

І.С. СКАРГА-БАНДУРОВА, д-р техн. наук, доц., проф., СНУ
ім. В. Даля, Северодонецьк,

Т.О. БІЛОБОРОДОВА, асп., СНУ ім. В. Даля, Северодонецьк

ПОШУКОВИЙ АНАЛІЗ ДАНИХ ДЛЯ ВИЗНАЧЕННЯ РЕЛЕВАНТНИХ ФАКТОРІВ ГІПОКСИЧНОГО УРАЖЕННЯ ПЛОДА

Обґрунтовано актуальність пошукового аналізу даних для вивчення властивостей досліджуваних даних перебігу вагітності. Проведено обробку даних з використанням статистичних, графічних, формальних параметричних та непараметричних методів дослідження. За результатами кластерного аналізу виділено групу факторів, з обґрунтованим значущим впливом на розвиток патології плоду. Підтверджено наявність взаємозв'язку між окремими виділеними факторами перебігу вагітності та наявністю досліджуваної патології у новонародженого, для досліджуваного набору даних отримано нові угруповання факторів перебігу вагітності. Іл.: 2. Табл.: 3. Бібліогр.: 16 назв.

Ключові слова: пошуковий аналіз даних, кластерний аналіз, патологія у новонародженого, угруповання факторів.

Опис проблеми. Дослідження останніх десятиліть і технології аналізу даних відкрили нові ідеї і розширили розуміння причин патологій розвитку плода. Разом з тим, ряд захворювань, серед яких гіпоксичні ураження плода, залишаються в центрі уваги медиків і аналітиків. Внутрішньоутробна гіпоксія пов'язана з різними показниками здоров'я матері, плаценти і плода, які можуть проявлятися по-різному і мати різні наслідки. У цьому сенсі своєчасне виявлення показників перебігу вагітності, здатних прогнозувати розвиток гіпоксії плода, є актуальною задачею. Відправною точкою в цих дослідженнях є пошуковий аналіз, який дозволяє проводити відбір діагностичних ознак і попередню інтерпретацію даних.

Об'єктом даного дослідження є ретроспективні набори даних, отримані в процесі спостереження за жінками, що перебували на 12 – 38 тижнях вагітності.

Аналіз літературних даних. Пошуковий аналіз даних є необхідним етапом вивчення властивостей обраних факторів, визначення дійсно потенційно корисних і важливих кореляцій і закономірностей в наявних даних. Однією з головних цілей пошукового аналізу даних є зменшення розмірності вихідних даних для досягнення їх оптимальної компактності за умови мінімальної втрати вихідної інформації. Результатом такого аналізу є перехід від безлічі вихідних змінних до істотно меншого числа нових змінних [1]. Так, автори [2] представили в своїй роботі методику

кількісного аналізу медичних даних, яка допоможе на початковому етапі дослідження виявити наявну прогностичну інформацію в досліджуваних даних. Першим кроком в інтелектуальному аналізі даних автори пропонують графічний метод визначення нормальності даних за допомогою гістограм, та, на підставі отриманих результатів, подальше визначення статистичної значущості показників за допомогою коефіцієнта кореляції Спірмана. В роботі також надано методiku зменшення кількості пояснюючих змінних, що включає застосування статистичних тестів, алгоритмів класифікації, використання крос-валідації. В роботі [3] надано приклад застосування пошукового аналізу до даних нейровізуалізації. Автори доводять перевагу використання графічного методу аналізу даних за допомогою функціональної вічкової діаграми для аналізу періодограмм спектральної області даних часових рядів електро-енцефалограми, в порівнянні з використанням класичної діаграми. Аналіз материнської смертності з застосуванням статистичного аналізу даних наведено в роботі [4], де безперервні змінні досліджені за допомогою критерію U Манна-Уїтні та представлені у вигляді медіани і діапазону. Категоріальні змінні автори порівнювали за допомогою точного критерію Фішера і представили у вигляді частот. Аналіз робіт в області методології пошукового аналізу даних в медичних дослідженнях показав потенційну можливість використання даного підходу для пошуку релевантних факторів гіпоксичних уражень плода, разом з тим, спеціалізовані медичні дослідження, зокрема [5 – 7] вказують на те, що визначення наявності або відсутності гіпоксії з набору даних у ряді випадків є проблематичним (немає жодної змінної, що чітко ідентифікує гіпоксію). Таким чином, для наявного набору даних поточне завдання полягає в пошуку і визначенні значущих діагностичних ознак і їх ранжування.

Мета і задачі дослідження. Метою дослідження є виділення релевантних факторів перебігу вагітності для розвитку гіпоксії у новонародженого. Використовуючи методи аналізу даних необхідно виконати пошук і виділити релевантні фактори та з'ясувати, які набори даних і в якій послідовності можуть служити показниками і (або) предикторами розвитку досліджуваної патології.

