

## ЗАСТОСУВАННЯ КЛАСТЕРНИХ ТА КЛАСИФІКАЦІЙНИХ МЕТОДІВ DATAMINING У ДОСЛІДЖЕННІ МЕХАНІЗМІВ ВЗАЄМОБТЯЖЕННЯ РЕАКТИВНОГО АРТРИТУ ТА ХРОНІЧНОГО ПІЄЛОНЕФРИТУ ЗА ПОКАЗНИКАМИ МЕТАБОЛІЗМУ СПОЛУЧНОЇ ТКАНИНИ

О.В. Залявська, О.С. Хухліна, Л.П. Ткач, О.М. Ніка

Буковинський державний медичний університет, м. Чернівці, Україна

**Ключові слова:** реактивні артрити, хронічний пієлонефрит, кластерні та класифікаційні методи аналізу DataMining.

Буковинський медичний вісник. 2021. Т. 25, № 4 (100). С. 23-29.

DOI: 10.24061/2413-0737.XXV.4.100.2021.4

E-mail: zaliavska.olena@bsmu.edu.ua

**Резюме.** Порушення обмінних процесів у системі «протеоглікани-колаген», зміна ферментативних реакцій у хворих на реактивний артрит (РеА) найчастіше передують багатьом ускладненням, у тому числі і порушенню функції нирок. Класичні регресійні методи малоінформативні у визначенні взаємобтяження РеА та хронічного пієлонефриту (ХП), а також первинності виникнення хвороби.

**Мета роботи** – дослідити можливі взаємозв'язки між РеА і ХП з метою встановлення ранніх критеріїв прогнозування розвитку ХП на тлі РеА за показниками метаболізму сполучної тканини.

**Матеріал і методи.** Обстежено 113 пацієнтів, яких розподілено на дві групи: перша група – хворі на урогенний РеА, активність I-III ст. ФНС I-III ст. ( $n=65$ ); друга група – хворі на урогенний РеА та ХП I-II ст.: пієлонефрит у фазі загострення, ( $n=48$ ). Контрольну групу склали 20 практично здорових осіб. Середня тривалість захворювання обстежених пацієнтів складала  $24,4 \pm 4,7$  місяця. Середній вік пацієнтів складав  $32,5 \pm 1,2$  року. Для обробки отриманих результатів досліджень використано кластерні та класифікаційні методи аналізу DataMining.

**Результати.** У результаті кластеризації методами  $k$ -середніх та нечіткої кластеризації отримано однакові результати належності до кластерів. Зокрема, вірно діагностовано хворобу РеА в 48 випадках, що становить 74 % всіх діагностованих, 17 осіб (26 %) віднесено до кластеру «2». У другому кластері є теж невідповідність. Зокрема з 32 осіб, в яких діагностовано РеА+ХП, у цей кластер, відповідно до математичних розрахунків, потрапило 28 осіб, тобто 88 %. Чотири пацієнти (13 %) віднесено до кластеру «1», тобто до осіб, в яких має бути діагностована лише одна хвороба РеА. Це свідчить про те, що межа між кластерами РеА та РеА+ХП є децю розмита. А це є підставою встановлення того факту, що хвороба РеА поступово може призвести до ХП.

**Висновки.** За допомогою методів кластеризації DataMining встановлено найбільшу значущість у діагностичному алгоритмі прогресування РеА таких показників, як вміст у крові вільного оксипроліну (ВОП), білокзв'язаного оксипроліну (БЗОП), ступінь колагенолітичної активності (КЛА) плазми крові (інтенсивність лізису азоколу), котрі показали пряму залежність від ступеня активності запального процесу. Зростання ВОП  $>13,8$  мкмоль/л, БЗОП  $>65,0$  мкмоль/л та КЛА (за азоколом)  $>0,85$  мкг/мл за 1 год є вірогідними чинниками ризику прогресування та ранніми критеріями тяжкого перебігу РеА та ХП.

