Confusion Matrix

	Good	Bad	None
Good	2	1	1
Bad	2	0	2

Average Classification Accuracy [%]

	Correct	Incorrect	None
Good	50	25	25
Bad	0	50	50
Average	25	37.5	37.5

Fig. 7. Cross-validation results based on the data scenario with derived feature F2_F3.

Table 14
Classification quality (CQ), reflexive classification quality (RCQ), combined classification quality (CCQ), and average classification accuracy (CA) for the data set in Table 13

	F1	F2	F3	F4	F1_F2	F1_F3	F2_F3
CQ	0.75	0.375	0.375	1	1	1	1
RCQ	0.67	0.4	1	0.67	0.25	0.4	0.29
CCQ	0.5	0.15	0.375	0.67	0.25	0.4	0.29
Av CA	75%	0%	87.5%	100%	37.5%	75%	37.5%

(c) The value of the combined classification quality for the derived feature F1_F3 is the highest (CCQ = 0.4) among the three derived features F1_F2, F1_F3, and F2_F3 and it has led to the highest value of the classification accuracy CA = 75%.

Based on the values of the combined classification quality provided in Table 14 an additional data set with features F4 (CCQ =0.67) and F1_ F3 (CCQ =0.4) has been created. Each of the four cross-validation schemes, the one-out-of n ($n = 8$) and ($k = 2$, 3, 4) fold schemes, has resulted in perfect classification accuracy. Such perfect cross-validation results could not be obtained for any other data sets.

6. Benefits of the combined classification quality measure

The combined classification quality measure offers the following benefits:

(a) Feature selection: The combined classification quality measure evaluates features in the same way as the entropy measure, Gini index, and other metrics.
(b) Quick and low computational cost evaluation of the newly introduced features: The computational complexity of the combined classification quality measure is much lower than the cross-validation task involving multiple runs of machine learning algorithms.
(c) Tool for deriving new features by applying various feature transformation schemes: New features can be created by forming joint features (e.g., $F_i\_F_j\_F_k$), computing feature ratios (e.g., F_i/F_j), feature subtraction (e.g., $F_i - F_j$), and so on.
(d) Ease of consideration of user preferences: One of the tenants of machine learning is to generate knowledge in the form acceptable by the user. The feature transformation mentioned in item (c) above is one of the most effective ways of meeting the user expectations.
(e) Temporal knowledge: One of the goals of feature transformation might to derive features that are invariant in time and therefore would be suitable for temporal data sets. For example, the knowledge derived based on the ratio F_i/F_j of two features might be time invariant.
(f) Knowledge evolution: The form of new features may be suitable for evolution. For example, the function f of the derived feature $f(F_i - F_j)$ may evolve while the function argument may remain unchanged, two existing joint features $F_i\_F_j$ and $F_k\_F_l$ may be merged, and so on.

7. Computational results

To demonstrate that data mining algorithms could be applied to discover knowledge in the postoperative care of HLHS patients we designed a five staged experiment. The experiment demonstrates the effectiveness of both the data mining approach and the use of the proposed metric, CCQ, to select features for transformation and the benefits of using transformed features. The experiment was set up as follows:

Step 1: Calculate the CQ, RCQ, and CCQ for all of the collected features in the data set.
Step 2: Determine features with the highest values of CCQ.
Step 3: Use the features with the highest CCQ to create derived feature pairs.
Step 4: Create three subsets by removing features from the original data set. Each subset (scenarios A, B, and C discussed later in this section) contained fewer features than the number of features in the previous scenario. Furthermore, derived features could not be removed from any scenario.
Step 5: Apply the data mining algorithm to each scenario twice, once without any of the derived features included and once with the entire set of derived features included.