Виконання завдань реалізовано у вигляді наступних етапів: проведено аналіз відсутніх значень; визначено основні міри мінливості досліджуваних факторів; визначено розподіл частот ймовірностей досліджуваних даних за допомогою графічного методу та з застосуванням формального тесту Колмогорова-Смірнова; визначення статистичної значимості досліджуваних показників перебігу вагітності; кластерний аналіз показників з побудовою дендограмм для виділення груп показників перебігу вагітності; визначення значущості виділених

груп показників перебігу вагітності за допомогою критерію Краскала-Уолліса.

Опис досліджуваних даних. За результатами вивчення медичної літератури про розвиток гіпоксії у новонародженого, виділено 29 показників які, за даними дослідників [5 – 7], мають високу ймовірність впливу на розвиток даної патології у новонародженого: вік, показники швидкості осідання еритроцитів (ШОЕ) та рівня гемоглобіну на 12, 21, 30 тижнях вагітності, протромбіновий індекс, систолічний та діастолічний тиск на 24, 28, 30, 34, 36, 38 тижнях вагітності, вертикальний розмір амніотичної рідини, ступінь зрілості і товщина плаценти, частота серцебиття плода на 20 – 24 і 30 – 38 тижнях вагітності, біофізичний профіль плода на 30 – 38 тижнях вагітності, діагноз новонародженого. Всі показники є кількісними, крім протромбінового індексу, ступеня зрілості плаценти на 20 – 24, 30 – 38 тижнях вагітності і біофізичного профілю плода, які є порядковими. Методи розвідувального аналізу застосовувалися для даних перебігу вагітності та стану новонароджених у 186 жінок. Обсяг порівнюваних наборів даних приблизно однаковий: 81 випадок новонароджених з патологією і 105 випадків новонароджених без патології.

Аналіз відсутніх значень. Відсутні значення в даних присутні в більшості статистичних даних [8]. Потенційна небезпека таких даних складається у тому, що вони можуть зменшити ефективність дослідження та привести до необ'єктивних результатів. Для того, щоб вирішити, як вчинити з відсутніми даними, корисно знати, причини їх відсутні. У досліджуваному випадку це пов'язано з відсутністю деяких протоколів лікування, відсутністю законодавчої документації, яка встановлює обов'язкові для заповнення дані та відсутність міри покарання за невиконання цих вимог, недбалість медперсоналу при заповненні документації, відсутність у вагітної жінки коштів для проходження необхідних, призначених лікарем досліджень тощо. Відповідно до типу відсутніх даних [9, 10], значення відсутні випадково. Кількість і процентне співвідношення відсутніх значень в даних за показниками представлені в табл. 1.

Аналіз кількості та процентного співвідношення відсутніх значень в даних за показниками, показав, що в більшості показників їх кількість не є критичною. Найбільша кількість відсутніх значень в даних за показником біофізичного профілю плода на УЗД 30 – 38 тижнів вагітності, настільки велика, що на думку авторів, проаналізувавши результати проведеного аналізу, в подальшому його можливо доведеться виключити. Також високий відсоток відсутніх даних у показників вертикального розміру амніотичної рідини на УЗД 30 – 38 тижнів і

протромбінового індексу. Грунтуючись на співвідношенні невеликого розміру досліджуваного набору даних до кількості відсутніх даних, було прийнято рішення для більшості використаних методів аналізу, описаних нижче, використовувати ігнорування таких відсутніх значень окремо, не ігноруючи при цьому спостереження повністю.

Таблиця 1

Кількісні показники і процентне співвідношення відсутніх значень в даних

Фактор перебігу вагітності (тижні вагітності* – т.в.)	Дані			
	Включені		Відсутні	
	Кількість	Процент	Кількість	Процент
Вік вагітної жінки	186	100,0%	0	0,0%
ШОЕ на 12 (т.в.)	157	84,4%	29	15,6%
ШОЕ на 21 т.в.	171	91,9%	15	8,1%
ШОЕ на 30 т.в.	150	80,6%	36	19,4%
Гемоглобін на т.в.	162	87,1%	24	12,9%
Гемоглобін на 21 т.в.	172	92,5%	14	7,5%
Гемоглобін на 30 т.в.	155	83,3%	31	16,7%
Протромбіновий індекс	128	68,8%	58	31,2%
Систоличний тиск на 24 т.в.	173	93,0%	13	7,0%
Діастолічний тиск на 24 т.в.	173	93,0%	13	7,0%
Систоличний тиск на 28 т.в.	171	91,9%	15	8,1%
Діастолічний тиск на 28 т.в.	171	91,9%	15	8,1%
Систоличний тиск на 30 т.в.	177	95,2%	9	4,8%
Діастолічний тиск на 30 т.в.	177	95,2%	9	4,8%
Систоличний тиск на 34 т.в.	168	90,3%	18	9,7%
Діастолічний тиск на 34 т.в.	168	90,3%	18	9,7%
Систоличний тиск на 36 т.в.	161	86,6%	25	13,4%
Діастолічний тиск на 36 т.в.	161	86,6%	25	13,4%
Систоличний тиск на 38 т.в.	154	82,8%	32	17,2%
Діастолічний тиск на 38 т.в.	154	82,8%	32	17,2%
Вертикальний розмір амніотичної рідини на 20-24 т.в.	139	74,7%	47	25,3%
Зрілість плаценти на 20-24 т.в.	168	90,3%	18	9,7%
Товщина плаценти на 20-24 т.в.	167	89,8%	19	10,2%
Серцебиття плоду на 20-24 т.в.	165	88,7%	21	11,3%
Вертикальний розмір амніотичної рідини на 30-38 т.в.	116	62,4%	70	37,6%
Зрілість плаценти на 30-38 т.в.	146	78,5%	40	21,5%
Товщина плаценти на 30-38 т.в.	146	78,5%	40	21,5%
Біофізичний профіль плоду на 30-38 т.в.	68	36,6%	118	63,4%
Серцебиття плоду на 30-38 т.в.	146	78,5%	40	21,5%