## ПРИМЕНЕНИЕ КЛАСТЕРНЫХ И КЛАССИФИКАЦИОННЫХ МЕТОДОВ DATAMINING В ИССЛЕДОВАНИИ МЕХАНИЗМОВ ВЗАИМООТЯГОЩЕНИЯ РЕАКТИВНОГО АРТРИТА И ХРОНИЧЕСКОГО ПИЕЛОНЕФРИТА ПО ПОКАЗАТЕЛЯМ МЕТАБОЛИЗМА СОЕДИНИТЕЛЬНОЙ ТКАНИ

О.В. Залявская, О.С. Хухлина, Е.П. Ткач, О.М. Ника

**Ключевые слова:** реактивные артриты, хронический пиелонефрит, кластерные и классификационные методы

**Резюме.** Нарушения обменных процессов в системе «протеогликианы-колаген», изменение ферментативных реакций у больных реактивным артритом (РеА) чаще всего предшествуют многим осложнениям, в том числе и нарушению функции почек. Классические регрессионные методы

## Оригінальні дослідження

анализа DataMining.

Буковинський медичний вісник. 2021. Т. 25, № 4 (100). С. 23-29.

малоінформативны в определении взаимоотношения РеА и хронического пиелонефрита (ХП), а также первичности возникновения болезни.

**Цель работы** – исследовать возможные взаимосвязи между РеА и ХП с целью установления ранних критериев прогнозирования развития ХП на фоне РеА по показателям метаболизма соединительной ткани.

**Материал и методы.** Обследовано 113 пациентов, которые распределены на две группы: первая группа – больные урогенным РеА, активность I-III ст. ФНС I-III ст. (n=65); вторая группа – больные урогенным РеА и ХБП I-II ст.: пиелонефрит в фазе обострения, (n=48). Контрольную группу составили 20 практически здоровых лиц. Средняя продолжительность заболевания обследованных пациентов составляла  $24,4 \pm 4,7$  месяца. Средний возраст пациентов составлял  $32,5 \pm 1,2$  лет. Для обработки результатов исследований использованы кластерные и классификационные методы анализа DataMining.

**Результаты.** В результате кластеризации методами k-средних и нечеткой кластеризации получены одинаковые результаты принадлежности к кластерам. В частности, верно диагностирована болезнь РеА в 48 случаях, что составляет 74% всех диагностированных, 17 человек (26%) отнесены к кластеру «2». Во втором кластере есть тоже несоответствие. В частности, из 32 человек, у которых диагностирована РеА+ХП, в этот кластер, согласно математическим расчетам, попало 28 человек, то есть 88%. Четыре пациента (13%) отнесены к кластеру «1», то есть к лицам, у которых должна быть диагностирована только одна болезнь РеА. Это свидетельствует о том, что граница между кластерами РеА и РеА+ХП несколько размыта. А это основание установления того факта, что болезнь РеА постепенно может привести к ХП.

**Выводы.** С помощью методов кластеризации DataMining установлена наибольшая значимость в диагностическом алгоритме прогрессирования РеА таких показателей, как содержание в крови свободного оксипролина (СОП), белоксвязанного оксипролина (БСОП), степень коллагенолитической активности (КЛА) плазмы крови (интенсивность лизиса азокола), показавшие прямую зависимость от степени активности воспалительного процесса. Рост СОП  $>13,8$  мкмоль/л, БСОП  $>65,0$  мкмоль/л и КЛА (за азоколом)  $>0,85$  мкг/мл за 1ч являются возможными факторами риска прогрессирования и ранними критериями тяжелого течения РеА и ХП.

## APPLICATION OF DATAMINING CLUSTERING AND CLASSIFICATION METHODS IN THE STUDY OF THE MECHANISMS OF MUTUAL BURDENING OF REACTIVE ARTHRITIS AND CHRONIC PYELONEPHRITIS IN TERMS OF CONNECTIVE TISSUE METABOLISM

O.V. Zaliavska, O.S. Khukhlina, Ye.P. Tkach, O.M. Nika

**Key words:** reactive arthritis, chronic pyelonephritis, DataMining clustering and classification methods.

Bukovinian Medical Herald. 2021. V. 25, № 4 (100). P. 23-29.

**Resume.** Disorders of metabolic processes in the "proteoglycans-collagen" system, changes in enzymatic reactions in patients with reactive arthritis (ReA) are often preceded by many complications, including impaired renal function. Classical regression methods are not very informative in determining the mutual burden of ReA and chronic pyelonephritis (CP), as well as the primacy of the disease.