Table 15
Classification quality, reflexive classification quality, and combined classification quality measures for features selected for transformation

	CVP	SaO_2	Pulse	Milri_dose
RCQ	0.278	0.278	0.278	0.278
CQ	0.062	0.087	0.043	0.027
CCQ	0.0172	0.0241	0.0119	0.0075

Table 16
Classification quality, reflexive classification quality, and combined classification quality measures for transformed features

	CVP_SaO_2	Pulse_CVP	SaO_2_Milri_dose	Pulse_SaO_2	CVP_Milri_dose
RCQ	1.167	1.278	0.556	1.333	0.5
CQ	0.076	0.064	0.094	0.022	0.077
CCQ	0.0889	0.0817	0.0522	0.0293	0.0385

The RCQ, CQ, and CCQ were calculated for each feature. Features that had highest combined classification quality values were then selected for transformation. To calculate the RCQ continuous features were discretized. Feature values were discretized according to their frequency. That is, there were an equal number of observations in each of the discretized intervals. This was needed due to the fact that if two continuous features were formed into, e.g., a pair, there would be too many combinations, which could lead to over-fitting. In the data set considered in this research, four features had the same value of reflective classification quality, as each feature was discretized into five intervals. The four features that had the highest CCQ values were: CVP, SaO_2, Pulse, and Epi_dose as shown in Table 15.

Based on the four features of Table 15 with the highest CCQ values, five new features were derived (see Table 16).

Three computational scenarios were designed to demonstrate the significance of the CCQ in the selection of features that impact classification accuracy. In each scenario several features were removed from the data set before mining. The removal of the features was required due to the fact the algorithm produced accurate results in it original form. None of the original features involved in the derived features were removed from the data set (i.e., CVP, SaO_2, Pulse, Milri_dose). The scenarios were then analyzed for the classification accuracy twice, once without any derived features (in this case pairs) included and once with all the derived features included. The selection of features for removal was arbitrary.

In Scenario A, the features Resp_Rate, pH, pCO_2, and Pacer_on_off were removed from the original data set. The classification accuracy results are shown in Table 17.

In Scenario B, in addition to the features removed in Scenario A, three additional features MT, Dopa_dose, and Rad_ABP removed from the original data set. The classification accuracy results are shown in Table 18.

In Scenario C the features pO_2, HCO_3, Base_Xs were removed from the original data set in addition to all the features that were removed in scenario B. The classification accuracy results are shown in Table 19.

Table 17
Average classification accuracy (%) for Scenario A

	Correct	Incorrect	None
No pairs	87.00	10.78	2.22
All pairs	87.00	9.01	3.99

Table 18
Average classification accuracy (%) for Scenario B

	Correct	Incorrect	None
No pairs	87.10	9.83	3.07
All pairs	86.75	9.36	3.89

Table 19
Average classification accuracy (%) for Scenario C

	Correct	Incorrect	None
No pairs	86.77	10.37	2.86
All pairs	86.55	9.87	3.58

Table 20
Number of objects supporting each value of the wellness score

Wellness Score	0	0.5	1	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	Total
Number of Objects	0	4	1	6	18	4	41	39	77	760	462	658	655	1202	567	547	49	48	5138

The overall classification accuracy does not appear to be affected by the addition of the derived pairs; however the improvement can be seen by looking at the data in more detail. The majority of observations, 99.4%, that were collected had the value of wellness score of 3.5 or higher (see Table 20). The occurrences of wellness score values 0 through 3 was small and therefore difficult to predict accurately. Focusing on the wellness score values of 3.5 to 9 will demonstrate the power and effectiveness of the features derived based on the combined classification accuracy metric.

The classification accuracy for scores 0 though 3 is low regardless whether the derived features had been included. This is due to the small number of objects supporting these scores. The data results were analyzed for wellness scores higher than 3. The difference in classification accuracy for data sets with pairs and without pairs for each of the three scenarios is shown in Table 21. The percentage of correct predictions increased for each scenario by an average of 1.83%. Furthermore, the percentage of incorrect predictions decreased by an average of 4.43% and the percentage instances in which the algorithm could not make predictions increased by an average of 2.6%. The percentage increase in the “None” category over the

Table 21
Difference in classification accuracy (%) for wellness score values of 3.5 through 9

Scenario	Correct	Incorrect	None
A	1.14	−6.96	5.83
B	2.06	−2.31	0.26
C	2.30	−4.01	1.71
Average	1.83	−4.43	2.60

“Incorrect” is preferred as it allows more cautious approach to decision making. For an information tool in medical applications it is more beneficial not to make predictions rather than predict an outcome in error.

The overall classification accuracy of the three scenarios was not greatly effected by the addition of the derived features selected using the combined classification quality metric. The power of the CCQ was demonstrated by the overall improvement in the classification accuracy of the wellness score values that had a significant number of supporting data objects (observations).

