Journal of Medical Systems (2019) 43: 8
https://doi.org/10.1007/s10916-018-1125-0

SYSTEMS-LEVEL QUALITY IMPROVEMENT

Extracting Interesting Rules from Gestation Course Data for Early Diagnosis of Neonatal Hypoxia

Inna Skarga-Bandurova [1] · Tetiana Biloborodova [1] · Maksym Nesterov [1]

Received: 5 May 2018 / Accepted: 20 November 2018 / Published online: 1 December 2018

Abstract
The topic of neonatal hypoxia is of paramount importance to anyone who cares during pregnancy and childbirth. Modern medicine associates this pathology with severe problems in the prenatal period. Underlying diseases of the mother during pregnancy, her anamnesis of life are the leading causes of complications in the newborn. Nevertheless, patterns of fetal hypoxia and neonatal hypoxia, as well as mechanisms of hypoxic-ischemic encephalopathy in newborns, remains poorly known and require further research. This study is focused on finding risk factors related to the chronic fetal hypoxia and defining a group of signs for diagnosing neonatal hypoxia. The real data of 186 pregnant women at the gestation age from 12 to 38 weeks were analyzed. A methodology for discovering interesting associations in gestation course data is proposed. Technique for association rules mining and rules selection by the neonatal hypoxia under study is discussed. The rules suggest that a strong relationship exists between the specific sets of attributes and the diagnosis. As a result, we set up a profile of the pregnant woman with a high likelihood of hypoxia of the newborn that would be beneficial to medical professionals.

Keywords Neonatal hypoxia · Gestation course data · Mining association rules · Measure of interestingness

Introduction

The main priorities of obstetrics and gynecology are women reproductive health, quality care throughout pregnancy, childbirth and postnatal period, preventing deaths of mothers and newborn babies. Despite significant progress in maternal health care in recent years, the etiology of certain diseases has not been fully explored. One of such disease is hypoxia (s.a. antepartum fetal hypoxia, neonatal hypoxia or hypoxia of the newborn). Diagnosis of fetal hypoxia and neonatal hypoxia is based on the information provided by the series of pregnancy-related examination, a thorough antenatal screening, ultrasonography observations, patient intakes, case history, and chart analysis. However, detection of hypoxia using this information remains a strong challenge especially in the early stage of the disease.

Neonatal hypoxia is a reduction in the supply of oxygen to the brain and other organs. It some cases it gives rise to neonatal hypoxic-ischemic encephalopathy (HIE), a brain dysfunction provoking epilepsy, autisms, different cognitive issues, delays in development, etc. According to [1], HIE occurs in about 20–30% of births in the developed countries and associated with significant morbidity and mortality and is an important predictor of long-term neurodevelopmental disability.

Various problems or medical complications during pregnancy in the antepartum period may cause HIE. Some of them rise from maternal states (e.g., diabetes, preeclampsia, drug and alcohol abuse), some caused by fetus states (problems with blood circulation, congenital infections of the fetus, fetal anemia, lung injury, etc.).

HIE can also affect infants during labor and delivery. In addition, infants can suffer from the effects of HIE in the postpartum period. In some cases, there are no identifiable causes for neonatal HIE [2, 3]. Moreover, HIE generally cannot be determined until the baby reaches three years of age.

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Inna Skarga-Bandurova
skarga_bandurova@ukr.net

Tetiana Biloborodova
biloborodova@snu.edu.ua

Maksym Nesterov
m.nesterov@snu.edu.ua

[1] Volodymyr Dahl East Ukrainian National University, Severodonetsk, Ukraine

Medical research aimed at identifying the causes and mechanisms of fetal hypoxia as well as early recognition the factors leading to hypoxia highly disagree on this point [2, 4–9]. Many researchers point out different groups of diagnostic signs. Thus, for example in [5], the following risk factors are defined: civil status; intrauterine release of meconium; operative delivery; breech delivery; an increased content of oxytocin in blood; pathology of the umbilical cord; application of vacuum extraction of the fetus; cardiotocography; fetal heart rate; the weight of newborn. At the same time, in [6] distinguishes other features such as the number of pregnancies (gravidities); number of births; the presence in the anamnesis of spontaneous abortions, rapid or protracted births, gynecological and somatic diseases; polyhydramnios; hypamnion; meconium in amniotic fluid; changes in the parameters of cardiotocography. According to work [9], the distribution of diagnostic signs is as follows: cardiotocography; cord entanglement; amniotic fluid staining with meconium. Various problems and medical complications cause disagreements in the opinions regarding the diagnostic markers of fetal hypoxia and require careful analysis of the data to their determination.

Due to a large variety the causes of neonatal hypoxia, the crucial task is to group specific factors and find critical combinations of parameters raise the risk of this condition. One way around this problem is to take advantage of the data mining technique and discover interesting associations in data.

The study is aimed at extracting interesting rules from gestation course data by applying an additional measure of interestingness besides having well-known minimum support and confidence, useful for early diagnosis of neonatal hypoxia. We use the association rule mining technique to find parameters combinations of the pregnancy risk factors that potentially related to newborns hypoxia.

For this paper, we will concentrate on data obtained during pregnancy in the antepartum period and related electronic health records (EHR) in first hours after delivery.