**Objective** - to investigate the possible relationship between ReA and CP in order to establish early criteria for predicting the development of CP on the background of ReA on the indicators of connective tissue metabolism.

**Material and methods.** 113 patients were examined, which were divided into two groups: the first group - patients with urogenic ReA, activity I-III. FTS I-III st. (n = 65); the second group - patients with urogenic ReA and CKD I-II: pyelonephritis in the acute phase (n = 48). The control group consisted of 20 healthy individuals. The average duration of the disease of the examined patients was  $24.4 \pm 4.7$  months. The mean age of patients was  $32.5 \pm 1.2$  years. DataMining clustering and classification methods were used to process the obtained research results.

**Results.** As a result of clustering methods (k-means and fuzzy clustering), the same

results of cluster membership were obtained. In particular, ReA disease was correctly diagnosed in 48 cases, which is 74% of all diagnosed, 17 people (26%) were assigned to cluster "2". There is also a mismatch in the second cluster. In particular, out of 32 people diagnosed with ReA + CP, according to mathematical calculations, 28 people (88%) got into this cluster. Four patients (13%) were assigned to cluster "1" – to the group of persons in whom only one ReA disease should be diagnosed. This indicates that the boundary between the ReA and ReA + CP clusters are somewhat blurred. And this is the basis for establishing the fact that ReA disease can gradually lead to CP.

**Conclusions.** Using DataMining clustering methods the greatest significance was defined in the diagnostic algorithm of ReA progression of such indicators as free oxypoline (FOP), protein-bound oxypoline (PBOP), the degree of collagenolytic activity (CLA) (intensity of azocol lysis), which showed a direct dependence on the degree of activity of the inflammatory process. Increases in FOP > 13.8  $\mu\text{mol/l}$ , PBOP > 65.0  $\mu\text{mol/l}$  and CLA (azocol) > 0.85  $\mu\text{g/ml}$  in 1 h are probable risk factors for progression and early criteria for severe ReA and CP.

**Вступ.** Більшість лабораторних показників, що використовуються в діагностиці реактивних артритів (ReA), мають неспецифічний характер, відображаючи наявність і активність запального процесу [1]. Великою інформативністю про перебіг патологічного процесу в уражених суглобах і про ступінь зниження інтенсивності даного процесу під впливом проведеного лікування володіють методики, які характеризують обмін специфічних компонентів сполучної тканини – колагену та протеогліканів. Порушення обмінних процесів у системі «протеоглікани-колаген», зміна ферментативних реакцій найчастіше передують багатьом ускладненням, у тому числі і порушенню функції нирок [2].

Класичні регресійні методи малоінформативні у визначенні взаємобтяження ReA та хронічного піелонефриту (ХП), а також первинності виникнення хвороби.

**Мета роботи** – дослідити можливі взаємозв'язки між ReA і ХП з метою встановлення ранніх критеріїв прогнозування розвитку ХП на тлі ReA за показниками метаболізму сполучної тканини.

**Матеріал і методи.** Обстежено 113 пацієнтів, яких розподілено на дві групи: перша група – хворі на урогенний ReA, активність I-III ст. ФНС I-III ст. (n=65); друга група – хворі на урогенний ReA та ХХН I-II ст.: піелонефрит у фазі загострення, (n=48). Контрольну групу склали 20 практично здорових осіб (ПЗО). Середня тривалість захворювання обстежених пацієнтів складала 24,4 $\pm$ 4,7 місяця. Середній вік пацієнтів складав 32,5 $\pm$ 1,2 року.

Зміни метаболізму вуглеводно-білкових компонентів позаклітинного матриксу визначали за вмістом у крові вільного оксипроліну (ВОП) за методикою С. С. Тетянець (1985), білковозв'язаного оксипроліну (БЗОП) за методом М. А. Осадчука (1979), гексуранових кислот (ГК) і гексозамінів (ГА) за методом О. Г. Архіпової (1988), серомукоїду (СМ), сіалових кислот (СК) за методом Л. А. Муравйової, Е. Ю. Волкова (1988), рівень колагенолітичної активності (КЛА) плазми за методом П. Н. Шараєва (1987).

