

УДК 681.004

DOI 10.36910/775.24153966.2023.76.5

О. І. Провотар, О. П. Ількун

Київський національний університет імені Тараса Шевченка

ПРО УДОСКОНАЛЕННЯ МЕТОДУ НЕЧІТКОГО ЛОГІЧНОГО ВИВЕДЕННЯ В СИСТЕМАХ ЗАСНОВАНИХ НА ЗНАННЯХ

В статті розглядаються два методи розробки та дослідження систем нечіткого виведення для діагностики стану пацієнта на основі множини нечітких симптомів. Описано задачу визначення хвороби пацієнта відповідно до даних діагностики стану здоров'я і поточних симптомів. Для вирішення цієї задачі в контексті нечіткої логіки запропоновано систему нечіткого логічного виведення, що використовує інформацію залежностей між хворобами, симптомами, рівнями симптомів. Представлено таблицю відповідності рівнів конкретних симптомів та хвороби. Для кожної лінгвістичної змінної поточного симптому та рівня хвороб побудовано нечіткі модифіковані трикутні функції належності. Розглянуто кілька способів задання правил, які окреслено у рамках двох методів. Перший метод використовує стандартний підхід до побудови системи нечіткого виведення. В його основі лежить поняття лінгвістичної змінної та нечіткого правила, яке створюється на основі експертних знань. Сформовано інтуїтивно зрозумілі прямі правила, які представлені у математичній формі. Використовуючи цю систему правил можна застосувати різні механізми виведення, наприклад, виведення Мамдані. Другий метод, в свою чергу, представляє удосконалену нечітку модель, що включає додаткові лінгвістичні змінні та правила з метою кількісного врахування правильно вказаних симптомів у множині всіх симптомів хвороби пацієнта. Дана система правил оперує лінгвістичними змінними, що використовують кількість правильно наданих симптомів відносно загальної кількості симптомів хвороби. Виконано порівняння описаних систем логічного виведення на тестовому наборі даних. Запропоновані методи аналізуються на предмет їх ефективності та точності при встановленні діагнозу. Результати дослідження можуть бути важливими для розробки новітніх методів діагностики, заснованих на нечіткій логіці, які можуть виявитись більш ефективними у порівнянні з традиційними методами.

Ключові слова: нечітка множина, нечіткі правила, системи нечіткого виведення, експертні системи.

O.I. Provota, O.P. Ilkun

ON THE IMPROVEMENT OF THE METHOD OF FUZZY INFERENCE IN KNOWLEDGE-BASED SYSTEMS

The article discusses two methods of developing and researching fuzzy inference systems for diagnosing a patient's condition based on a set of fuzzy symptoms. The problem of determining the patient's disease according to the data of the diagnosis of the state of health and current symptoms is described. To solve this problem in the context of fuzzy logic, a system of fuzzy logical inference is proposed, which uses the information of dependencies between diseases, symptoms, and levels of symptoms. A table of correspondence between the levels of specific symptoms and the disease is presented. Fuzzy modified triangular membership functions were constructed for each linguistic variable of current symptom and disease level. Several ways of setting rules are considered, which are outlined within the framework of two methods. The first method uses a standard approach to building a fuzzy inference system. It is based on the concept of a linguistic variable and a fuzzy rule, which is created on the basis of expert knowledge. Intuitive direct rules are formed, which are presented in mathematical form. Using this rule system, various derivation mechanisms can be applied, such as Mamdani derivation. The second method, in turn, presents an improved fuzzy model that includes additional linguistic variables and rules to quantify the correctly specified symptoms in the set of all symptoms of the patient's disease. This system of rules operates on linguistic variables that use the number of correctly given symptoms relative to the total number of symptoms of the disease. A comparison of the described logical inference systems was made on the test data set. The proposed methods are analyzed for their effectiveness and accuracy in establishing a diagnosis. The results of the study may be important for the development of new diagnostic methods based on fuzzy logic, which may prove to be more effective compared to traditional methods.