Оцінка мінливості досліджуваних факторів. На другому етапі дослідження проведено аналіз мір мінливості кількісних факторів

перебігу вагітності для норми і патології та визначення наступних показників досліджуваних даних: середнє значення [11, 12], стандартне відхилення, дисперсія [13]. Результати розрахунків міри мінливості даних для показників перебігу вагітності з наявністю і відсутністю гіпоксії у новонародженого представлені в табл. 2.

Таблиця 2

Основні числові статистичні показники факторів перебігу вагітності з наявністю і відсутністю гіпоксії у новонародженого

Фактор перебігу вагітності (тижні вагітності* – т.в.)	Новонароджений з гіпоксією			Новонароджений без гіпоксії		
	Середнє значення	Стандартне відхилення	Дисперсія	Середнє значення	Стандартне відхилення	Дисперсія
Вік вагітної жінки	27.30	4.72	22.31	27.18	5.65	31.99
ШОЕ на 12 т.в.	22.14	12.18	148.49	23.34	12	144.2
ШОЕ на 21 т.в.	30.78	14.76	218.03	29.01	13.96	194.96
ШОЕ на 30 т.в.	31.7	13.86	192.09	33.65	14.7	216.12
Гемоглобін на 12 т.в.	113.51	13.38	179.01	115.26	16.92	286.48
Гемоглобін на 21 т.в.	110.86	11.9	141.69	109.18	12.13	147.3
Гемоглобін на 30 т.в.	109.18	12.13	147.3	103.13	11.76	138.28
Сістолічний тиск на 24 т.в.	106.86	11.49	132	104.45	10.56	111.51
Діастолічний тиск на 24 т.в.	66.8	8.44	71.36	64.75	9.8	96.1
Сістолічний тиск на 28 т.в.	105.87	10.74	115.4	105.5	11.01	121.36
Діастолічний тиск на 28 т.в.	66.81	9.07	82.33	65.69	8.76	76.75
Сістолічний тиск на 30 т.в.	107.39	11.6	134.59	105	10.18	125
Діастолічний тиск на 30 т.в.	66.99	7.94	63.1	65.39	7.43	55.24
Сістолічний тиск на 34 т.в.	110.29	10.96	120.2	108.72	11.26	126.96
Діастолічний тиск на 34 т.в.	70.21	8.57	73.5	67.96	8.58	73.62
Сістолічний тиск на 36 т.в.	112.8	15.48	239.75	108	17.7	97.79
Діастолічний тиск на 36 т.в.	72.28	12.26	150.42	69.69	9.88	97.79
Сістолічний тиск на 38 т.в.	113.03	11.56	133.75	110.57	13.33	177.83
Діастолічний тиск на 38 т.в.	71.24	9.68	93.72	70.45	9.84	96.91
Вертикальний розмір амніотичної рідини на 20-24 т.в.	80.71	43.79	1918.17	82.46	42.25	1785.48
Товщина плаценти на 20-24 т.в.	24.5	5.69	32.45	23.92	4.72	22.35
Серцебиття плоду на 20-24 т.в.	146.71	7.69	59.26	145.59	5.95	35.44
Вертикальний розмір амніотичної рідини на 30-38 т.в.	87.6	58.25	3393.85	104.17	53.26	2837.56
Товщина плаценти на 30-38 т.в.	34.07	3.87	15.01	36.92	7.42	55.19
Серцебиття плоду на 30-38 т.в.	141.39	8.81	77.62	143.26	7.02	49.38

Вивчивши міри мінливості досліджуваних даних, зроблено висновок про відмінність середнього значення для показника вертикального розміру амніотичної рідини на УЗД 30 – 38 тижнях вагітності і про відсутність принципових відмінностей в інших показниках, які б могли вказати на фактори, що мають вагомий вплив на виникнення патології.

Аналіз розподілу частот ймовірностей. На наступному етапі дослідження було проведено визначення розподілу частот ймовірностей досліджуваних даних. Цей етап необхідний для визначення методів подальшого дослідження даних, тому що для нормально розподілених даних математичною статистикою вироблені ефективні методи дослідження, які, однак, не придатні для даних які не відповідають нормальному розподілу.

За допомогою стандартного відхилення визначено значення факторів, що мають більший вплив на ймовірність розвитку гіпоксії у новонародженого. Так, наприклад, стандартне відхилення показника перебігу вагітності "гемоглобін" на 12 тижні вагітності для новонародженого з діагнозом "гіпоксія" дорівнює 13.38.

