

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)

Кафедра Інформаційних управляючих систем
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів виявлення аномальних ситуацій у бізнес-процесах
медичних закладів

(тема)

Виконав:

здобувач 2 року навчання,
групи ІУСТм-23-1

Мартиненко Олексій Володимирович
(прізвище, ім'я, по батькові)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Інформаційні управляючі системи та технології
(повна назва освітньої програми)

Керівник: проф. каф. ІУС Чалий С.Ф.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ІУС


(підпис)


Костянтин ПЕТРОВ
(власне ім'я, прізвище)

2025 р.

Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наукКафедра Інформаційних управляючих системРівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки
(код і повна назва)Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)Освітня програма Інформаційні управляючі системи та технології
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри 
(підпис)

“ 09 ” грудня 2024 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**здобувачеві Мартиненку Олексію Володимировичу
(прізвище, ім'я, по батькові)1. Тема роботи Дослідження методів виявлення аномальних ситуацій у бізнес-процесах медичних закладів

затверджена наказом по університету від “ 27 ” листопада 2024 р. № 1249Ст


2. Термін подання здобувачем роботи до екзаменаційної комісії “ 18 ” січня 2025 р.


3. Вихідні дані до роботи Науково-технічна література, публікації та інтернет-ресурси, що стосуються теми кваліфікаційної роботи; матеріали звіту з практики4. Перелік питань, що потрібно опрацювати у роботі Аналіз медичних бізнес-процесів та методів виявлення аномальних ситуацій; виявлення переваг та недоліків цих методів; удосконалення підходу до виявлення аномальних ситуацій на основі процесного аналізу з використанням систем нечіткого виведення; розробка інформаційної технології з урахуванням удосконаленого методу; програмна реалізація та експериментальна перевірка отриманого методу

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Видача завдання на кваліфікаційну роботу	09.12.2024	Виконано
2	Проведення аналізу предметної області	09.12.2024-14.12.2024	Виконано
3	Формування постановки задачі	15.12.2024	Виконано
4	Розробка удосконаленого методу виявлення	16.12.2024-22.12.2024	Виконано
5	Розробка інформаційної технології з використанням	23.12.2024-29.12.2024	Виконано
6	Практична реалізація удосконаленого методу	30.12.2024-02.01.2025	Виконано
7	Експериментальна перевірка удосконаленого	03.01.2025-05.01.2025	Виконано
8	Оформлення пояснювальної записки	06.01.2025-15.01.2025	Виконано
9	Захист кваліфікаційної роботи	21.01.2025	Виконано

Дата видачі завдання 09 грудня 2024 р.

Здобувач 
(підпис)

Керівник роботи 
(підпис)

проф. каф. ІУС Сергій ЧАЛИЙ
(посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 77 с., 16 рис., 7 табл., 1 дод., 30 джерел.

АНОМАЛІЯ, БІЗНЕС-ПРОЦЕС, ВІДХИЛЕННЯ, ГЕНЕТИЧНИЙ АЛГОРИТМ, ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ, МЕДИЧНИЙ ЗАКЛАД, ПРОЦЕСНИЙ АНАЛІЗ, СИСТЕМА НЕЧІТКОГО ВИВЕДЕННЯ

Об'єктом дослідження кваліфікаційної роботи є моделі медичних бізнес-процесів.

Предмет дослідження: методи виявлення аномальних ситуацій у медичних бізнес-процесах.

Мета роботи: дослідження методів виявлення аномальних ситуацій у медичних бізнес-процесах з урахуванням як послідовності процесів, так і їхніх даних, для підвищення якості лікування пацієнтів.

Результатом дослідження є опис удосконаленого методу виявлення аномальних ситуацій у медичних бізнес-процесах на основі процесного аналізу та систем нечіткого виведення. Експериментальна перевірка практичної реалізації довела підвищення точності виявлення аномальних ситуацій з використанням удосконаленого методу.

Наукова новизна дослідження полягає в удосконаленні методу виявлення аномальних ситуацій у медичних бізнес-процесах шляхом аналізу логу, формування нечітких причинно-наслідкових правил для дій бізнес-процесу та виявлення аномалій з використанням цих правил.

ABSTRACT

Mater's thesis: 77 pages, 16 figures, 7 tables, 1 appendix, 30 sources.

ANOMALY, BUSINESS-PROCESS, OUTLIER, GENETIC ALGORITHM, INFORMATION TECHNOLOGY, HEALTHCARE FACILITY, PROCESS MINING, FUZZY INFERENCE SYSTEM.