Keywords: fuzzy set, fuzzy rules, fuzzy inference systems, expert systems.

Постановка проблеми. Медична діагностика є важливою складовою сучасної медицини, що допомагає у виявленні та визначенні різноманітних хвороб. Однак, існують численні виклики, такі як велика кількість можливих симптомів, а також невизначеність і неясність інформації, отриманої від пацієнтів, що можуть ускладнити процес діагностики. У світлі цього нечітка логіка та системи нечіткого виведення [1][2] виступають як потенційно ефективні інструменти для обробки такої нечіткої інформації, покращуючи точність діагностики.

В рамках цієї статті розглядається застосування нечіткої логіки для діагностики медичних станів на основі симптомів, вказаних пацієнтами. Основна увага приділяється розробці та порівнянню двох різних систем нечіткого виведення, кожна з яких оперує на основі сету лінгвістичних змінних і правил, створених з метою моделювання і аналізу медичних даних.

Перша система базується на стандартному підході до нечіткого виведення, де симптоми і хвороби представлені за допомогою лінгвістичних змінних, і правила формулюються на основі експертних знань. Проте, цей метод має обмеження, зокрема з огляду на «прокляття розмірності» [3] та потенційні проблеми із неповними даними.

Друга система запропонована як варіант вдосконалення, в якому введено нові лінгвістичні змінні та правила, що дозволяють більш точно враховувати кількість правильно вказаних симптомів відносно загальної кількості симптомів хвороби.

Аналіз останніх досліджень і публікацій. Формулювання наукової думки в окресі діагностики стану пацієнта на основі множини нечітких симптомів є різномірним та масштабним. У сучасній науковій площині з'являються роботи присвячені дослідженням сервісів та алгоритмів їх впровадження у сферу охорони здоров'я для підвищення рівня якості діагностики хвороб.

Н. І. Мельникова [4] вирішила важливу науково-прикладну проблему розроблення та удосконалення моделей, методів і засобів машинного навчання в задачах класифікації, кластеризації, прогнозування та візуалізація результатів опрацювання персональних даних для адаптації медичних рішень до пацієнта.

О. О. Василенко [5] розглянув інтелектуальні методи та засоби експертних систем медичної діагностики. Автором досліджено методи застосування потенціалу штучного інтелекту в медицині та проведення діагностики на основі байєсівських мереж. Реалізовані інтелектуальні обчислювальні засоби у вигляді медичної експертної системи на підставі моделей та запропонованих у дослідженні засобів діагностування.

У [6] запропоновано підхід, який дозволяє описувати нечіткі знання, які стосуються як діагностики, так і захворювання. Особливістю запропонованої діагностичної системи є наявність інтелектуального компонента, який забезпечує реалізацію різних функціональностей виключно за рахунок власних можливостей без зміни програмного коду.

Із зарубіжних авторів варто відмітити роботи таких науковців як: Фам Хай, Лун Цу, Кхань Фан, Ха Куок [7], Ду Цюяо, Ван Сяо, Чень Цзюньюй, Ван Іран, Лю Веньлань, Ван Ліпін, Лю Хуейхуей, Цзян Ліся, Не Цзунсю [8], Лекшмі І., Реджіс М. [9] та інших.

Однак незважаючи на масштабність наукових досліджень питання актуальності даної роботи не викликає сумнівів.

Метою даної роботи є дослідження ефективності обох систем через їх порівняння та визначення оптимальної стратегії для використання нечіткої логіки у медичній діагностиці. Дослідження спрямоване на виявлення переваг та недоліків кожної системи, з метою виявлення наявності покращеного методу діагностики хвороб.

Виклад основного матеріалу. Розглянемо задачу визначення хвороби пацієнта відповідно до даних діагностики стану здоров'я і поточних симптомів.