Критеріями включення в дослідження були хворі на урогенний реактивний артрит із супровідною хронічною хворобою нирок I-II ст.: піелонефритом у фазі загострення, особи віком від 18 до 35 років, верифіковані за допомогою клінічних, лабораторних та інструментальних методів дослідження, інформована письмова згода пацієнта на участь у дослідженні.

Для обробки даних використано три різних методи кластеризації, що базуються на абсолютно різних математичних теоріях: класичний метод k-середніх, карти Кохонена (нейронні мережі), нечіткі k-середні (нечітка логіка). Як вхідні дані виступає множина чинників, результатом кластеризації служить номер кластеру, до якого відноситься пацієнт. Однак на відміну від двох перших методів, нечітка кластеризація дозволяє встановити ступінь належності пацієнта до того чи іншого кластеру.

*Метод k-середніх*

Метод k-середніх (k-means) – це метод кластеризації, мета якого – розділити n спостережень на k кластерів так, щоб кожне спостереження належало до кластеру з найближчим до нього середнім значенням. Метод базується на мінімізації суми квадратів відстаней між кожним спостереженням та центром його кластеру, тобто функції

$$\sum_{i=1}^N d(x_i, m_j(x_i))^2$$

де d – метрика,  $x_i$  – i-й об'єкт даних, а  $m_j(x_i)$  – центр кластеру, якому на j-й ітерації приписаний елемент  $x_i$ .

Алгоритм кластеризації такий:

1. Існуючий масив спостережень (пацієнтів), кожен з яких має множину ознак (чинників), розташовується у багатовимірному просторі.
2. Визначається кількість кластерів, що необхідно утворити.
3. Випадковим чином обирається k спостережень, які на цьому кроці вважаються

## Оригінальні дослідження

центрами кластерів.

4. Кожне спостереження «приписується» до одного з  $n$  кластерів – того, відстань до якого найкоротша.

5. Розраховується новий центр кожного кластеру як елемент, ознаки якого розраховуються як середнє арифметичне ознак об'єктів, що входять у цей кластер.

6. Проводиться така кількість ітерацій (повторюються кроки 3-4), поки центри кластерів стануть стійкими (тобто при кожній ітерації в кожному кластері опиняться одні й ті самі об'єкти), дисперсія всередині кластеру буде мінімізована, а між кластерами – максимізована.

### *Карти Кохонена*

Самоорганізована карта Кохонена (англ. Self-organizing map - SOM) - змагальна нейронна мережа з навчанням без учителя, що виконує задачу візуалізації та кластеризації. Ідея мережі запропонована фінським вченим Т. Кохоненом, є методом проєкціювання багатовимірного простору в простір із більш низькою розмірністю (найчастіше двовимірний). Є однією з версій нейронних мереж Кохонена.

Самоорганізована карта складається з компонентів, так званих вузлів або нейронів. Їх кількість задається аналітиком. Кожен із вузлів описується двома векторами. Перший – вектор ваги  $m$ , що має таку ж розмірність, що і вхідні дані. Другий – вектор  $r$ , координати вузла на карті. Зазвичай вузли розташовують у вершинах регулярної решітки з квадратними або шестикутними комірками.

Спочатку будується початковий варіант карти. У процесі навчання вектори ваги вузлів наближаються до вхідних даних. Для кожного спостереження вибирається найбільш схожий по вектору ваги вузол, і значення його вектора ваги наближається до спостереження. Також до спостереження наближаються вектори ваги декількох вузлів, розташованих поруч таким чином, якщо в множині вхідних даних два спостереження були схожі, на карті їм будуть відповідати близькі вузли. Циклічний процес навчання закінчується по досягненні картою допустимої (заздалегідь заданою аналітиком) похибки, або при виконанні заданої кількості ітерацій.

### *Нечітка кластеризація*

Алгоритм нечіткої кластеризації називають FCM-алгоритмом (Fuzzy Classifier Means, Fuzzy C-Means). Метою FCM-алгоритму кластеризації є автоматична класифікація множини об'єктів, які задаються векторами ознак у просторі ознак. Іншими словами, такий алгоритм визначає кластери і відповідно класифікує об'єкти. Кластери представляються нечіткими множинами, і, крім того, границі між кластерами також є нечіткими.