The object of the research of the qualification work is healthcare business-processes.

Subject of the research: methods for detecting anomalies in healthcare business-processes.

Objective of the research: to investigate methods for detecting anomalies in medical business-processes, taking into account both the sequence of processes and their data, to improve the quality of patient care.

The result of the research is a description of an improved method for detecting anomalies in healthcare business-processes based on process mining and fuzzy inference systems. Experimental verification of the practical implementation has demonstrated increased accuracy in detecting anomalies using the improved method.

The scientific novelty of the research lies in the enhancement of a method for detecting anomalies in healthcare business-processes through log analysis, the formulation of fuzzy cause-and-effect rules for business-process actions, and the identification of anomalies using these rules.

ЗМІСТ

Скорочення та умовні позначки	7
Вступ.....	8
1 Аналіз предметної області та постановка задачі	10
1.1 Аналіз типів та функцій медичних закладів	10
1.2 Аналіз моделей бізнес-процесів медичних закладів	14
1.3 Аналіз методів виявлення аномальних ситуацій	19
1.4 Постановка задачі дослідження.....	25
2 Розробка методу виявлення аномальних ситуацій у медичних бізнес-процесах з використанням методології Process Mining	27
2.1 Розробка підходу до виявлення аномальних ситуацій у бізнес-процесах медичних закладів.....	27
2.2 Удосконалення методу виявлення аномальних ситуацій на основі процесного аналізу бізнес-процесів медичних закладів	31
3 Розробка інформаційної технології виявлення аномальних ситуацій у медичних бізнес-процесах	44
4 Програмна реалізація та експериментальна перевірка удосконаленого методу виявлення аномальних ситуацій.....	49
4.1 Програмна реалізація модулю виявлення аномалій з використанням системи нечіткого виведення	49
4.2 Експериментальна перевірка удосконаленого методу виявлення аномальних ситуацій.....	54
Висновки.....	61
Перелік джерел посилання	62
Додаток А Графічний матеріал кваліфікаційної роботи.....	66

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ІС – інформаційна система

ІТ – інформаційна технологія

МОЗ – Міністерство охорони здоров'я

ПЗ – програмне забезпечення

BPM – Business Process Management, методологія управління бізнес-процесами

CBLOF – Cluster-Based Local Outlier Factor, метод виявлення відхилень на основі методу Local Outlier Factor та кластеризації

COG – Centre of Gravity, метод центру тяжіння

COPOD – Copula-Based Outlier Detection, метод виявлення відхилень на основі копул

ECOD – Empirical-Cumulative-distribution-based Outlier Detection, метод виявлення відхилень на основі емпіричних функцій кумулятивного розподілу

LODA – Lightweight Online Detector of Anomalies, метод виявлення аномалій на основі проекції багатовимірних даних на одновимірний простір

LOF – Local Outlier Factor, метод виявлення відхилень на основі оцінки щільності сусідніх областей даних

MO_GAAL – Multiple-Objective Generative Adversarial Active Learning, метод виявлення аномалій на основі генеративних змагальних мереж, що використовує багатоцільові функції

OSCVN – One-Class Support Vector Machine, метод виявлення аномалій на основі методу опорних векторів

SO_GAAL – Simple-Objective Generative Adversarial Active Learning, метод виявлення аномалій на основі генеративних змагальних мереж, що використовує одну цільову функцію

ВСТУП

Сфера охорони здоров'я усіх розвинених країн світу стає основним об'єктом занепокоєння у політичній, соціальній і особливо економічній складових усього сучасного суспільства. Це породжує необхідність постійного розвитку та поліпшення медичних бізнес-процесів в усіх їхніх аспектах. До них, зокрема, включаються як клінічні процеси, так і адміністративні. Проте ключовими серед них є надання медичних послуг пацієнтам – клієнтам медичних закладів. Цей бізнес-процес є головним для одного з типів медичних закладів, а саме для лікувально-профілактичних: госпіталів, лікарень, амбулаторій, поліклінік, диспансерів та санаторіїв.