Для вирішення цієї задачі в контексті нечіткої логіки потрібно побудувати систему нечіткого логічного виведення, що використовує інформацію залежностей між хворобами, симптомами, рівнями симптомів.

Для прикладу, нехай кількість хвороб 3, а кількість симптомів 3.

Позначимо множину хвороб D :

$$D = \{Хвороба_1, Хвороба_2, Хвороба_3\}$$

Також введемо множину симптомів S і терм-множину рівнів симптомів L :

$$S = \{Симптом_1, Симптом_2, Симптом_3\}, L = \{Низький, Середній, Високий\}$$

Тоді належність одного чи іншого симптома s з відповідним рівнем l до хвороби d можна визначити за допомогою множини пар R_d :

$$R_d = \{(s, l) \mid s \in S, l \in L\}$$

Аналогічно, дану залежність можна позначити у табличному вигляді, де рядки відповідають за хворобу, а колонки за симптоми:

Табл. 1

Таблиця відповідності рівнів конкретних симптомів до хвороби

Хвороби	Симптом_1	Симптом_2	Симптом_3
Хвороба_1	Середній	Високий	-
Хвороба_2	Низький	Середній	Середній
Хвороба_3	-	Високий	Низький

Додатково введемо терм-множину рівнів визначеності хвороби M :

$$M = \{Малий, Помірний, Великий\}$$

Використовуючи вищезазначену початкову інформацію про предметну область і залежності в ній можна безпосередньо приступати до нечіткого моделювання. Перш за все введемо лінгвістичні змінні та функції належності, для поточних рівнів симптомів і рівнів визначеності хвороби:

{Поточний_Симптом_1, Поточний_Симптом_2, Поточний_Симптом_3} – поточні симптоми.

{Рівень_Хвороба_1, Рівень_Хвороба_2, Рівень_Хвороба_3} – рівні хвороб.

Для кожної лінгвістичної змінної поточного симптома та рівня хвороб побудуємо наступні нечіткі модифіковані трикутні функції належності[10] відповідно:

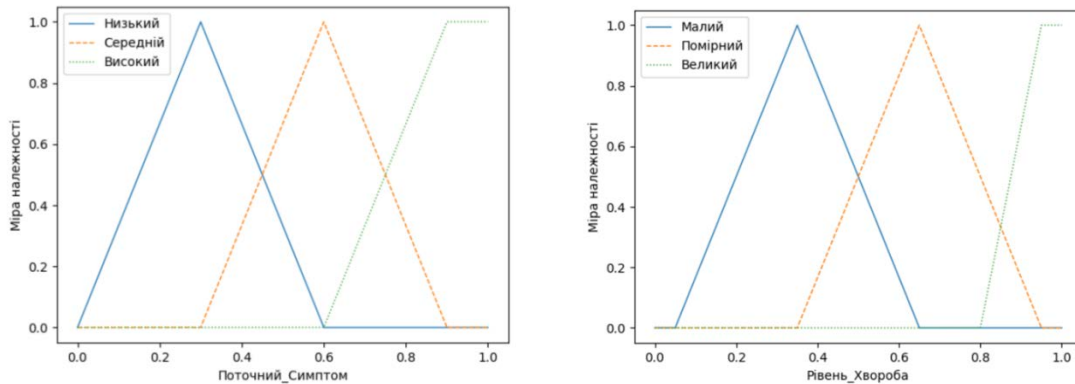


Рис. 1. Функції належності для лінгвістичних змін поточного симптому та рівня хвороби

Аналогічно, такі функції можна задати трійками значень, а саме:

- Поточний_Симптом: Низький – (0, 0.3, 0.6); Середній – (0.3, 0.6, 0.9); Високий – (0.6, 0.9, 0.9)
- Рівень_Хвороба: Малий – (0.05, 0.35, 0.65); Помірний – (0.35, 0.65, 0.95); Великий – (0.65, 0.95, 0.95)

Наступний крок – це побудова нечітких правил, саме від них залежить якість і точність виведення в системах логічного виведення. Ці правила визначають, як система реагує на різноманітні вхідні дані та обставини, тому їх правильне формування є ключовим для отримання бажаних результатів.