The remaining paper is organized as follows. The second section provides a brief review of related works. The methodologies, materials, and methods are described in section three. The approach utilized to obtaining and reducing association rules from the gestation course data is discussed. Experimental results with medical data sets are present in section four. Section five contains the conclusions of this paper and the future of work.

Background and related work

Association rules are widely used for medical data since their introduction in [10]. Detailed review of fundamental algorithms with current trends of association and frequent pattern mining for medical application, their advantages and disadvantages can be found in [11]. Association rule mining for the early diagnosis of Alzheimer's disease is discussed in [12]. Clinical decision support system (CDSS) for the treatment of encephalopathy using associative rules is provided in [13]. CDSS generate decisions on the basis of both clinical and physiological data applying association rule data mining algorithm and clustering algorithm. Its distinctive feature is a wide range of data on possible pathological conditions and the real-time decision support. Authors [14] explored the discovering of association rules in medical data. In the study, they presented two algorithms to association rules mining. Several medically important association rules are discussed as well as rules which are not interesting. In [15] a unified approach to quantify the association rules is proposed. Authors suggest a comprehensive definition of the interestingness of the rules. For rules with calculated importance, they used the property of anti-monotonicity. The method of associative rule mining to refine existing rules and assessing the risk of drug side effects was described in [16]. To clarifying rules the sets of organism responses to a particular medication were used. To obtained (refined) rules the risk assessment of adverse effects on the body on particular drug was made. To this, the reliability parameters, such as the lift, and the chi-square value (the significance of the association) were used.

Discovering interesting association rules in the medical domain often involves real challenges arising from their nature and the field of application. Rules and associations involve many items that hard to interpret and generate different outcomes. Worthwhile, noting that despite all the algorithms improvements, the obtained association rules can either be too obvious or contradict a priori knowledge, or contain redundant information [17]. The task of mining association rules is to generate minimal set of rules providing complete coverage of diagnostic signs with support and confidence greater or equal than some pre-specified thresholds of minimum support and minimum confidence, respectively. Another important thing here is the fact that interestingness is genuinely human construct and strongly depended from the context, which cannot be tackled automatically; hence, a human-in-the-loop [18, 19] should be taken into consideration.

Methodologies, materials, and methods

When searching for the interesting associations in gestation course data, we assume that signs of neonatal hypoxia are the frequently occurring sequences of indicators, markers, and indexes obtained during pregnancy-related examination recorded in pregnancy profile. Thus, our strategy is to find these indicators for neonatal hypoxia and reduce insignificant patterns in the results.

To achieve this goal, the process discovering interesting associations in gestation course data is divided into six steps:

1. Selection the significant diagnostics parameters of the risk factors in pregnancy, affecting to the emergence of neonatal hypoxia;

2. Transformation numerical data to nominal scale according to the medical thresholds;
3. Association rules mining and rules selection;
4. Pruning the number of candidates by the lift;
5. Calculation of rules interestingness with respect to each other;
6. Creating the profile of a pregnant woman with the risk of neonatal hypoxia.

Selection the significant diagnostics parameters

For this research, we collected the real gestation course data of 186 pregnant women at the gestation age from 12 to 38 weeks, among them 81 datasets with neonatal hypoxia and 105 datasets without this pathology. The fragment of the structure of input data is presented in the Table. 1.

In our previous study [20], we discovered the significant parameters associated with risk factors for pregnancy and affecting the emergence of neonatal hypoxia. In decreasing order of influence, they are: placental grading at 30 to 38 weeks, placenta thickness at 30 to 38 weeks, blood prothrombin index, amniotic fluid index (AFI) at 30 to 38 weeks, erythrocyte sedimentation rate (ESR) at 21 weeks, ESR at 30 weeks.

Transformation numerical data to nominal scale according to the medical thresholds

On this stage, all quantitative data were transformed to nominal scale according to normal limits and reference ranges for laboratory tests. As a result, the following gradations were obtained:

- placental grading at 30 to 38 weeks: null (0), first (I), second (II), third (III);
- placenta thickness at 30 to 38 weeks: low (0–24 mm), normal (25–45 mm), high (over 46 mm);
- prothrombin index: four (IV), five (V), six (VI);
- AFI at 30 to 38 weeks: low (0–82 mm), normal (83–268 mm), high (over 269 mm);
- ESR at 21 weeks: low (0–5 mm/s), normal (6–25 mm/s), high (over 26 mm/s);
- ESR at 30 weeks: low (0–4 mm/s), normal (5–40 mm/s), high (over 40 mm/s).

Association rule mining for gestation course data

The study was conducted with the R, a free software environment for statistical computing and graphics and R package arules [21]. For association rule mining, we use Apriori algorithm [22]. The following rule evaluation metrics were set: minimum support threshold is 0.1 and minimum confidence is 0.2.

On the first stage, we obtained 473 rules and sorted them according to the label diagnosis 'neonatal hypoxia' As a result, 29 rules remained (see Fig. 1). The complete set of source data and rules is available in [23].

Table 2 presents the best results achieved by Apriori algorithm.

In general case, association analysis algorithms generate a huge number of items and can produce up to hundreds of association rules. An association rule is an implication expression R: X → Y, where X denoted antecedent and Y denotes consequent $X \cap Y = \varnothing$. Both X and Y are considered as a set of conjuncts of the form $c_1, c_2, \ldots, c_k$. The strength of the association rule is measured in terms of its support (s), confidence [24] and interestingness [14].