Узагальнена послідовність дій в бізнес-процесі лікування пацієнта складається з реєстрації пацієнта, його первинного огляду, визначення плану лікування, проходження лікування, повторного огляду та виписки пацієнта за умови завершення лікування, або ж циклічне продовження попередніх етапів. Крім того, обидва етапи огляду та проходження лікування можуть поділятися на свої власні підпроцеси, зокрема такі як проходження фізичних тестів, аналізів, виконання хірургічних операцій тощо. Кожний з них у якомусь вигляді представлено в інформаційних системах медичних закладів задля автоматизації управління та моніторингу. Завдяки їх наявності поліпшення медичних бізнес-процесів відбувається з використанням відповідних підходів та методів, що впроваджуються на рівні апаратного та програмного забезпечення інформаційних систем.

Одним зі способів покращення процесів лікування пацієнтів у таких системах є постійний їх моніторинг та виявлення відхилень від заданих процедур та стандартів – виявлення аномальних ситуацій. До них можуть належати як відхилення в послідовності дій, що виконуються у межах бізнес-процесу, так і відхилення у даних, що були використані відповідним бізнес-процесом. Однак існуючі методи виявлення аномалій зосереджуються окремо або на

послідовностей дій процесу, або на його даних, що породжує відповідну проблему інтеграції алгоритмів виявлення аномальних ситуацій в єдину систему моніторингу бізнес-процесів.

Метою роботи є дослідження методів виявлення аномальних ситуацій у медичних бізнес-процесах з урахуванням як послідовності процесів, так і їхніх даних, задля підвищення якості лікування пацієнтів.

Відповідними задачами дослідження є: аналіз медичних бізнес-процесів та методів виявлення аномальних ситуацій; розробка методу, що поєднуватиме обидва аспекти медичних бізнес-процесів; програмна реалізація запропонованого методу та його експериментальна перевірка.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Аналіз типів та функцій медичних закладів

Відповідно до наказу Міністерства охорони здоров'я (МОЗ) України [1] медичні заклади, або ж заклади охорони здоров'я, з урахування напряму їхньої діяльності загалом можна класифікувати за наступними трьома основними категоріями:

- лікувально-профілактичні: основними завдання закладів є надання медичної допомоги пацієнтам (клієнтам) та профілактика захворювань;
- санітарно-профілактичні: метою закладів є запобігання захворювань та підтримка здорових умов життя через контроль за санітарно-гігієнічними умовами, проведення профілактичних заходів та моніторинг стану довкілля;
- фармацевтичні (аптечні): заклади зберігання, розподілу та реалізації лікарських засобів.

Окрім загального розподілу на три окремі класи, лікувально-профілактичні та санітарно-профілактичні медичні заклади також поділяються на підкатегорії, які вже більш предметно визначають розподіл за типами діяльності.

Зокрема, лікувально профілактичні заклади поділяються на:

- лікарняні (багатопрофільні, однопрофільні, спеціалізовані та особливого типу): госпіталі, лікарні, клініки, диспансери;
- амбулаторно-поліклінічні: амбулаторії, поліклініки, діагностичні центри;
- заклади переливання крові, швидкої та екстреної медичної допомоги: станції переливання крові та центри екстреної медичної допомоги;
- санаторно-курортні: оздоровчі центри, курортні поліклініки, санаторії.

Санітарно-профілактичні поділяються на:

- санітарно-епідеміологічні: дезінфекційні та санітарно-епідеміологічні станції;
- заклади санітарної просвіти: центри здоров'я;

– заклади громадського здоров'я: лабораторні центри, центри громадського здоров'я, центри контролю та профілактики хвороб.

Відповідно, можна виокремити наступні основні функції, що виконуються медичними закладами кожного з попередньо визначених класів:

а) функції лікувально-профілактичних закладів:

- 1) діагностика та лікування захворювань;
- 2) проведення профілактичних оглядів;
- 3) вакцинація;
- 4) реабілітація після лікування;

б) функції санітарно-профілактичних закладів:

- 1) контроль за якістю довкілля;
- 2) виявлення та ліквідація джерел інфекційних захворювань;
- 3) проведення епідеміологічного нагляду;
- 4) розробка і впровадження програм санітарної освіти;

в) функції фармацевтичних закладів:

- 1) забезпечення населення лікарськими засобами;
- 2) консультація пацієнтів щодо застосування ліків;
- 3) контроль за якістю та зберіганням медикаментів.

У таблиці 1.1 наведено зведену характеристику попередньо визначених типів та функцій медичних закладів.