FCM-алгоритм кластеризації припускає, що об'єкти належать всім кластерам із певною функцією належності. Ступінь належності визначається

відстанню від об'єкта до відповідних центрів кластів. Даний алгоритм ітераційно обчислює центри кластерів і нові ступені належності об'єктів.

### *Класифікаційний аналіз*

Методи кластеризації дають змогу підтвердити зв'язок між кластерами. Однак вони не дають можливості встановити правила співвідношення одного кластера до того чи іншого, а отже, проводити висновок без уведення інформаційних систем, в які запрограмовані вищезазначені методи кластеризації. Ці правила зручно побудувати у вигляді дерев рішень, які можна отримати за допомогою класифікаційного аналізу нейронними мережами. Древа рішень - це спосіб представлення правил в ієрархічній, послідовній структурі, де кожному об'єкту відповідає єдиний вузол, що дає рішення.

У роботі був використаний метод CART (Classification and Regression Tree) – це алгоритм побудови бінарного дерева рішень – дихотомічної класифікаційної моделі. В алгоритмі CART кожен вузол дерева рішень має двох нащадків. На кожному кроці побудови дерева правило, сформоване у вузлі, ділить задану множину прикладів (навчальну вибірку) на дві частини - частина, в якій виконується правило (нащадок - right) і частина, в якій правило не виконується (нащадок - left). Для вибору оптимального правила використовується функція оцінки якості розбиття. Реалізація цього методу здійснюється нейронними мережами.

### **Результати дослідження та їх обговорення**

Дослідження показників стану сполучної тканини у хворих на РеА із коморбідним перебігом ХП показали вірогідне підвищення вмісту в крові ВОП, БЗОП, показників колагенолітичної та протеолітичної активності плазми, що свідчить про одночасне підсилення синтезу та розпаду колагену на тлі активації колагенолізу [3].

Найбільшу значущість у діагностичному алгоритмі прогресування та клінічний перебіг РеА мають такі показники, як вміст у крові вільного та білковозв'язаного оксипроліну, гексозамінів, ступінь колагенолітичної активності плазми крові (інтенсивність лізису азоколу), котрі показали пряму залежність від ступеня активності запального процесу.

У групі хворих на РеА із супровідним ХП спостерігається підвищення ступеня активності РеА порівняно з ізольованим перебігом РеА за показниками запалення та реакції сполучної тканини, що сприяє його прогресуванню [3].

З метою дослідження можливих взаємозв'язків між РеА і ХП та встановлення критеріїв прогнозування розвитку ХП на тлі РеА за показниками метаболізму сполучної тканини, використано кластерні та класифікаційні методи DataMining [4, 5]. Фактично завдання зводиться до кластеризації об'єктів за множиною вищезазначених множин чинників на три кластери. Якщо в результаті кластеризації один кластер міститиме лише здорових людей, другий – хворих на РеА, а третій – на

PeA+ХП, це буде підтвердженням гіпотези. Подальший класифікаційний аналіз допоможе встановити важливість факторів та побудувати дерево рішень, згідно з яким можна діагностувати хворобу.

Також з метою вивчення можливих взаємозв'язків між PeA і прогресування ХП на тлі PeA при дослідженні метаболізму сполучної тканини і встановлення критеріїв тяжкості використано три різних методи кластеризації, що базуються на абсолютно різних математичних теоріях: класичний метод k-середніх, карти Кохонена (нейронні мережі), метод k-середні (нечітка логіка).

У результаті кластеризаційного аналізу пацієнти розподілилися таким чином (табл. 1). Всі здорові особи потрапили в окремий кластер (кластер «0»). Такі висновки отримані згідно з усіма трьома методами кластеризації. Відмінності спостерігаються в діагностуванні таких хвороб, як PeA та PeA+ХП.

Згідно з результатами, отриманими при побудові карти Кохонена, особи, в яких діагностували PeA+ХП, потрапили в окремий кластер (кластер «2»). Особи, в яких діагностовано лише хвороба PeA (65 пацієнтів), згідно з картами Кохонена, в основному потрапили в кластер «1» (54 пацієнти – 83%). Деякі з них (11 пацієнтів – 17 %) класифіковані як такі, що належать до кластеру «2». Тобто, мають мати ознаки двох хвороб – PeA+ХП, хоча насправді в них діагностовано лише PeA. Це може означати, що ця група осіб знаходиться в групі ризику розвитку ХП. Тобто, такий аналіз може дати підстави вважати, що PeA є причиною прогресування ХП.