При відсутності у новонародженого гіпоксії, ця міра дорівнює 16.92. Виходячи з цього, при оцінці ймовірності виникнення гіпоксії новонародженого, можна зробити висновок, що при значеннях показника перебігу вагітності "гемоглобін" на 12 тижні вагітності від 100.13 до 126.89 ймовірність виникнення гіпоксії у новонародженого вища. Найбільше стандартне відхилення зафіксовано для показників вертикального розміру амніотичної рідини та товщини плаценти на 30 – 38 тижня вагітності, що свідчить про високу варіативність показників. Дисперсія цих показників також вказує на їх високу мінливість.

В результаті, виділено один фактор, який, з певною часткою ймовірності, може впливати на виникнення патології – вертикальний розмір амніотичної рідини на УЗД 30 – 38 тижнях вагітності.

Для визначення розподілу частот ймовірностей даних і перевірки на відповідність даних нормальному розподілу використано графічний спосіб – гістограми [14, 15]. Отримані гістограми для кількісних показників перебігу вагітності представлені на рис. 1.

Отримані гістограми розподілу частот кількісних показників зіставлені з кривою нормального розподілу. Розподіл даних в більшій чи меншій мірі відхиляється від кривої нормального розподілу.

При зіставленні гістограм можна говорити про те, що більша частина факторів не відповідає нормальному розподілу. До факторів, розподіл яких більшою мірою відповідає нормальному розподілу можна віднести: вік, рівень гемоглобіну на 12 і 21 тижнях вагітності, товщина

плаценти і частота серцебиття плода на УЗД 20 – 24 тижнях вагітності, товщина плаценти на УЗД 30 – 38 тижнях вагітності.

Скориставшись критерієм Колмогорова-Смірнова проведено уточнення отриманих при використанні графічного методу результатів. Для показників перебігу вагітності ШОЕ на 21 тиждні вагітності, ШОЕ на 30 тиждні вагітності, гемоглобін на 21 тиждні вагітності, гемоглобін на 30 тиждні вагітності значення статистичної помилки більше, ніж рівень $p = 0,05$, отже, за цими атрибутам розподіл істотно не відрізняється від нормального.

Для решти показників статистична помилка менше, ніж рівень $p = 0,05$, отже, розподіл істотно відрізняється від нормального розподілу.

На підставі отриманих результатів, можна зробити висновок про те, що параметричні методи аналізу доцільно використовувати для таких показників перебігу вагітності: ШОЕ на 21 тиждні вагітності, ШОЕ на 30 тиждні вагітності, гемоглобін на 21 тиждні вагітності, гемоглобін на 30 тиждні вагітності. Для показників, що не відповідають нормальному розподілу частот ймовірностей, слід застосовувати непараметричні методи.

Статистична значущість показників. Подальше дослідження полягає у визначенні статистичної значущості виділених факторів перебігу вагітності для діагнозу "гіпоксія" у новонародженого. Для факторів, відповідних нормальному розподілу: ШОЕ на 21 тиждні вагітності, ШОЕ на 30 тиждні вагітності, гемоглобін на 21 тиждні вагітності, гемоглобін на 30 тиждні вагітності, було застосовано однофакторний дисперсійний аналіз.

Встановивши коректність і обґрунтованість використання однофакторного ANOVA за показником р-рівнем критерію Лівіня, який для ШОЕ на 21 тиждні вагітності, ШОЕ на 30 тиждні вагітності, гемоглобін на 21 тиждні вагітності, гемоглобін на 30 тиждні вагітності становить 0.482, 0.397, 1.106, 0.138 відповідно, що є більше 0,05. Р-рівень критерію Лівіня говорить про те, що дисперсії порівнюваних показників однорідні, тобто між ними немає якісних відмінностей. На цій підставі можна зробити висновок про те, що результати однофакторного ANOVA можуть бути використані для порівняння цих показників.

Встановлена значимість впливу показників на рівень розвитку змінної "діагноз новонародженого" для порівнюваних показників ШОЕ на 21 тиждні вагітності, – 0.644, ШОЕ на 30 тиждні вагітності – 0.688, гемоглобін на 21 тиждні вагітності – 3.016, гемоглобін на 30 тиждні вагітності – 9.899. З чого зроблено висновок про незначний вплив показників гемоглобін на 21 тиждні вагітності, гемоглобін на 30 тиждні вагітності на діагноз гіпоксія у новонародженого і про дуже незначний вплив показників ШОЕ на 21 тиждні вагітності, ШОЕ на 30 тиждні вагітності. Для визначення статистичної значущості показників які не

відповідають нормальному розподілу використано непараметричний критерій U Манна-Уїтні. Спираючись на нульову та альтернативну гіпотези, нульова гіпотеза для показників протромбіновий індекс, вертикальний розмір амніотичної рідини, ступень зрілості плаценти та її товщина при проведенні УЗД на 30 – 38 тижнях вагітності відхиляється й приймається альтернативна гіпотеза, яка свідчить про зв'язок діагнозу новонародженого та цих показників. Таким чином, виділені значущі фактори для подальшого дослідження. Рівень значущості для показників: протромбіновий індекс – 0.006, вертикальний розмір амніотичної рідини на 30 – 38 тижня вагітності – 0.020, ступінь зрілості плаценти на 30 – 38 тижня вагітності – 0.001, товщина плаценти на 30 – 38 тижня вагітності – 0.002.