8. Conclusions

In this paper, a data mining approach to postoperative management of infants with hypoplastic left heart syndrome was considered. To efficiently analyze the data, a new metric for assessment of data utility, called the combined classification quality (CCQ) measure, was developed. The power of the combined classification quality measure has resulted in improvement of classification accuracy of the wellness score.

The use of a data mining approach with the combined classification quality metric has proven to be successful in the discovery of complex relationships in the postoperative care of neonates with hypoplastic left heart syndrome. Medical experts have recognized the discovered rules (relationships) as novel and useful in clinical practice. These rules described previously unknown relationships between physiological parameters and interventions in postoperative care. With further investigation and more data, these rules should allow transferring “wisdom” to low-volume intuitions and improve in the care of neonates with hypoplastic left heart syndrome.

The use of the combined classification quality metric as a method for feature selection has proven to be highly effective. The combined classification quality has provided a low computational complexity method for selection of features in their original or derived form for improving classification accuracy. This method provides an efficient and low-cost alternative to computationally intensive “trial and error” methods for deriving new feature sets. Furthermore, the combined classification quality measure can be used with other measures as entropy, Gini index, and so on to evaluate features.

Acknowledgements

The authors would like to express appreciation to A. Glick for organizing some of the data sets used in the study, C.F. Yu for design and coding the data collection system, and Y. Gan for design and development of the user interface. The research has been partially funded by the Children’s Miracle Network.

Appendix A. Classification quality

Parameters

U	set of all objects
A	set of all features
n	number of classes
f_a	information function, $f_a\colon U \rightarrow V_a$ such that for any $a \in A$ and $x \in U$, $f_a(x) \in V_a$
I_P	indiscernibilty relation for a set P of features defined as $x I_P y \Leftrightarrow f_a(x) = f_a(y), \forall a \in A$
Y	scenario of U
P	subset of features, $P \subseteq A$
Y_i	subset i of objects, where $U_{i=1,n} Y_i = Y$, i.e., the sets Y_i are disjunctive classes of Y
$\underline{P}(Y_i)$	P-lower approximation of Y_i, $\underline{P}(Y_i)\{y \in U : I_P(y) \subseteq Y_i\}$
$\underline{P}(Y)$	$\{\underline{P}(Y_1), \ldots, \underline{P}(Y_i), \ldots, \underline{P}(Y_n)\} = P$-lower approximation of Y

Using the above notation the classification quality measure, $\mathrm{CQ}_P(Y)$, of rough set theory is defined as

$$\mathrm{CQ}_P(Y) = \Sigma_{i=1,n} \underline{P}(Y_i)|/|U|,$$

where $|\cdot|$ is the cardinality of the set $\cdot$.

This ratio of all P-correctly classified objects to all objects expresses the quality of approximation of classification Y by the set of features P. In this paper, the rough set notation $\mathrm{CQ}_P(Y)$ is abbreviated as CQ.

References

[1] M. McConnell, M.E. Elixson, The neonate suspected with congenital heart disease, Crit. Care Nurs. Q. 25 (3) (2002) 17–25.

[2] H.P. Gutgeell, J. Gibgon, Management of hypoplastic left heart syndrome in the 1990s, Am. J. Cardiol. 89 (7) (2002) 842–846.

[3] C. Wright, Cardiac surgery 2002: staged repair of hypoplastic left heart syndrome, Crit. Care Nurs. Q. 25 (3) (2002) 72–77.

[4] A. Tulloh, G. Sharland, J. Simpson, S. Rollings, E. Baker, S. Qureshi, E. Rosenthal, C. Austin, D.P. Anderson, Outcome of staged reconstructive surgery for hypoplastic left heart syndrome following antenatal diagnosis, Arch. Dis. Childhood 85 (6) (2001) 474–477.

[5] O. Barnea, E. Salloum, S. Chien, S. Koenig, A. Rossi, E. Austin, W.P. Santamore, Estimation of oxygen delivery in newborns with a univentricular circulation, Circulation 98 (1998) 1407–1413.

[6] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning About Data, Kluwer, Boston, MA, 1991.