Розглянемо декілька способів задання правил:

1) Метод 1

В контексті нашої проблеми визначення хвороби пацієнта за симптомами і початкової інформації про залежність симптомів і хвороб має сенс використати наступні інтуїтивно зрозумілі прями правила за наступним форматом:

1. Якщо (Поточний_Симптом_1 $\in R_1$ (Поточний_Симптом_1)) І (Поточний_Симптом_2 $\in R_1$ (Поточний_Симптом_2)) ... І (Поточний_Симптом_М $\in R_1$ (Поточний_Симптом_К)) ТО Рівень_Хвороба_1 \in Великий;

2. Якщо (Поточний_Симптом_1 $\in R_2$ (Поточний_Симптом_1)) І (Поточний_Симптом_2 $\in R_2$ (Поточний_Симптом_2)) ... І (Поточний_Симптом_М $\in R_2$ (Поточний_Симптом_К)) ТО Рівень_Хвороба_2 \in Великий;

...

N. Якщо (Поточний_Симптом_1 $\in R_N$ (Поточний_Симптом_1)) І (Поточний_Симптом_2 $\in R_N$ (Поточний_Симптом_2)) ... І (Поточний_Симптом_М $\in R_N$ (Поточний_Симптом_К)) ТО Рівень_Хвороба_N \in Великий.

де N – кількість хвороб, K – кількість симптомів, $R_i(x)$ – функція, що визначає рівень симптому $x \in S$ для хвороби $i = 1..N$.

Для наведеного вище прикладу загальна система правил нечіткого виведення матиме вигляд:

1. Якщо (Поточний_Симптом_1 \in Середній) І (Поточний_Симптом_2 \in Високий) ТО Рівень_Хвороба_1 \in Великий;

2. Якщо (Поточний_Симптом_1 \in Низький) І (Поточний_Симптом_2 \in Середній) І (Поточний_Симптом_3 \in Середній) ТО Рівень_Хвороба_2 \in Великий;

3. Якщо (Поточний_Симптом_2 \in Високий) І (Поточний_Симптом_3 \in Низький) ТО Рівень_Хвороба_3 \in Великий.

Використовуючи цю систему правил можна застосувати різні механізми виведення, наприклад, виведення Мамдані [11][12].

Даний підхід простий і зрозумілий, водночас, при такій системі правил можуть виникати, проблеми. Наприклад, нерідко під час опитування пацієнта можна припуститись помилки – пацієнт може надати лише частину симптомів і не надати їх на вхід для обробки програмою.

Оскільки, особливість роботи нечітких правил з оператором "І" під час процесу виведення така, що якщо хоча б одна з вхідних лінгвістичних змінних не відповідає терму в правилі, то результат для цього конкретного правила буде дуже низьким, оскільки мінімальний оператор відображає найнижчий ступінь належності серед змінних. Таку проблему можна спробувати вирішити додатковими, компенсуючими правилами, проте, коли кількість вхідних змін зростає, кількість комбінацій правил може швидко збільшуватися, призводячи до проблеми "прокляття розмірності".

Для підвищення ефективності процесу визначення хвороби за нечіткими правилами пропонується розглянути дещо іншу систему.

2) Метод 2

Дана система правил буде оперувати лінгвістичними змінними, що використовують кількість правильно наданих симптомів відносно загальної кількості симптомів хвороби.

Введемо додаткову терм-множину відносної кількості правильно наданих симптомів C :

C

$= \{ \text{Низька_Відносна_Кількість}, \text{Середня_Відносна_Кількість}, \text{Велика_Відносна_Кількість} \}$

Варто зазначити, що значення лінгвістичних змін $\text{Відносна_Кількість}_i$, $i = 1..N$ обчислюється як кількість наявних симптомів на вході які є i для Хвороби $_i$ поділене на кількість всіх симптомів Хвороби $_i$.