У результаті кластеризації методами k-середніх та

Таблиця 1

#### Результати класифікаційного аналізу за методом нейронних мереж

Групи обстежених хворих із реальними діагнозами				
	Класифіковано			
Фактично	Здоровий	PeA	PeA+ХП	Разом
Здоровий	20			20
PeA		65		65
PeA+ХП			32	32
Разом	20	65	32	117
Карта Кохонена				
	Класифіковано			
Фактично	Здоровий	PeA	PeA+ХП	Разом
Здоровий	20			20
PeA		54		54
PeA+ХП			43	43
Разом	20	54	43	117
k-середні (Нечітка кластеризація)				
	Класифіковано			
Фактично	Здоровий	PeA	PeA+ХП	Разом
Здоровий	20			20
PeA		52		52
PeA+ХП		1	44	45
Разом	20	53	44	117

нечіткої кластеризації отримано однакові результати належності до кластерів. Однак ці результати дещо відрізняються від попереднього методу. Зокрема, вірно діагностовано хвороба PeA в 48 випадках, що становить 74 % всіх діагностованих, 17 осіб (26 %) віднесено до кластеру «2». У другому кластері є теж невідповідність. Зокрема з 32 осіб, в яких діагностовано PeA+ХП, у цей кластер, відповідно до математичних розрахунків, потрапило 28 осіб, тобто 88 %. Чотири пацієнти (13 %) віднесено до кластеру «1», тобто до осіб, в яких має бути діагностована лише одна хвороба PeA. Це свідчить про те, що межа між кластерами PeA та PeA+ХП є дещо розмита. А це є підставою встановлення того факту, що хвороба PeA поступово може призвести до ХП. Це добре аналізувати саме методом нечіткої кластеризації. Зокрема, якщо в результаті кластеризації для нового пацієнта отримано такі дані:  $\mu_0 = 0,95$ ,  $\mu_1 = 0,08$ ,  $\mu_2 = 0,02$  – це означатиме, що людина є здоровою, її можна співвіднести з кластером «0» на 95 %. У випадку ж  $\mu_0 = 0,01$ ,  $\mu_1 = 0,56$ ,  $\mu_2 = 0,43$  можна стверджувати, що в людини є PeA. Якщо формально він належить до кластеру 1, його показники є близькими до кластеру «2». Тобто, у подальшому в нього можна буде діагностувати і хворобу ХП. Аналогічно у випадку лікування цим методом можна відслідковувати «переміщення» пацієнта між кластерами.

Методи кластеризації дали змогу підтвердити зв'язок між хворобою PeA та ХП. Однак вони не дають можливості встановити правила співвідношення нових пацієнтів до того чи іншого кластеру, а отже, проводити попередній діагноз без уведення інформаційних систем, в які запрограмовані вищезазначені методи кластеризації. Ці правила зручно побудувати у вигляді дерев рішень, які можна отримати за допомогою класифікаційного аналізу нейронними мережами. Древа рішень – це спосіб представлення правил в ієрархічній, послідовній структурі, де кожному об'єкту відповідає єдиний вузол, що дає рішення.

У роботі використаний метод CART (Classification and Regression Tree) – це алгоритм побудови бінарного дерева рішень – дихотомічної класифікаційної моделі. В алгоритмі CART кожен вузол дерева рішень має двох нащадків. На кожному кроці побудови дерева правило, сформоване у вузлі, ділить задану множину прикладів (навчальну вибірку) на дві частини – частина, в якій виконується правило (нащадок – right) і частина, в якій правило не виконується (нащадок – left). Для вибору оптимального правила використовується функція оцінки якості розбиття. Реалізація цього методу здійснюється нейронними мережами.