Таблиця 1.1 – Зведена характеристика типів та функцій медичних закладів

Тип закладу	Підтип	Медичний заклад	Функція медичного закладу
1	2	3	4
Лікувально-профілактичний	Лікарняний	Госпіталь	Діагностика та лікування військовослужбовців та ветеранів війни
		Лікарня	Діагностика та лікування пацієнтів

Продовження таблиці 1.1

1	2	3	4
Лікувально-профілактичний	Лікарняний	Клініка	Надання медичної допомоги та здійснення навчальної і наукової роботи
		Диспансер	Надання медичної допомоги для певних груп захворювань із забезпеченням диспансеризації пацієнтів
	Амбулаторно-поліклінічний	Амбулаторія	Надання медичної допомоги з однієї або декількох лікарських спеціальностей
		Поліклініка	Надання медичної допомоги з багатьох лікарських спеціальностей
		Діагностичний центр	Діагностика захворювань пацієнтів
	Заклад переливання крові, швидкої та екстреної медичної допомоги	Станція переливання крові	Проведення операцій з переливання крові
	Заклад переливання крові, швидкої та екстреної медичної допомоги	Центр екстреної медичної допомоги	Надання швидкої медичної допомоги
	Санаторно-курортний	Оздоровчий центр	Збереження та покращення загального стану здоров'я
		Курортна поліклініка	Амбулаторна допомога пацієнтам, що перебувають на лікуванні або відпочинку
		Санаторій	Лікування та профілактика хронічних захворювань

Кінець таблиці 1.1

1	2	3	4
Санітарно-профілактичний	Санітарно-епідеміологічний	Дезінфекційна станція	Евакуація та дезінфекція підвідомчих територій
		Санітарно-епідеміологічна станція	Здійснення санітарного нагляду на підвідомчій території
	Заклад санітарної просвіти	Центр здоров'я	Здійснення просвітницької діяльності
	Заклад громадського здоров'я	Лабораторний центр	Здійснення лабораторних досліджень
		Центр громадського здоров'я	Збереження та зміцнення здоров'я окремих громад
		Центр контролю та профілактики хвороб	Контроль та профілактика хвороб населення
Фармацевтичний (аптечний)	-	Аптека	Розповсюдження медичного обладнання та ліків
		База (склад)	Зберігання медичного обладнання та ліків
		Лабораторія	Дослідження та контроль якості медичного обладнання та ліків

Відповідно, кожний з визначених типів медичних закладів має певні особливості у функціях, що вони виконують, а тому моделі їхніх бізнес-процесів адаптуються в залежності від мети функціонування кожного окремого типу закладу, хоча загалом мають подібну структуру у межах одного типу.

1.2 Аналіз моделей бізнес-процесів медичних закладів

При розгляді найбільш розповсюдженого типу медичних закладів, а саме лікувально-профілактичного, можна виділити два окремі класи їхніх бізнес-процесів [2]:

- клінічні, або лікувальні: ключові процеси, пов'язані з пацієнтами та базуються на діагностично-терапевтичних циклах;
- адміністративні: усі процеси, що підтримують клінічну діяльність медичних закладів.

Основними особливостями обох груп бізнес-процесів, кожна з яких є нетривіальною за своєю складністю побудови, є наступні характеристики [3]:

- висока динамічність: часті зміни через появи технології, медичні відкриття чи адміністративні вимоги;
- висока складність: нетривіальний процес винесення медичних рішень, велика кількість даних, непередбачуваність реакцій пацієнтів;
- багатопрофільність: велика кількість відділів різних профілів із медичної сфери;
- варіативність: висока залежність від знань медичного персоналу та стану здоров'я пацієнтів, що може спричиняти відхилення від звичного процесу роботи з ними.

Для формування детальної характеристики бізнес-процесів медичних закладів, зокрема задля визначення їхніх моделей, було виділено відповідні ключові клінічні та адміністративні бізнес-процеси:

а) клінічні процеси:

- 1) діагностика та лікування;
- 2) хірургічні процедури;
- 3) фармацевтичні послуги;
- 4) моніторинг пацієнтів;
- 5) реабілітаційні сервіси;

б) адміністративні процеси:

- 1) управління медичними закладами;
- 2) управління персоналом;
- 3) управління фінансами;
- 4) обслуговування пацієнтів;
- 5) управління медичних обладнанням.

У таблиці 1.2 наведено сформовану характеристику визначених ключових бізнес-процесів лікувально-профілактичних медичних закладів.