Таким чином, на даному етапі дослідження, виділені фактори, які безпосередньо впливають на стан новонародженого в контексті досліджуваної патології. В порядку зменшення впливу на діагноз новонародженого, це: ступінь зрілості плаценти при УЗД на 30 – 38 тижнях вагітності, товщина плаценти при УЗД на 30 – 38 тижнях вагітності, протромбіновий індекс крові, вертикальний розмір амніотичної рідини при УЗД на 30 – 38 тижнях вагітності, ШОЕ крові на 21 тижні вагітності, ШОЕ крові на 30 тижні вагітності.

Кластерний аналіз. Подальші кроки дослідження спрямовані на виявлення угруповань показників перебігу вагітності, що вірогідно впливають на ймовірність розвитку гіпоксії у новонародженого. Для визначення груп показників була застосована ієрархічна кластеризація з визначенням відстаней за допомогою квадрата відстані Евкліда за допомогою методу найдальшого сусіда. Цей метод дозволив визначити прийнятну кількість кластерів для досліджуваного набору даних і згрупувати значимо важливі параметри перебігу вагітності для розвитку патології у новонародженого. Проаналізувавши кроки агломерації визначено раціональне для досліджуваного набору даних число кластерів. Стрибокподібне збільшення значень коефіцієнта агломерації відбулося на 21 стадії роботи алгоритму, оптимальне для досліджуваного набору даних число кластерів – 7. По результатам кластерного аналізу отримано наступне об'єднання показників:

Кластер 1: ступінь зрілості плаценти при УЗД 20 – 24 і 30 – 38 тижнях вагітності, протромбіновий індекс, біофізичний профіль плода.

Кластер 2: ШОЕ на 12, 21, 30 тижнях вагітності, товщина плаценти при УЗД 30 – 38 тижнях вагітності.

Кластер 3: частота серцебиття плода при проведенні УЗД на 20-24 і 30-38 тижнях вагітності.

Кластер 4: систолічний тиск на 24, 28, 30, 34, 36 і 38 тижнях вагітності, рівень гемоглобіну в крові на 12, 21, 30 тижнях вагітності.

Кластер 5: діастолічний тиск на 24, 28, 30, 34, 36, 38 тижнях вагітності.

Кластер 6: вертикальний розмір амніотичної рідини на УЗД 20 – 24 тижнів вагітності.

Кластер 7: вертикальний розмір амніотичної рідини на УЗД 30 – 38 тижнів вагітності.

На рис. 1, наведено дендограму кластеризації показників перебігу вагітності на основі наявності або відсутності досліджуваної патології у новонародженого.

Для визначення значущості впливу отриманих угруповань чинників, проведено порівняння ступеня значимості відмінностей діагнозу новонародженого в отриманих семи угрупованнях. Зважаючи на об'єм досліджуваного набору даних, відповідно до їх розподілу, обрано критерій Краскала-Уолліса.