To demonstrate the interpretation of the rules obtained let's analyze rule 286. Support determines how often a rule is applicable to a given data set and gives us the ratio of number pregnant women with following items occur together: 'ESR indicators at week 21 of pregnancy = increased', 'the vertical size of the amniotic fluid on ultrasound scanning during 30-38 weeks of gestation = decreased', 'the thickness of the placenta

Table 1 Fragment of the input data set

Dataset ID	years, ESR at 12 week, ESR at 21 week, ESR at 30 week, Hemoglobin at 12 week, Hemoglobin at 21 week, Hemoglobin at 30 week, Prothrombin index, Systolic pressure at 24 week, Diastolic pressure at 24 week, Systolic pressure at 28 week, Diastolic pressure at 28 week, Systolic pressure at 30 week, Diastolic pressure at 30 week, Systolic pressure at 34 week, Diastolic pressure at 34 week, Systolic pressure at 38 week, Diastolic pressure at 38 week, AFI at 20 to 24 weeks, Placental grading at 20 to 24 weeks, Placenta thickness at 20 to 24 weeks, Fetal heart rate, at 20 to 24 weeks, AFI at 30 to 38 weeks, Placental grading at 30 to 38 weeks, Placenta thickness at 30 to 38 weeks, Fetal biophysical profile (FBP) at 30 to 38 weeks, Fetal heart rate at 30 to 38 weeks, Diagnosis
001	27,,,,,,,,110,70,95,60,90,65,105,70,85,60,100,65,49,0,23,148,70,1,26,148,pathology
002	24,25,16,39,110,108,102,5100,70,90,60,90,60,90,60,,,120,80,48,0,24,146,64,3,36,9145,pathology
003	20,17,44,16,132,97,98,120,70,110,60,120,75,120,55,120,60,115,70,46,0,20,139,64,1,35,8151,pathology
004	24,12,33,43,118,102,6100,70,,,110,70,110,70,,,110,70,100,1,25,140,69,2,27,10,137,pathology
005	35,46,45,44,106,106,96,5120,70,110,70,110,70,110,70, 110,70,110,70,140,1,27,148, 70,2,34,150,pathology
…	….
086	19,16,17,15,117,101,102,108,70,128,70,120,60,115,70,115,70,125,85,49,0,23,148,110,3,36,9140,norm

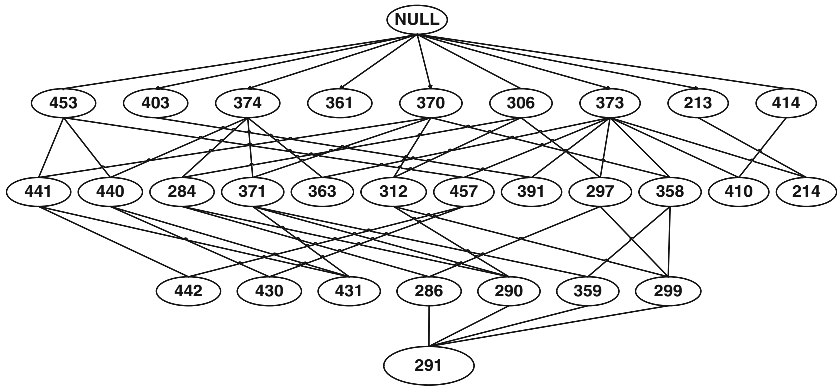


Fig. 1 The set of 29 rules-candidates for early diagnosis of neonatal hypoxia

on ultrasound scanning during 30-38 weeks of gestation = normal' with diagnosis = 'neonatal hypoxia' to the total number of pregnant women. Thus, we can see that there are 26 pregnancy women with a similar set of parameters.

The confidence determines how frequently items in set of consequences Y appear in cases containing antecedents X, this indicator shows us that the neonatal hypoxia is diagnosed in 59% of cases with the presence of increased ESR index at 21 weeks of pregnancy, a lowered vertical amniotic fluid size, and a normal placental thickness index during 30–38 weeks of pregnancy.

The lift indicates the probability of a neonatal hypoxia and is calculated by the rules that have all items from rule 286 together. This probability for rule 286 is 1.35 or 35%. More formally, lift computes the ratio between the confidence of rule and the support of the set in the rule consequent [25].

Pruning the number of rule-candidates

For pair of rule-candidates R_1 and R_2 the lift is equivalent to interest parameter and defined as follows:

$$I(R_1, R_2) = \frac{s(R_1, R_2)}{s(R_1) \cdot s(R_2)}. \quad (1)$$

The measure of interestingness in this case can be interpreted as follows:

$$I^I(R_1, R_2) \begin{cases} = 1, if\, R_1 \text{ and } R_2 \text{ are independent}, \\ > 1, if\ R_1 \text{ and } R_2 \text{ are positively correlated}, \\ < 1, if\, R_1 \text{ and } R_2 \text{are negatively correlated}. \end{cases} \quad (2)$$

Pruning the number of candidates according to the lift less than 1 decreases 29 rules-candidates to 18 for label diagnosis 'neonatal hypoxia' (see Fig. 2). The numbers of the graph nodes correspond to the numbers of the rules listed in Table. 1.