Таблиця 1.2 – Характеристика ключових бізнес-процесів лікувально-профілактичних медичних закладів

Бізнес-процес	Мета	Вхідні дані	Вихідні дані	Зацікавлені сторони
1	2	3	4	5
Діагностика та лікування	Діагностика захворювань пацієнтів та їхнє лікування	Медична історія, симптоми, результати аналізів	Діагноз, план лікування, рецепти, рекомендації	Пацієнт, доктор, допоміжний медичний персонал
Хірургічні процедури	Проведення операцій у процесі лікування	Медична історія, перед-операційні приписання	Пост-операційні приписання	Пацієнт, хірург, допоміжний медичний персонал
Фармацевтичні послуги	Надання послуг з купівлі фармацевтичних препаратів	Рецепти, інвентар препаратів	Видані препарати, інструкція користування	Пацієнт, фармацевт, доктор
Моніторинг пацієнтів	Відстеження стану пацієнтів	Дані стану пацієнтів	Повідомлення про аномальні ситуації, корегування планів лікування	Пацієнт, доктор, медичний персонал
Реабілітаційні сервіси	Відновлення пацієнтів після травм та операцій	Медичні записи	План реабілітації, звіти з прогресу	Пацієнт, доктор, терапевт

Кінець таблиці 1.2

1	2	3	4	5
Управління медичними закладами	Дотримання регуляторних норм у медичних закладах	Розклад обслуговування закладів, протоколи очищення, регуляторні норми	Звіти про дотримання регуляторних норм	Менеджери медичних закладів, прибиральники
Управління персоналом	Підтримання працездатності медичного закладу	Дані працівників	Робочий розклад, дані продуктивності персоналу, відгуки пацієнтів	Працівники HR відділу, менеджери, медичний персонал
Управління фінансами	Підвищення доходу та оптимізація витрат	Дані про доходи, дані про витрати	Квартальна звітність, аналітичні відомості	Працівники бухгалтерського відділу, аналітики
Обслуговування пацієнтів	Обслуговування пацієнтів під час прийомів	Контактні дані пацієнтів, звернення про допомогу	Розклад лікарських прийомів, запис на прийом	Реєстраційний персонал, пацієнт, доктор
Управління медичним обладнанням	Управління медичним інвентарем медичних закладів	Запити на медичне обладнання, препарати, поточний інвентар	Замовлення обладнання, препаратів, правила зберігання	Працівники складу, постачальники і обладнання

З урахуванням попередньо визначених типів бізнес-процесів було також сформовано узагальнене представлення бізнес-процесу медичних закладів (див. рис. 1.1), яка у подальшому стане цільовою під час розробки відповідного підходу до виявлення аномальних ситуацій.



Рисунок 1.1 – Узагальнене представлення бізнес-процесів медичних закладів

Однією з проблем, що виникають під час виконання наведених вище бізнес-процесів медичних закладів, є визначення та застосування ефективного підходу до управління ними.

Загалом прийнято виділяти три підходи до управління бізнес-процесами підприємств [4]:

- процесний;
- системний;
- ситуативний.

Процесний підхід розглядає управління як суму взаємопов'язаних функцій, а тому спрямований на покращення інформаційних потоків між бізнес-процесами задля пришвидшення виконання кожного з них, що допомагає задовольнити потреби клієнтів.

У рамках системного підходу підприємство розглядається як система взаємозв'язаних елементів, що завжди впливають одне на одного, а тому формують єдину систему, яка має вхід (мету діяльності), вихід (результат діяльності), внутрішні і зовнішні зв'язки та незалежні фактори впливу. У такому підході ефективність підприємства залежить від ефективності кожного окремого бізнес-процесу, а не лише окремих найрезультативніших його складових.

Ситуативний підхід зосереджується на аналізі конкретних ситуацій, що виникають у процесі діяльності підприємства, та знаходженні оптимальних рішень їхніх проблем. Такий підхід є найбільш ефективним для великих підприємств, що мають величезну кількість бізнес-процесів, загальне управління якими виявляється досить складним завданням.

На рисунку 1.2 відображено відповідну характеристику кожного з підходів.