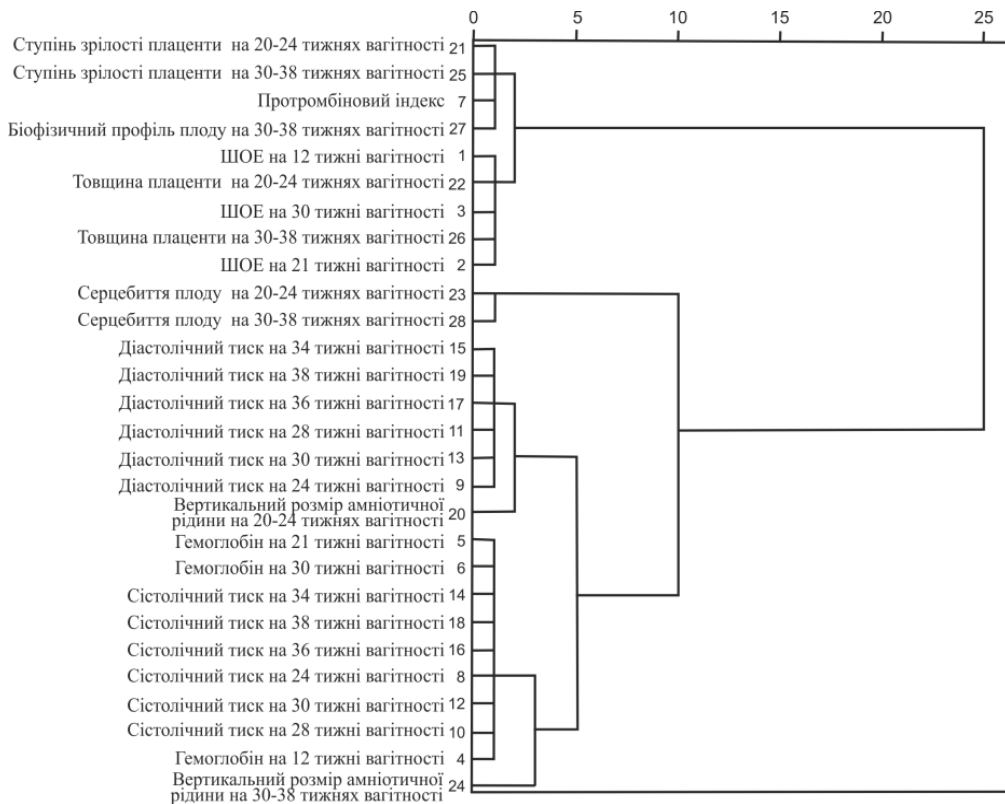


Рис. 1. Дендограма кластеризації показників перебігу вагітності

Результати порівняння ступеней значущості відмінностей діагнозу новонародженого в отриманих семи угрупованнях із застосуванням критерію Краскала-Уолліса показали, що, як це видно з табл. 3,

найбільший вплив на діагноз новонародженого мають угруповання факторів перебігу вагітності 6 і 7. Але р-рівень рівняється 0.691, тобто більше 0.05, вказує на відсутність значущих відмінностей між групами.

Таким чином, в результаті проведеного дослідження виділено угруповання 6 факторів, які безпосередньо впливають на діагноз новонародженого в контексті досліджуваної патології. В порядку убутання впливу це: ступінь зрілості плаценти при УЗД на 30 – 38 тижнях вагітності, товщина плаценти при УЗД на 30 – 38 тижнях вагітності, протромбіновий індекс крові, вертикальний розмір амніотичної рідини при УЗД на 30 – 38 тижнях вагітності, ШОЕ крові на 21 тижні вагітності, ШОЕ крові на 30 тижні вагітності.

Таблиця 3

Отриманий ранг при застосуванні критерію Краскала-Уолліса

Кластери	Середній ранг
1	93,19
2	99,13
3	72,00
4	78,20
5	87,50
6	134,00
7	134,00

Висновки. На підставі аналізу медичної літератури, виділені показники перебігу вагітності, які потенційно впливають на виникнення гіпоксії у новонародженого. Сформовано набір даних показників перебігу вагітності жінок для випадків виникнення патології та її відсутності. За результатами оцінки статистичних характеристик даних зроблено висновок про відмінність показника вертикального розміру амніотичної рідини на УЗД 30-38 тижнях вагітності для спостережень з патологією і без. Вперше виділено групу факторів, з обґрунтованим значущим впливом на розвиток патології у новонародженого. В результаті роботи підтверджено наявність взаємозв'язку між окремими виділеними факторами перебігу вагітності та наявністю досліджуваної патології у новонародженого, отримано нові угруповання факторів перебігу вагітності. Для досліджуваної патології це ступінь зрілості плаценти при УЗД на 30 – 38 тижнях вагітності, товщина плаценти при УЗД на 30 – 38 тижнях вагітності, протромбіновий індекс крові,

вертикальний розмір амніотичної рідини при УЗД на 30 – 38 тижнях вагітності, ШОЕ крові на 21 тижні вагітності, ШОЕ крові на 30 тижні вагітності. Подальше дослідження буде направлено на виявлення менш залежних факторів розвитку патології з пограничних значень.