Calculation of rules interestingness

As it mentioned above, in medical applications a large number of rules are determined. To increase the information importance of rules, it is necessary to reduce their number and focus on potentially interesting ones. Further exploration of interestingness enables to discover different subjective and probabilistic measures of interestingness. When specifying the interestingness of the rule, various probabilistic measures can be used: support; confidence; Goodman and Kruskal statistics; Pyatetsky-Shapiro measure; Laplace transforms; etc. [26, 27]. In the present study, for reducing the number of rules, we applied three level technique proposed in [15] beginning with the detection of deviations in data, then testing of differences among adjusted attributes and finally, quantifying the interestingness of association rules.

1. Detecting deviations in data is performed as follows

The every conjunct c_j from association rule set is represented in the form <A = V>, where A is an item name (attribute), Dom (A) is the domain of A, and V (value) ∈ Dom (A). Degree of deviation is defined as deviation between two conjuncts $\Delta(c_i, c_j)$ and is calculated following the comparison between the items of the two conjuncts. For conjuncts c_i, c_j deviation of c_i with respect to c_j is defined as a Boolean function as follows:

$$\Delta(c_i, c_j) = \begin{cases} 0, if\, A_i = A_j \text{ and } V_i = V_j, \\ 1, if\, A_i = A_j \text{ and } V_i \neq V_j. \end{cases} \quad (3)$$

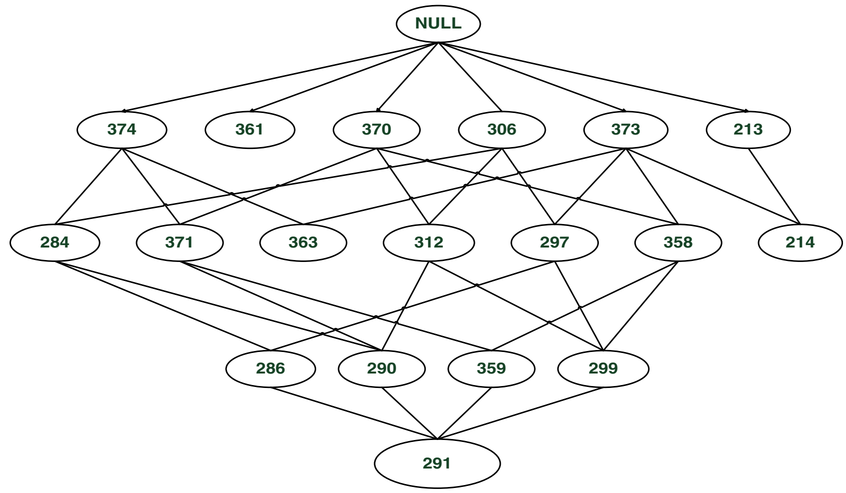
To demonstrate the interpretation of the rules obtained let's analyze rule 213 {Placental grading at 30 to 38 weeks = 1} = > {Diagnosis = neonatal hypoxia} and rule 250 {Placental grading at 30 to 38 weeks = 1} = > {AFI at 30 to 38 weeks = low}. These rules have identical attribute {Placental grading at 30 to 38 weeks} and the same values ($V_{213} = 1$ and $V_{250} = 1$), therefore the degree of deviation is equal to 0, i.e. the comparable elements are identical.

The degree of deviation at the lowest level for the rule 213 {Placental grading at 30 to 38 weeks = 1} = > {Diagnosis = neonatal hypoxia} and rule 230 {Placental grading at 30 to 38 weeks = 3} = > {Diagnosis = normal} is equal to 1. In this case we have identical attribute {Placental grading at 30 to

Table 2 Support, confidence and lift values to the assigned associated rules

Rule No.	Antecedent (Item name)	Support	Confidence	Lift
213	Placental grading at 30 to 38 weeks = 1	0.1344	0.6944	1.5946
214	Placental grading at 30 to 38 weeks = 1, Placenta thickness at 30 to 38 weeks = normal	0.1344	0.6944	1.5946
284	ESR at 21 week = high, AFI at 30 to 38 weeks = low	0.1398	0.5909	1.3569
286	ESR at 21 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.1398	0.5909	1.3569
290	ESR at 21 week = high, ESR at 30 week = high, AFI at 30 to 38 weeks = low	0.1129	0.5833	1.3395
291	ESR at 21 week = high, ESR at 30 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.1129	0.5833	1.3395
297	AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.2097	0.5735	1.3169
299	ESR at 30 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.1290	0.5714	1.3121
306	AFI at 30 to 38 weeks = low	0.2150	0.5634	1.2937
312	ESR at 30 week = high, AFI at 30 to 38 weeks = low	0.1290	0.5581	1.2816
358	ESR at 30 week = high, Placenta thickness at 30 to 38 weeks = normal	0.2097	0.4937	1.1336
359	ESR at 21 week = high, ESR at 30 week = high, Placenta thickness at 30 to 38 weeks = normal	0.1667	0.4920	1.1299
361	ESR at 30 week = normal	0.1290	0.4898	1.1247
363	ESR at 21 week = high, Placenta thickness at 30 to 38 weeks = normal	0.2097	0.4875	1.1194
370	ESR at 30 week = high	0.2527	0.4700	1.0792
371	ESR at 21 week = high, ESR at 30 week = high	0.1989	0.4683	1.0755
373	Placenta thickness at 30 to 38 weeks = normal	0.3441	0.4571	1.0497
374	ESR at 21 week = high	0.2527	0.4563	1.0478
391	ESR at 21 week = normal, Placenta thickness at 30 to 38 weeks = normal	0.1075	0.4255	0.9771
403	ESR at 21 week = normal	0.1344	0.3968	0.9112
410	Placental grading at 30 to 38 weeks = 2, Placenta thickness at 30 to 38 weeks = normal	0.1236	0.3833	0.8802
414	Placental grading at 30 to 38 weeks = 2	0.1344	0.3788	0.8698
430	ESR at 21 week = high, Prothrombin index = 5, Placenta thickness at 30 to 38 weeks = normal	0.1075	0.3509	0.8057
431	ESR at 21 week = high, ESR at 30 week = high, Prothrombin index = 5	0.1129	0.3500	0.8037
440	ESR at 21 week = high, Prothrombin index = 5	0.1344	0.3333	0.7654
441	ESR at 30 week = high, Prothrombin index = 5	0.1290	0.3333	0.7654
442	ESR at 30 week = high, Prothrombin index = 5, Placenta thickness at 30 to 38 weeks = normal	0.1021	0.3333	0.7654
453	Prothrombin index = 5	0.1667	0.3039	0.6978
457	Prothrombin index = 5, Placenta thickness at 30 to 38 weeks = normal	0.1236	0.2949	0.6771