Введемо лінгвістичні змінні та функції належності для відносної кількості коректно наданих симптомів для наданих вище хвороб:

$\{ \text{Відносна_Кількість}_1, \text{Відносна_Кількість}_2, \text{Відносна_Кількість}_3 \}$ – відносні кількості хвороб.

Задамо кожній такій лінгвістичній змінній наступні нечіткі модифіковані трикутні функції належності відповідно до терм-множини:

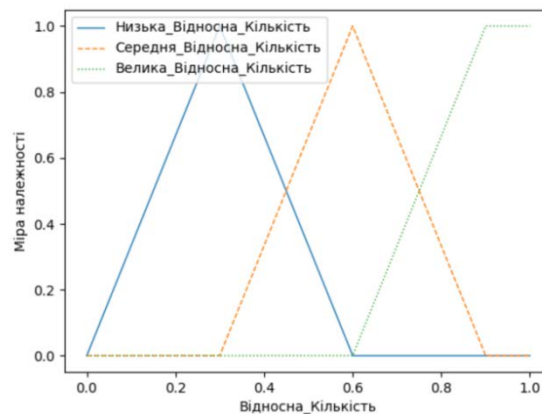


Рис. 2. Функція належності лінгвістичних змін відносної кількості коректно наданих симптомів хвороби

Аналогічно, такі функції змінної $\text{Відносна_Кількість}$ можна задати трійками значень, а саме:

Низька_Відносна_Кількість – (0, 0.3, 0.6);

Середня_Відносна_Кількість – (0.3, 0.6, 0.9);

Висока_Відносна_Кількість – (0.6, 0.9, 0.9).

Можна побудувати нечіткі правила логічного виведення наступним чином:

1.1 Якщо ($\text{Відносна_Кількість}_1$ є Висока_Відносна_Кількість) ТО Рівень_Хвороба_1 є Великий;

1.2 Якщо ($\text{Відносна_Кількість}_1$ є Середня_Відносна_Кількість) ТО Рівень_Хвороба_1 є Помірний;

1.3 Якщо (*Відносна_Кількість_1* є Низька_Відносна_Кількість) ТО *Рівень_Хвороба_1* є Малий;

...

N.1 Якщо (*Відносна_Кількість_N* є Висока_Відносна_Кількість) ТО *Рівень_Хвороба_N* є Великий;

N.2 Якщо (*Відносна_Кількість_N* є Висока_Відносна_Кількість) ТО *Рівень_Хвороба_N* є Помірний;

N.3 Якщо (*Відносна_Кількість_N* є Висока_Відносна_Кількість) ТО *Рівень_Хвороба_N* є Малий.

де N – кількість хвороб, *Відносна_Кількість_i*, $i = 1..N$ відносні кількості коректних симптомів хвороб

Для наведеного вище прикладу загальна система правил нечіткого виведення матиме вигляд:

1.1 Якщо (*Відносна_Кількість_1* є Висока_Відносна_Кількість) ТО *Рівень_Хвороба_1* є Великий;

1.2 Якщо (*Відносна_Кількість_1* є Середня_Відносна_Кількість) ТО *Рівень_Хвороба_1* є Помірний;

1.3 Якщо (*Відносна_Кількість_1* є Низька_Відносна_Кількість) ТО *Рівень_Хвороба_1* є Малий;

2.1 Якщо (*Відносна_Кількість_2* є Висока_Відносна_Кількість) ТО *Рівень_Хвороба_2* є Великий;

2.2 Якщо (*Відносна_Кількість_2* є Середня_Відносна_Кількість) ТО *Рівень_Хвороба_2* є Помірний;

2.3 Якщо (*Відносна_Кількість_2* є Низька_Відносна_Кількість) ТО *Рівень_Хвороба_2* є Малий;