[7] A. Alonso-Betanzos, E. Mosqueira-Rey, V. Moret-Bonillo, B. Baldonedo del Río, Applying statistical, uncertainty-based and connectionist approaches to the prediction of fetal outcome: a comparative study, Artif. Intell. Med. 17 (1) (1997) 37–57.

[8] C.L. Tsien, I.S. Kohane, N. McIntosh, Multiple signal integration by decision tree induction to detect artifacts in the neonatal intensive care unit, Artif. Intell. Med. 19 (3) (2000) 189–202.

[9] K.J. Cios (Ed.), Special issue on data mining in medicine, IEEE Eng. Med. and Biol. Mag. 19(4) (2000).

[10] N. Shan, W. Ziarko, H.J. Hamilton, N. Cercone, Using rough sets as tools for knowledge discovery, in: U.M. Fayyad, R. Uthurusamy (Eds.), Proceedings of the First International Conference on Knowledge Discovery and Data Mining, AAAI Press, Menlo Park, CA, 1995, pp. 263–268.

[11] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, Menlo Park, CA, 2001.
[12] M. Stone, Cross-validatory choice and assessment of statistical predictions, J. R. Stat. Soc. 36 (1974) 111–147.
[13] A. Kusiak, Feature transformation methods in data mining, IEEE Trans. Electron. Packag. Manuf. 24 (3) (2001) 214–221.

Alex Burns graduated from the University of Michigan with a BSE degree in Industrial and Operations Engineering. He is currently pursing his Master's degree in industrial engineering at The University of Iowa. His fields of interest include data mining, feature transformations, and knowledge discovery.

Christopher Caldarone is an Associate Professor at the University of Toronto and a Staff Cardiovascular Surgeon at the Hospital for Sick Children in Toronto. As a full time congenital heart surgeon, one of his primary clinical interests is the care of neonates. He has published in the numerous journals including the *Journal of Thoracic and Cardiovascular Surgery*, the *Annals of Thoracic Surgery*, and *Circulation*. He is a member of the American College of Surgeons, the American Academy of Pediatrics, and the Society of Thoracic Surgeons. His email address is christopher.caldarone@sickkids.ca.

Michael Kelleher is an Associate Professor of Pediatrics at the Northwestern University and is engaged in the practice of pediatric intensive care medicine at Children's Memorial Hospital in Chicago, Illinois. He is interested in medical informatics and the application of computational intelligence to the care of critically ill children. He is a Fellow of the American Academy of Pediatrics, a member of the American Thoracic Society, the American Medical Informatics Association, and the Society for Critical Care Medicine. He has published in journals sponsored by the ATS and serves as a reviewer for *Critical Care Medicine*. His E-mail address is m-kelleher@northwestern.edu.

Andrew Kusiak is a Professor of Industrial Engineering at the University of Iowa, Iowa City. He is interested in theory and applications of computational intelligence, data mining, and optimization in healthcare, pharmaceutical industry, product development, and manufacturing. He has published research papers in journals sponsored by AAAI, IEEE, IIE, INFORMS, ESOR, IFIP, IFAC, IPE, ISPE, and SME. He speaks frequently on international meetings, conducts professional seminars, and consults for industrial corporations. He serves on the editorial boards of 18 journals, and edits book series. He is the Editor-in-Chief of the *Journal of Intelligent Manufacturing.* His E-mail address is andrew-kusiak@uiowa.edu.

Fred S. Lamb is an Associate Professor of Pediatrics at the University of Iowa, Iowa City, Iowa. Trained as a Pediatric Cardiologist, he is the head of the Division of Pediatric Critical Care and Medical Director of the Pediatric Intensive Care Unit. His clinical interests include management of postoperative congenital heart repair patients with a particular emphasis on factors regulating vascular tone. His basic science research laboratory is funded by the National Institutes of Health and the American Heart Association to study the physiologic role of chloride ion channels in determining the contractility of vascular smooth muscle. His E-mail address is fred-lamb@uiowa.edu.

Thomas Persoon is a Management Engineer in the Department of Pathology and an Adjunct Instructor of Industrial Engineering at the University of Iowa, Iowa City, Iowa. He graduated from the University of Iowa with an MS degree in Industrial Engineering. He is interested in process modeling, data flow, project management, and data mining.