Fig. 2 Pruned set of candidates by the lift



38 weeks} and different values ($V_{213} = 1$ and $V_{230} = 3$). This means the comparable elements are different.

2. The differences among adjusted attributes are determined using the following formula:

$$\overline{d}(R_1, R_2) = \begin{cases} 0, \quad \text{if } |R_1| = |R_2| \, \forall c_i \in R_1, \exists\, c_j \in R_2, \text{ that } \Delta(c_i, c_j) = 0, \\ 1 \qquad\qquad \forall c_i \in R_1, \neg\exists\, c_j \in R_2, \text{ that } \Delta(c_i, c_j) = 1, \\ \dfrac{\sum\limits_{c_i \in R_1, c_j \in R_2} \min \Delta(c_i, c_j)}{|R_1|}, \text{ othervise}, \end{cases} \tag{4}$$

where R_1 and R_2 are considered as two sets of conjuncts c_i and c_j.

Parameter value $\overline{d} = 0$ indicates that R_1 and R_2 are identical, $\overline{d} = 1$ indicates the maximum deviation between rule sets, and the other $\overline{d}$ values between 0 and 1 are defined as a transient deviation.

For example, for the rule 283 {ESR at 21 week = high, AFI at 30 to 38 weeks = low} => {Placental maturity degree at 30 to 38 weeks = second} and rule 284 {ESR at 21 week = high, AFI at 30 to 38 weeks = low} => {Diagnosis = neonatal hypoxia} the number of elements is equal. The elements of rule 283 and their values belong to the set of elements and values of rule 284 and have a degree of deviation at the lowest level equal to 0. Thus, the degree of deviation between the comparable sets of elements on the average level is equal to 0, and, consequently, the set of elements of the conditions are identical.

For rule 297 {AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal} => {Diagnosis = neonatal hypoxia} and rule 173 {AFI at 30 to 38 weeks = normal, Placental maturity grade at 30 to 38 weeks = second} => {Diagnosis = normal}, in the set of values of elements of the rule 297 there is no value belonging to the set of values of elements of the rule 173 with the degree of deviation at the lowest level = 1. In this case, the degree of deviation between the compared sets of elements on the average level = 1, and, therefore, the sets are different.

In case of non-compliance with the above conditions, for the rule 290 {ESR at 21 week = high, ESR at 30 week = high, AFI at 30 to 38 weeks = low} => {Diagnosis = neonatal hypoxia} and rule 150 {ESR at 21 week = high, AFI at 30 to 38 weeks = low, placental grading at 30 to 38 weeks = second} => {Prothrombin index = five}, with equal number of elements we can determine the quantitative estimation of the deviation for the conditions of the rules at the median level. The degree of deviation at the lowest level for the values of these elements is 0, 0, 1, respectively. By (4) the average deviation for these rules:

$$\frac{1}{3}(0, 0, 1) = 0.33 \cdot 0 + 0.33 \cdot 0 + 0.33 \cdot 1 = 0.33.$$

The interestingness of the rules is computed in terms of a certain deviation at the lowest and average level.

3. 3. Estimation of interestingness for association rules is performing as follows.

Let R_1: $X_1 \rightarrow Y_1$ and R_2: $X_2 \rightarrow Y_2$ be two association rules, then interestingness of a rule R_1 with respect to the rule R_2 is calculated as follows:

$$I^{II}(R_1, R_2) = \begin{cases} 0, & if\ \overline{d}(X_1, X_2) = 0\ and\ \overline{d}(Y_1, Y_2) = 0, \\ \left(\min\limits_{S \in R} \overline{d}(X_1, X_2) + \overline{d}(Y_1, Y_2)\right)/2, & if\ \overline{d}(X_1, X_2) \geq \overline{d}(Y_1, Y_2), \\ \left(\overline{d}(X_1, X_2) + \min\limits_{S \in R} \overline{d}(Y_1, Y_2)\right)/2, & if\ \overline{d}(X_1, X_2) < \overline{d}(Y_1, Y_2), \\ 1, & if\ \overline{d}(X_1, X_2) = 1\ and\ \overline{d}(Y_1, Y_2) = 1. \end{cases} \quad (5)$$

According to formula (5), $I^{II} = 0$ indicates that $R_1$and R_2 are identical, $I^{II} = 1$ denotes maximum deviation between $R_1$and R_2. Other cases indicate different deviations in the interestingness of association rules. To select interesting rules the user should specify the threshold of their interestingness.