Список літератури: 1. *Наследов А.Д.* SPSS 19: профессиональный статистический анализ данных / *А.Д. Наследов.* – СПб.: Питер, 2011. – 400 с. 2. *Tsanas A.* A methodology for the analysis of medical data. Chapter 7 in *Handbook of Systems and Complexity in Health* / *A. Tsanas, M.A. Little, P.E. McSharry* // Springer. – 2013. – P. 113-125. 3. *Duy Ngo* An exploratory data analysis of electroencephalograms using the functional boxplots approach / *Duy Ngo, Ying Sun, M. G. Genton, J. Wu, R. Srinivasan, S.C. Cramer and H. Ombao* // *Frontiers in Neuroscience.* – 2015. – Vol. 9. – P. 1-18. 4. *Hasegawa J.* Current status of pregnancy-related maternal mortality in Japan: a report from the Maternal Death Exploratory Committee in Japan / *J. Hasegawa, A. Sekizawa, H. Tanaka, et al.* // *BMJ Open.* – 2016. – Vol. 79. – Issue 3. – P. 1835-1840. 5. *Мерцалова О. В.* Перинатальні гіпоксичні ураження центральної нервової системи плода у вагітних високого ризику (діагностика, прогноз виходів, оптимізація ведення вагітності та пологів) : Дис... д-ра наук: 14.01.01. – Харків. – 2002. – 40 с. 6. *Martinez-Biarge, M.* Antepartum and intrapartum factors preceding neonatal hypoxic-ischemic encephalopathy / *Martinez-Biarge, M., Diez-Sebastian, J., Wusthoff, et al.* // *Pediatrics.* – 2013. – Vol. 132. – P. 952-959. 7. *Milsom I.* Influence of maternal, obstetric and fetal risk factors on the prevalence of birth asphyxia at term in a Swedish urban population / *I. Milsom, L. Ladfors, K. Thiringer, et al.* // *Acta Obstet Gynecol Scand.* – 2002. – Vol. 81. – P. 909-917. 8. *Bell M.L.* Handling missing data in RCTs; a review of the top medical journals / *M.L. Bell, M. Fiero, N.J. Horton and Hsu Chiu-Hsieh* // *BMC Medical Research Methodology.* – 2014. 9. *Cochrane Handbook for Systematic Reviews of Interventions – Version 5.1.0* [Електронний ресурс] – Режим доступу [www. URL: http://www.cochrane.org/handbook](http://www.cochrane.org/handbook) – 10.08.2016 р. 10. *Sauro J.* 7 Ways To Handle Missing Data – 2015 [Електронний ресурс] – Режим доступу [www. URL: http://www.measuringu.com/blog/handle-missing-data.php](http://www.measuringu.com/blog/handle-missing-data.php) – 10.08.2016 р. 11. *Descriptive Statistics and Exploratory Data Analysis – 2008* [Електронний ресурс] – Режим доступу [www. URL: http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture2.pdf](http://www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture2.pdf) – 10.08.2016 р. 12. *Stark P.B.* Selected by the Scout Report for Business and Economics – 2013 [Електронний ресурс] – Режим доступу [www. URL: http://www.stat.berkeley.edu/~stark/SticiGui/Text/location.htm](http://www.stat.berkeley.edu/~stark/SticiGui/Text/location.htm) – 10.08.2016 р. 13. *McDonald J.H.* Handbook of biological statistics / *McDonald J.H.* // Baltimore: Sparky House Publishing, 2008. – 291 p. 14. *Term H.* Descriptive Statistics for Research – 2008 [Електронний ресурс] – Режим доступу [www. URL: https://www.stats.ox.ac.uk/pub/bdr/IAUL/Course1Notes8.pdf](https://www.stats.ox.ac.uk/pub/bdr/IAUL/Course1Notes8.pdf) – 10.08.2016 р. 15. *Шеломоский В.В.* Математическая статистика – Електронний ресурс] – Режим доступу [www. URL: http://www.exponenta.ru/educat/systemat/shelomovsky/lab/lab04.asp](http://www.exponenta.ru/educat/systemat/shelomovsky/lab/lab04.asp) – 10.08.2016 р. 16. *Graham Hole* Research Skills Kruskal-Wallis handout – 2011 [Електронний ресурс] – Режим доступу [www. URL: http://users.sussex.ac.uk/~grahamh/RM1web/Kruskal-Wallis%20Handout2011.pdf](http://users.sussex.ac.uk/~grahamh/RM1web/Kruskal-Wallis%20Handout2011.pdf) – 10.08.2016 р.

References:

1. Nasledov, A.D. (2011), *SPSS 19: professional statistical data analysis*, SPb.: Piter, 400 p.
2. Tsanas, A., Little, M.A., and McSharry, P.E. (2013), "A methodology for the analysis of medical data", in Sturmborg, J.P. and Martin, C.M. (Ed.) *Chapter 7 in Handbook of Systems and Complexity in Health*, Springer, pp. 113-125.

3. Ngo, D., Sun, Y., Genton, M.G., Wu, J., Srinivasan, R., Cramer, S. C., and Ombao, H. (2015), "An exploratory data analysis of electroencephalograms using the functional boxplots approach", *Frontiers in Neuroscience*, Vol. 9, pp. 1-18.
4. Hasegawa, J., Sekizawa, A., Tanaka, H. et al. (2015), "Current status of pregnancy-related maternal mortality in Japan: a report from the Maternal Death Exploratory Committee in Japan", *Circulation Journal*, Vol. 79, pp. 1835-1840.
5. Mertsalova, O.V. (2002), *Perinatal hypoxic central nervous system of the fetus in high risk pregnant women (diagnosis, prognosis outputs, optimization of pregnancy and childbirth): dissertation: Dis... d-ra med. nauk.*, Harkiv, 40 p.
6. Martinez-Biarge, M., Diez-Sebastian, J., Wusthoff, et al. (2013), "Antepartum and intrapartum factors preceding neonatal hypoxic-ischemic encephalopathy", *Pediatrics*, Vol. 132, pp. 952 – 959.
7. Milsom, I., Ladfors, L., Thiringer, K. et al. (2002), "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*, Vol. 81, pp. 909- 917.
8. Bell, M. L, Fiero, M., Horton, N. J and Chiu-Hsieh Hsu (2014), "Handling missing data in RCTs; a review of the top medical journals", *BMC Medical Research Methodology*, Vol. 14, pp.
9. Cochrane Handbook for Systematic Reviews of Interventions – Version 5.1.0 (2011), available at: http://handbook.cochrane.org/front_page.htm
10. Sauro, J. (2015), 7 Ways To Handle Missing Data available at: www.measuringu.com/blog/handle-missing-data.php
11. Descriptive Statistics and Exploratory Data Analysis (2008), available at: www.gs.washington.edu/academics/courses/akey/56008/lecture/lecture2.pdf
12. Stark, P.B. (2013), Selected by the Scout Report for Business and Economics available at: www.stat.berkeley.edu/~stark/SticiGui/Text/location.htm
13. McDonald, J.H. (2008) *Handbook of biological statistics*, Baltimore: Sparky House Publishing, 291 p.
14. Term, H. (2008), Descriptive Statistics for Research available at: www.stats.ox.ac.uk/pub/bdr/IAUL/Course1Notes8.pdf.
15. Shelomovsky, V., "Mathematical statistics" available at: www.exponenta.ru/educat/systemat/shelomovsky/lab/lab04.asp
16. Graham, Hole (2011), Research Skills Kruskal-Wallis handout available at: <http://users.sussex.ac.uk/~grahamh/RM1web/Kruskal-Wallis%20Handoout2011.pdf>
15. Shelomovsky, V.V. Matemeticheskaya ststistika – [Elektronnij resurs] – Rezhim dostupu www. URL: <http://www.exponenta.ru/educat/systemat/shelomovsky/lab/lab04.asp> – 10.08.2016.
16. Graham, Hole, (2011), Research Skills Kruskal-Wallis handout –[Elektronnij resurs] – Rezhim dostupu www. URL: <http://users.sussex.ac.uk/~grahamh/RM1web/Kruskal-Wallis%20Handoout2011.pdf> – 10.08.2016.