3.1 Якщо (*Відносна_Кількість_3* є Висока_Відносна_Кількість) ТО *Рівень_Хвороба_3* є Великий;

3.2 Якщо (*Відносна_Кількість_3* є Середня_Відносна_Кількість) ТО *Рівень_Хвороба_3* є Помірний;

3.3 Якщо (*Відносна_Кількість_3* є Низька_Відносна_Кількість) ТО *Рівень_Хвороба_3* є Малий;

Порівняння описаних систем логічного виведення

Порівняємо два описаних типи систем з точки зору правильності визначення хвороби в умовах, коли на вхід системи можуть прийти не всі симптоми.

Для тестування скористаємось набором даних [13]. Використовуючи ці дані підготуємо наступні спотворені 11 наборів, кожен з яких містить $P = 2650$ прикладів хвороб і 370 потенційно спотворених значень симптомів для кожного з прикладів хвороби відповідно до таблиці 2.

Оскільки вихід системи нечіткого логічного виведення для кожного прикладу хвороби є набір пар (міра, *Рівень_Хвороба_i*), $i = 1..N$ з потенційно однаковими значеннями міри, введемо функцію γ_j , $j = 1..P$ для обчислення дійсного значення правильно визначеного j прикладу хвороби.

$$\gamma_j = \begin{cases} 0.0, & \text{якщо хвороби } j \text{ немає серед пар з максимальною мірою} \\ \frac{1.0}{T}, & \text{де } T \text{ кількість пар з однаковою максимальною мірою} \end{cases}$$

Тоді коректність роботи системи σ на тому чи іншому наборі даних будемо визначати як суму правильно визначених прикладів хвороб γ_j відносно загальної кількості прикладів хвороб у наборі, враховуючи можливість отримання однакової міри під час роботи програми.

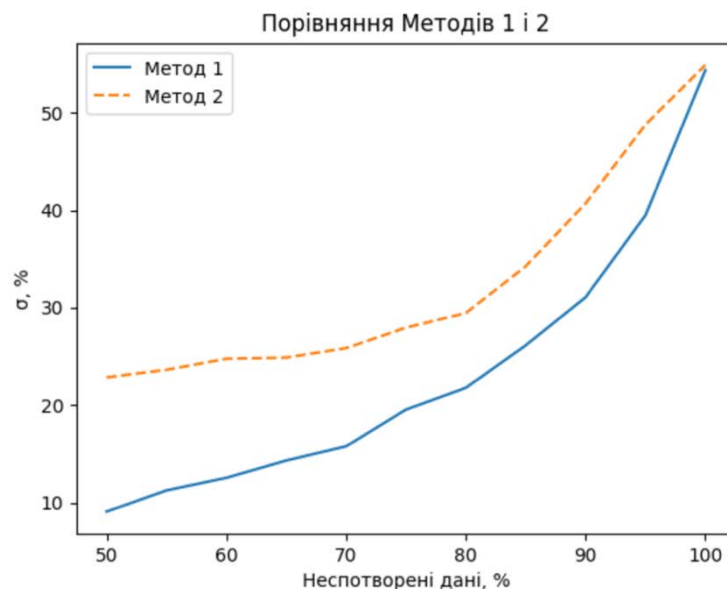
$$\sigma = \frac{\sum_{j=1}^P \gamma_j}{P}$$

Табл. 2.

Опис наборів тестових даних і результатів визначення хвороби кожним типом системи

Набір	% неспотверених симптомів	σ 1 метод, %	σ 2 метод, %	$\Delta\sigma$, %
data_50	50	9.06	22.84	13.78
data_55	55	11.24	23.62	12.38
data_60	60	12.53	24.76	12.23
data_65	65	14.31	24.87	10.56
data_70	70	15.76	25.84	10.08
data_75	75	19.53	27.95	8.42
data_80	80	21.76	29.44	7.68
data_85	85	26.13	34.24	8.11
data_90	90	31.96	40.70	8.74
data_95	95	39.48	48.78	9.3
data_100	100	54.36	54.90	0.54

Додатково побудуємо графіки відповідно до даних з таблиці:

Рис. 3. Порівня графіків σ для Метод 1 і Метод 2

Висновки. У даній статті було представлено та досліджено дві системи нечіткого виведення для діагностування медичних станів на основі симптомів, поданих пацієнтами. Показано, що ці системи можуть служити як фундамент для розробки ефективних методів діагностики у медицині, особливо у випадках, коли не всі симптоми хвороби є наявними чи точними.