For example, the estimation of interestingness for rule 286 {ESR at 21 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal} = > {Diagnosis = neonatal hypoxia} in relation to rule 290 {ESR at 21 week = high, ESR at 30 week = high, AFI at 30 to 38 weeks = low} = > {Diagnosis = neonatal hypoxia} can be performed as follows.

The deviation of the elements of the rule's conditions at the lowest level is (0, 0, 1). The deviation at the average level is 0.33. The deviation of the effects of the rules at the lowest level is 0. The deviation of the effects of the rules at the average level is 0. Since the deviation of the conditions at the average level is 0.33 and means more than the deviation of the effects on the average level of 0, then the interestingness of rule 286 in relation to rule 290 is $I^{II}(286, 290) = (0.33 + 0)/2 = 0.165$..

The anti-monotone property based on the threshold of the measure of interestingness can be applied to reduce the dimension of the resulting rule set. The anti-monotone property is that the measure of the interest of any set of elements should not exceed the minimal measure of interest of any of its subsets. This property greatly facilitates the mining rules.

Taking all these points together, we calculated the interestingness of the associative rules for the diagnosis of hypoxia of the newborn with a threshold of interest equal to 0.2.

Creating the profile of a pregnant woman with the risk of neonatal hypoxia

The profile of patient at risk of hypoxia is formed from the resulting table of interesting association rules with excluding all duplicate items. In general case, it consists of an item name (attribute), values of the item, and limits for normal ranges regarding medical norms and reference ranges for laboratory tests.

Findings and study outcome

Within the accuracy of observation for gestation course data of 186 pregnant women with 81 datasets with neonatal hypoxia, the following results were obtained.

The following is a detailed consideration of the reduction of candidate rules based on the quantification of the interestingness of the rules.

At the first stage, the one-element sets (candidates) were formed. They are node 374 {ESR at 21 week = high}; node 361 {ESR at 30 week = normal}; node 370 {ESR at 30 week = high}; node 306 {AFI at 30 to 38 weeks = low}; node 373 {Placenta thickness at 30 to 38 weeks = normal}; and node 213 {Placental grading at 30 to 38 weeks = 1}.

During processing, the node 361 was defined as infrequent and was excluded from the further set of the rules. The rest of the sets are considered as often encountered.

Then two-element sets (candidates) were put together. At this stage, two sets were identified as rarely encountered, they are node 363 {ESR at 21 weeks = high, Placenta thickness at 30 to 38 weeks = normal}; and node 214 {Placental grading at 30 to 38 weeks = 1, Placenta thickness at 30 to 38 weeks = normal}.

Since the rarely encountered sets were excluded, the four-element set (the candidate) was built from the three-element sets, which are often encountered. At the last stage, the node 291 {ESR at 21 weeks = high, ESR at 30 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal} is defined as the single four-element set.

The deviation in data at the first level for simple one-element rules 213, 306, 361, 370, 373, 374 is equal to 1. The deviation of the consequences is equal to 0.

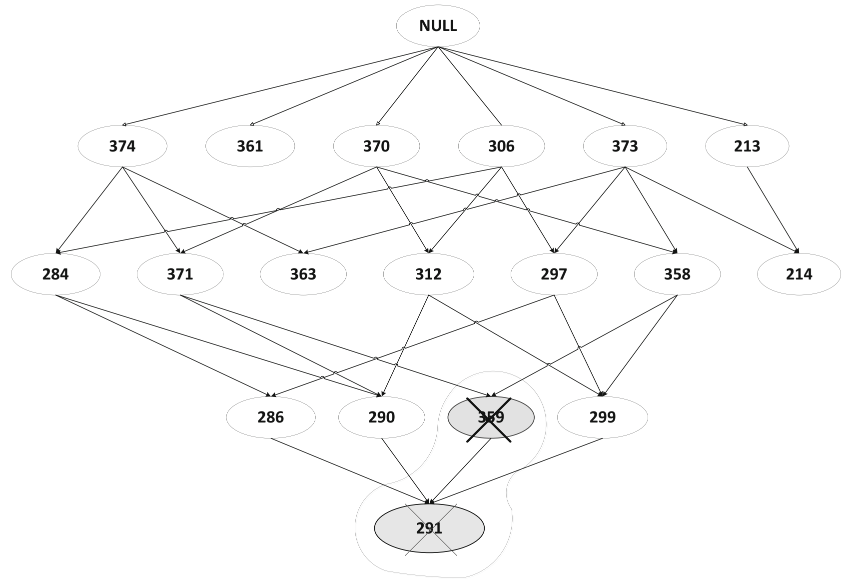


Fig. 3 A reduced number of rules in compliance with their interestingness

The difference among adjusted attributes at the second level is 0.5, and the deviation of the consequences is 0.