Статтю представив д-р техн. наук, проф. СХУ ім. Даля Рязанцев О.І.

Надійшла (received) 09.08.2016

Skarga-Bandurova Inna, D.Sci.Tech., Assoc. Professor, Professor
Volodymyr Dahl East Ukrainian National University
Avenue 59-a, Central, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: skarga_bandurova@ukr.net
ORCID ID: 0000-0003-3458-8730

Biloborodova Tetyana, PhD student
Volodymyr Dahl East Ukrainian National University
Avenue 59-a, Central, Severodonetsk, Luhansk region, Ukraine, 93400
tel./phone: (064) 522-89-97, e-mail: beloborodova.t@gmail.com
ORCID ID: 0000-0001-7561-7484

УДК 004.67:618.3

Пошуковий аналіз даних для визначення релевантних факторів гіпоксичного ураження плода / Скарга-Бандурова І.С., Білобородова Т.О. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2016. – № 44 (1216). – С. 102 – 115.

Обґрунтовано актуальність пошукового аналізу даних для вивчення властивостей досліджуваних даних перебігу вагітності. Проведено обробку даних з використанням статистичних, графічних, формальних параметричних та непараметричних методів дослідження. За результатами кластерного аналізу виділено групу факторів, з обґрунтованим значущим впливом на розвиток патології плода. Підтверджено наявність взаємозв'язку між окремими виділеними факторами перебігу вагітності та наявністю досліджуваної патології у новонародженого, для досліджуваного набору даних отримано нові угруповання факторів перебігу вагітності. Іл.: 2. Табл.: 3. Бібліогр.: 16 назв.

Ключові слова: пошуковий аналіз даних, кластерний аналіз, патологія у новонародженого, угруповання факторів.

УДК 004.67:618.3

Поисковый анализ данных для определения релевантных факторов гипоксического поражения плода / Скарга-Бандуров И.С., Белобородова Т.А. // Вестник НТУ "ХПИ". Серия: Информатика и моделирование. – Харьков: НТУ "ХПИ". – 2016. – № 44 (1216). – С. 102 – 115.

Обоснована актуальность поискового анализа данных для изучения свойств исследуемых данных течения беременности. Проведена обработка данных с использованием статистических, графических, формальных параметрических и непараметрических методов исследования. По результатам кластерного анализа выделена группа факторов, с обоснованным значимым влиянием на развитие патологии плода. Подтверждено наличие взаимосвязи между отдельными выделенными факторами течения беременности и наличием исследуемой патологии у новорожденного, для исследуемого набора данных получены новые группировки факторов течения беременности. Ил.: 2. Табл.: 3. Библиогр.: 16 назв.

Ключевые слова: поисковый анализ данных, кластерный анализ, патология у новорожденного, группировка факторов.

UDC 004.67:618.3

Exploratory data analysis to identifying meaningful factors of hypoxic fetal injuries / Skarga-Bandurova I.S., Biloborodova T.O. // Herald of the National Technical University "KhPI". Subject issue: Information Science and Modelling. – Kharkov: NTU "KhPI". – 2016. – № 44 (1216). – P. 102 – 115.

The study aimed to exploratory data analysis to study the properties of the retrospective data of pregnancy. In this paper, we performed data processing using statistical, graphical, formal parametric and nonparametric methods. The results of cluster analysis highlighted a group of factors with a significant influence on the sound development of the pathology of the newborn. It confirmed the presence of the relationship between the individual factors isolated pregnancy and the presence of the studied disease in the newborn, new grouping factors for the test data were obtained. Figs.: 2. Tabl.: 3. Refs.: 16 titles.

Keywords: exploratory data analysis, cluster analysis, pathology of the newborn, grouping factors.