Вхідними параметрами такого аналізу є ті самі чинники, а вихідним – номер кластеру. Для встановлення правил класифікаційний аналіз проведений для трьох вихідних полів: реальний діагноз, кластер згідно з картами Кохонена, кластер згідно з k-середніми (або нечіткої кластеризації). Як

## Оригінальні дослідження

видно з таблиці спряженості (див. табл. 1), у перших двох випадках, нейронна мережа зі 100% коректністю може класифікувати приналежність пацієнта до того чи іншого кластеру. У випадку нечіткої кластеризації нейронна мережа помилилася лише один раз, віднісши людину з хворобами РеА+ХП до першого кластеру.

Така точність результатів дає змогу представити дерево рішень та визначити значимість чинників для постановки діагнозу. У першому випадку, коли класифікували реальні діагнози, значимим є лише один чинник – ВОП.

Дерево рішень виглядає так:

Діагноз	ВОП, мкмоль/л
Здоровий	$\text{ВОП} < 13,925$
РеА	$13,925 \leq \text{ВОП} < 15,885$
РеА+ХП	$\text{ВОП} \geq 15,885$

У випадку кластеризації методом карт Кохонена, отримані наступні класифікаційні правила:

Діагноз	ВОП, мкмоль/л, ПАК за азоколом, мкг/мл за 1 год
Здоровий	$\text{ВОП} < 13,925$
РеА	$\text{ВОП} \geq 13,925$ $\text{ПАК за азоколом} < 0,855$
РеА+ХП	$\text{ВОП} \geq 13,925$ $\text{ПАК за азоколом} \geq 0,855$

У цьому випадку значимими є лише атрибути ВОП та ПАК за азоколом. З цих двох таблиць видно, що правила для діагностування здорового пацієнта збігаються. Відмінність є лише при діагнозі РеА та РеА+ХП, які визначаються за показником ПАК.

Трохи іншою є картина у випадку застосування нечіткої кластеризації.

Діагноз	ВОП, БЗОП, мкмоль/л
Здоровий	$\text{ВОП} < 13,835$
РеА	$\text{ВОП} \geq 13,835$ $\text{БЗОП} < 65,025$
РеА+ХП	$\text{ВОП} \geq 13,835$ $\text{БЗОП} \geq 65,025$

Основним фактом належності пацієнта до кластеру здорових осіб залишається ВОП. Однак його порогова величина є меншою на 0,1 мкмоль/л і становить 13,835. Ключовим же фактором до визначення належності до типу хвороби є фактор БЗОП.

Отже, використання цих трьох таблиць може бути основою для попереднього аналізу взаємообтяження хвороб РеА та ХП.

**Висновки.** За допомогою методів кластеризації DataMining (класичний метод k-середніх, карти Кохонена (нейронні мережі), нечіткі k-середні (нечітка логіка)) встановлено найбільшу значущість

у діагностичному алгоритмі прогресування реактивних артритів таких показників, як вміст у крові вільного оксипроліну, білокзв'язаного оксипроліну, ступінь колагенолітичної активності плазми крові (інтенсивність лізису азоколу), котрі показали пряму залежність від ступеня активності запального процесу. Зростання ВОП > 13,8 мкмоль/л, БЗОП > 65,0 мкмоль/л та ПАК (за азоколом) > 0,85 мкг/мл за 1 год є вірогідними чинниками ризику прогресування та ранніми критеріями тяжкого перебігу реактивних артритів та хронічного пієлонефриту.

## Список літератури

1. Martijn AH, Gupta ZD, Bijlsma WJ, Boonen A, Chau J, Courvoisier DS, et al. International Consortium for Health Outcome Measurement Set of Outcomes That Matter to People Living With Inflammatory Arthritis: Consensus From an International Working Group. *Arthritis Care Res.* 2019;71(12):1556-65.
2. Chang MH, Nigrovic PA. Antibody-dependent and -independent mechanisms of inflammatory arthritis. *JCI Insight.* 2019;4(5):e125278. <https://doi.org/10.1172/jci.insight.125278>.
3. Залявська ОВ, Хухліна ОС, Каушанська ОВ, Гайдичук ВС. Стан метаболізму компонентів сполучної тканини у хворих на реактивний артрит. *Здобутки клінічної та експериментальної медицини.* 2014;1:44-7.
4. Kaur G, Jagdev G. A study of clustering and classification techniques involved in data mining. *International Journal of Advanced Technology in Engineering and Science.* 2017;5(5):360-71.
5. Arumugam P, Christy V. Analysis of Clustering and Classification Methods for Actionable Knowledge. *Materials Today: Proceedings.* 2018;5(1):1839-45.