Обидві розглянуті системи нечіткого виведення показали свою ефективність адекватно та гнучко обробляти нечітку інформацію, що подавалась на вхід та знаходити правильні діагнози на основі наявних даних. Вдосконалена система, що включає нові лінгвістичні змінні та правила, показала себе дещо ефективніше, ніж традиційна, особливо при наявності обмеженої чи спотвореної інформації. Заснований на аналізі двох моделей нечіткого виведення досвід вказує на значний потенціал для подальшого вдосконалення та розвитку систем діагностування, заснованих на нечіткій логіці, для більш точного та надійного визначення хвороб.

Розроблені методики можуть внести вагомий вклад у поліпшення медичної діагностики, дозволяючи лікарям та медичним експертам здійснювати більш точні та обґрунтовані висновки на основі неповних чи нечітких даних.

Список літератури:

1. Zadeh L.A. (1965) Fuzzy Sets // Information and Control. – Vol.8.
2. Provotar A.I., Lapko A.V., Provotar A.A. (2013) Fuzzy inference systems and their applications. Int. Sci. J. Cybern. Syst. Anal. 49(4), pp. 517–525.

3. Keogh E., Mueen A. (2017) Curse of Dimensionality. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA. pp. 314–315.
4. Мельникова Н.І. Моделі та методи підтримки персоналізованих рішень у медичних системах. – На правах рукопису. Дисертаційна робота на здобуття наукового ступеня доктора технічних наук за спеціальністю 05.13.23 – системи штучного інтелекту. – Національний університет «Львівська політехніка» Міністерства освіти і науки України, Львів, 2023. 293 с.
5. Василенко О. О. Інтелектуальні методи та засоби експертних систем медичної діагностики. РИ, 2018. №4. С. 86-94.
6. Саввакін В. Д., Провотар О. І. Діагностична система на основі нечітких знань / Компьютерная математика. 2019. № 1. С.56-63.
7. Pham Hai, Long Cu, Khanh Phan, Ha Quoc. (2023). A Fuzzy Knowledge Graph Pairs-Based Application for Classification in Decision Making: Case Study of Preeclampsia Signs. Information. 14. 104. 10.3390/info14020104.
8. Du Qiuyao, Wang Xiao, Chen Junyu, Wang Yiran, Liu Wenlan, Wang Liping, Liu Huihui, Jiang Lixia, Nie Zongxiu. (2023). Machine learning encodes urine and serum metabolic patterns for autoimmune disease discrimination, classification and metabolic dysregulation analysis. The Analyst. 148. 10.1039/d3an01051a.
9. Lekshmi I., Regees M. (2022). review on fuzzy multi-criteria decision-making methods in medical diagnosis. International journal of health sciences. 2546-2554. 10.53730/ijhs.v6nS5.9198.
10. Alemu, Matiwos. (2018) A fuzzy model for chaotic time series prediction. International Journal of Innovative Computing, Information and Control. 14. pp. 1767-1786.
11. Mamdani, E. H. (1974) Application of fuzzy algorithms for the control of a simple dynamic plant. In Proc IEEE, pp. 121–159.
12. M.M. Gupta, J. Qi (1991) Theory of T-norms and fuzzy inference methods, Fuzzy Sets and Systems, Volume 40, Issue 3, pp. 431-450.
13. Boger, C.M. (2002) A Synoptic Key of the Materia Medica.