Calculations on the third level give us the interestingness of association rules 213, 306, 361, 370, 373, 374 is equal to 0.25,

Table 3 Total number of interesting association rules

Rule No.	Antecedent (Item name)	Interestingness
213	Placental grading at 30 to 38 weeks = 1	0.25
214	Placental grading at 30 to 38 weeks = 1, Placenta thickness at 30 to 38 weeks = normal	0.25
284	ESR at 21 week = high, AFI at 30 to 38 weeks = low	0.25
286	ESR at 21 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.33
290	ESR at 21 week = high, ESR at 30 week = high, AFI at 30 to 38 weeks = low	0.33
297	AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.25
299	ESR at 30 week = high, AFI at 30 to 38 weeks = low, Placenta thickness at 30 to 38 weeks = normal	0.33
306	AFI at 30 to 38 weeks = low	0.25
312	ESR at 30 week = high, AFI at 30 to 38 weeks = low	0.25
358	ESR at 30 week = high, Placenta thickness at 30 to 38 weeks = normal	0.25
361	ESR at 30 week = normal	0.25
363	ESR at 21 week = high, Placenta thickness at 30 to 38 weeks = normal	0.25
370	ESR at 30 week = high	0.25
371	ESR at 21 week = high, ESR at 30 week = high	0.25
373	Placenta thickness at 30 to 38 weeks = normal	0.25
374	ESR at 21 week = high	0.25

Table 4 A profile of the pregnant woman with a threat of hypoxia in a newborn

Item name (attribute)	Item values	Normal limits
ESR at 21 weeks of pregnancy (mm/s)	26–69	25
ESR at 30 weeks of pregnancy (mm/s)	4–69	30–35
Placental grading at 30 to 38 weeks (grade)	I	I-II
Placenta thickness at 30 to 38 weeks (mm)	25–45	25–45
AFI at 30 to 38 weeks (mm)	269–306	82–268

i.e., above the threshold of interest and therefore these rules are considered as interesting.

The interestingness of the rules 214, 284, 297, 312, 358, 363, 371 for two-element rule sets is 0.25. Three-element rules 286, 290, 299 have an interest of 0.33, and three-element rule 359 has an interest of 0.165, i.e., less than the assigned threshold of interestingness. Thus, all rules with interestingness less than the minimum assigned value 0.2 are considered as not interesting, and conformably to the property of anti-monotone, all subsets of these sets are not interesting too and can be excluded from further consideration. In Fig. 3 a grey background, highlights excluded rules.

As a result, the total number of association rules was reduced to 16 shown in Table 3.

These rules give us the profile of pregnant women with a high likelihood of hypoxia of the newborn, shown in Table 4.

The combination of the objective measures with the subjective measures of interestingness makes it possible to improve the results of interestingness and to discover associations among attributes that characterize the patterns of the norm and neonatal hypoxia as well as to use of them for the early diagnosis of this disease. These findings warrant further application and development of data mining techniques for medical informatics. In spite of positive trends observed, a number of topics need to be adjusted and special consideration such as sample size, algorithms for selecting the significant parameters related to the risk factors in pregnancy, accounting for missing values in datasets and many others should be taken into account.

Conclusions and future work

The research was targeted at improving association rule mining technique in the medical domain. In this paper, we covered the slow but steady progress toward defining a group of signs of neonatal hypoxia. We have proposed a methodology for sequential estimation of interestingness of association rules and applied two sets of criteria for evaluating the quality of association rules. The first set of criteria consists of objective interestingness measure based on the statistical indicators of support, confidence, and lift. The second set of criteria involves subjective measures of interestingness based on quantifying the interestingness of association rules. A three-level procedure for reducing the number of association rules and sifting through the candidates to identify the most interesting rules has been performed.

The hybrid framework to association rule mining has been tested on real gestation course dataset. Based on the results, we set up a profile of pregnant women with a high likelihood of hypoxia in the newborn and obtained encouraging results. From the obtained profile, the following combination of parameters gives a high risk of neonatal hypoxia, they are increased ESR at 21 weeks of pregnancy, low or normal ESR at 30 weeks of pregnancy, the low placental grading at 30 to 38 weeks of pregnancy, normal placenta thickness at 30–38 weeks of pregnancy, increased AFI at 30–38 weeks of pregnancy.

Two points are worth noting about the effects of using this approach. There is still a possibility of inaccuracy in results, due to narrow set and small sample size. The research result requires a medical expert assessment for determining the practical value of the rules obtained. Nevertheless, results bring us to the conclusion that this approach offers hope to improve our knowledge about latent signs of neonatal hypoxia.

Compliance with Ethical Standards

Conflict of interest None of the authors has any conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Kurinczuk, J. J., White-Koning, M., and Badawi, N., Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy. Early Hum. Dev. 86:329–338, 2010.
2. Martinez-Biarge, M., Diez-Sebastian, J., Wusthoff, C. J., Mercuri, E., and Cowan, F. M., Antepartum and intrapartum factors preceding neonatal hypoxic-ischemic encephalopathy. Pediatrics 132(4): e952–e959, 2013.
3. Allen, K. A., and Brandon, D. H., Hypoxic ischemic encephalopathy: pathophysiology and experimental treatments. Newborn Infant Nurs Rev 11(3):125–133, 2011.
4. Seikku, L., Rahkonen, L., and Tikkanen, M., Hämäläinen, E., Rahkonen, P., Andersson, S., Teramo, K., Paavonen, J., Stefanovic, V.: Amniotic fluid erythropoietin and neonatal outcome in pregnancies complicated by intrauterine growth restriction before 34 gestational weeks. Acta Obstet. Gynecol. Scand. 94(3): 288–294, 2015.
5. Milsom, I., Ladfors, L., Thiringer, K., Niklasson, A., Odeback, A., and Thornberg, E., Influence of maternal, obstetric and fetal risk factors on the prevalence of birth asphyxia at term in a Swedish urban population. Acta Obstet. Gynecol. Scand. 81(10):909–917, 2002.
6. Mertsalova, O.V., Perinatal hypoxic injuries of fetal CNS in high-risk pregnant women (diagnosis, prognosis of outcomes, optimization of pregnancy and delivery management), 2002.