## References

1. Martijn AH, Gupta ZD, Bijlsma WJ, Boonen A, Chau J, Courvoisier DS, et al. International Consortium for Health Outcome Measurement Set of Outcomes That Matter to People Living With Inflammatory Arthritis: Consensus From an International Working Group. *Arthritis Care Res.* 2019;71(12):1556-65.
2. Chang MH, Nigrovic PA. Antibody-dependent and -independent mechanisms of inflammatory arthritis. *JCI Insight.* 2019;4(5):e125278. <https://doi.org/10.1172/jci.insight.125278>.
3. Zaliavs'ka OV, Khukhlina OS, Kaushans'ka OV, Haidychuk VS. Stan metabolizmu komponentiv spoluchnoi tkanyuny u khvorykh na reaktyvnyi artryt [The state of metabolism of connective tissue components in patients with reactive arthritis]. *Zdobutky klinichnoi ta eksperymental'noi medytsyny.* 2014;1:44-7. (in Ukrainian).
4. Kaur G, Jagdev G. A study of clustering and classification techniques involved in data mining. *International Journal of Advanced Technology in Engineering and Science.* 2017;5(5):360-71.
5. Arumugam P, Christy V. Analysis of Clustering and Classification Methods for Actionable Knowledge. *Materials Today: Proceedings.* 2018;5(1):1839-45.

## Відомості про авторів

Залявська О.В. – канд. мед. наук, доцент кафедри внутрішньої медицини, клінічної фармакології та професійних хвороб, Буковинський державний медичний університет, м. Чернівці, Україна.

Хухліна О.С. – д-р мед. наук, професор, завідувачка кафедри внутрішньої медицини, клінічної фармакології та професійних хвороб, Буковинський державний медичний університет, м. Чернівці, Україна.

Ткач С.П. – канд. мед. наук, доцент кафедри внутрішньої медицини, клінічної фармакології та професійних

хвороб, Буковинський державний медичний університет, м. Чернівці, Україна.

Ніка О.М. – канд. мед. наук, доцент кафедри нервових хвороб, психіатрії та медичної психології, Буковинський державний медичний університет, м. Чернівці, Україна.

#### **Сведения об авторах**

Залаявская Е.В. – канд. мед. наук, доцент кафедры внутренней медицины, клинической фармакологии и профессиональных болезней, Буковинский государственный медицинский университет, г. Черновцы, Украина.

Хухлина О.С. – д-р мед. наук, профессор, заведующая кафедрой внутренней медицины, клинической фармакологии и профессиональных болезней, Буковинский государственный медицинский университет, г. Черновцы, Украина.

Ткач Е.П. – канд. мед. наук, доцент кафедры внутренней медицины, клинической фармакологии и профессиональных заболеваний, Буковинский государственный медицинский университет, г. Черновцы, Украина.

Ника О.М. – канд. мед. наук, доцент кафедры нервных болезней, психиатрии и медицинской психологии, Буковинский государственный медицинский университет, г. Черновцы, Украина.

#### **Information about the authors**

Zaliavska O.V. – Candidate of Medical Sciences, Associate Professor of the Department of Internal Medicine, Clinical Pharmacology and Occupational Diseases, Bukovinian State Medical University, Chernivtsi, Ukraine.

Khukhlina O.S. – Doctor of Medical Sciences, Professor, Head of the Department of Internal Medicine, Clinical Pharmacology and Occupational Diseases, Bukovinian State Medical University, Chernivtsi, Ukraine.

Tkach Ye.P. - Candidate of Medical Sciences, Associate Professor of Internal Medicine, Clinical Pharmacology and Occupational Diseases, Bukovinian State Medical University, Chernivtsi, Ukraine.

Nika O.M. – Candidate of Medical Sciences, Associate Professor of the Department of Nervous Diseases, Psychiatry and Medical Psychology, Bukovinian State Medical University, Chernivtsi, Ukraine.

*Надійшла до редакції 01.11.21*

*Рецензент – проф. Федів О.І.*

*© О.В. Залаявська, О.С. Хухліна, Л.П. Ткач, О.М. Ніка, 2021*