7. Antonucci, R., Porcella, A., and Pilloni, M. D., Perinatal asphyxia in the term newborn. Journal of Pediatric and Neonatal Individualized Medicine 3(2):e030269, 2014. https://doi.org/10.7363/030269
8. Hypoxic-Ischaemic Encephalopathy (HIE). Queensland Clinical Guidelines, 2016.
9. Palsdottir, K., Dagbjartsson, A., Thorkelsson, T., and Hardardottir, H., Birth asphyxia and hypoxic ischemic encephalopathy, incidence and obstetric risk factors. Laeknabladid 93(9):595–601, 2007.
10. Agrawal, R., Imielinski, T., and Swami, A., Mining association rules between sets of items in large databases. ACM SIGMOD Conference 207-216, 1993.
11. Kumar, K. P., and Arumugaperumal, S., Association Rule Mining and Medical Application: A Detailed Survey. Int. J. Comput. Appl. 80(17):0975–8887, 2013.
12. Chaves, R., Gorriz, J. M., Ramirez, J., Illan, I. A., Salas-Gonzalez, D., and Gomez-Rio, M., Efficient mining of association rules for the early diagnosis of Alzheimer's disease. Phys. Med. Biol. 56: 6047–6063, 2011.
13. Cheng, Chih-Wen, Chanani, N., Enugopalan, J., Maher, K., Wang, M. D., icuARM - An ICU Clinical Decision Support System Using Association Rule Mining. IEEE Journal of Translational Engineering in Health and Medicine, 1, 2013.
14. Ordonez, C., Santana, C. A., Levien de Braal: Discovering Interesting Association Rules in Medical Data. ACM DMKD workshop, 78–85, 2000.
15. Kaur, H., Wasan, S. K., Al-Hegami, A. S. and Bhatnagar, V., A Unified Approach for Discovery of Interesting Association Rules in Medical Database. In Advances in Data Mining. Lecture Notes in Artificial Intelligence. - SpringerVerlag, Berlin, Heidelberg, 4065, 2006.
16. Reps, J. M., Aickelin, U., Ma, J., Zhang, Y. Refining Adverse Drug Reactions using Association Rule Mining for Electronic Healthcare Data. IEEE International Conference on Data Mining Workshop, 763–770, 2014.
17. Simovici, D.A. Data Mining of Medical Data: Opportunities and Challenges in Mining Association Rules. Retrieved from www.cs.umb.edu/~dsim/papersps/dmmd.pdf, 2012.
18. Holzinger, A., Interactive Machine Learning for Health Informatics: When do we need the human-in-the-loop? Brain Informatics 3(2):119–131, 2016.
19. Yildirim, P., Ekmekci, I., and Holzinger, A., On Knowledge Discovery in Open Medical Data on the Example of the FDA Drug Adverse Event Reporting System for Alendronate (Fosamax). In: Holzinger, A., Pasi, G. (Eds), Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data, Lecture Notes in Computer Science, LNCS 7947. Berlin: Springer, 2013, 195–206.
20. Skarga-Bandurova, I., Biloborodova, T., Exploratory Data Analysis to Identifying Meaningful Factors of Hypoxic Fetal Injuries. Herald of the NTU "KhPI". Series: Informatics and Modelling. – Kharkov: NTU "KhPI", 44 (1216), 122–135, 2016.
21. Hahsler, M., Grün, B., Hornik, K., Buchta, C., Introduction to arules - A computational environment for mining association rules and frequent item sets, 2010. http://cran.r-project.org/web/packages/arules/vignettes/arules.pdf. Accessed 23 Nov 2018.
22. Hegland, M., The Apriori Algorithm – a Tutorial. WSPC/Lecture Notes Series, 9(7), 2005.
23. Biloborodova, T.O., Models and methods of information technology for automated systems of medical data analysis, PhD diss., V. Dahl EUNU, 2018. https://snu.edu.ua/wp-content/uploads/2018/08/Biloborodova.Dysertatsiya.pdf
24. Doddi, S., Marathe, A., Ravi, S.S., Torney, D.C.: Discovery of Association Rules in Medical Data http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.4047&rep=rep1&type=pdf
25. Zhao, Y., Zhang, C., Cao, L., Post-mining of association rules: techniques for effective knowledge extraction. In Information Science Reference, 2009.
26. Masood, A., Ouaguenouni, S.: Probabilistic Measures for Interestingness of Deviations – A Survey. International Journal of Artificial Intelligence & Applications (IJAIA), 4, 2013.
27. Billing, V. A., Association rule mining for medical diagnostics. Software Solutions and Systems 2:146–157, 2016.