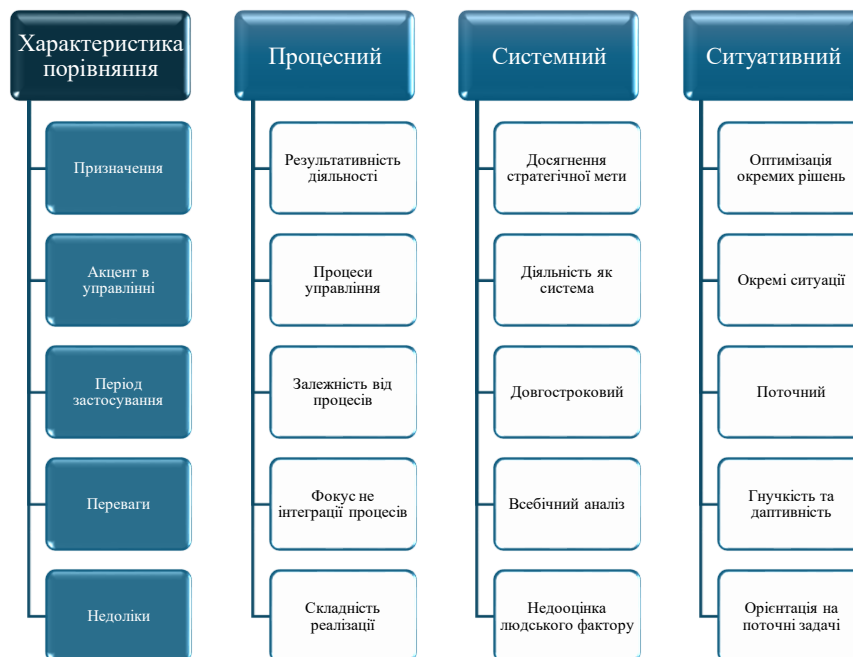


Рисунок 1.2 – Порівняння підходів до управління бізнес-процесами

Окремим підходом до управління бізнес-процесами підприємств можна також виділити гібридний, який поєднує у собі аспекти двох або більше інших підходів. Зокрема, добре поєднуються системний та ситуативний підходи,

завдяки чому бізнес-процеси підприємства на загальному рівні розглядаються як суцільна система взаємозв'язаних елементів, та водночас вони можуть розглядатись окремо у випадках виникнення ситуативних проблем.

Враховуючи особливості бізнес-процесів медичних закладів, найбільш адекватним підходом до управління ними можна вважати саме гібридний: поєднання системного та ситуативного підходів із залученням особливостей кожного з них у процес управління. Завдяки цьому підходу можливі одночасна реалізація стратегічного планування діяльності медичних закладів та розвитку кожного з їхніх ключових бізнес-процесів, а також швидке реагування на проблеми, що можуть ситуативно виникати у них. До таких проблем, зокрема, відносяться і аномальні ситуації будь-якого характеру відповідно до характеристик та особливостей кожного окремого бізнес-процесу відповідного медичного закладу [5].

Також завдяки системному підходу у рамках методології управління бізнес процесами (Business Process Management – BPM), що спрямована на безперервне покращення процесів медичних закладів з використанням інформаційних технологій (ІТ), досягається найбільш точне представлення усіх бізнес-процесів як єдиною системи, що спрощує подальше перенесення її відповідних елементів у відповідні інформаційні системи [6].

1.3 Аналіз методів виявлення аномальних ситуацій

З реалізацією системного підходу до управління бізнес-процесами медичних закладів з використанням інформаційних технологій для побудови відповідних інформаційних систем (ІС) з'являється певна низка можливостей щодо цільового покращення ефективності ведення діяльності закладів охорони здоров'я.

Зокрема, завдяки впровадженню ІС становиться можливим активне

виявлення аномальних ситуацій – відхилення у певних наборах даних, що використовуються у бізнес-процесах або є їхніми кінцевими результатами. Такі аномальні ситуації можуть бути як і невеликими помилками, що виникають через незначні збої у системі або людський фактор, або ж доволі значними відхиленнями від норми, що впливають як на загальну працездатність медичних закладів, так і на критично важливі клінічні процеси, пов'язані з обслуговуванням клієнтів – пацієнтів.

Загалом можна виділити п'ять типів аномальних ситуацій [7]:

- точкові: окремі точкові дані, що значно відрізняються від усього набору даних; такі типи аномалій можуть виникати через випадкові вимірвальні помилки, помилки введення даних;

- контекстуальні: такі аномальні ситуації вважаються аномальними у рамках якогось певного контексту, у той час як в усіх інших випадках вони є нормальними;

- колективні: відхилення цілих наборів даних від загального; такі аномалії можуть виникати через збої у системі;

- просторові: аномальні ситуації, що пов'язані з географічними даними;

- темпоральні: відхилення даних від очікуваних шаблонів, що базуються на основі часових рядів.

У рамках діяльності медичних закладів з урахуванням особливостей їхніх бізнес-процесів можуть виникати наступні типи аномальних ситуацій: точкові, колективні та темпоральні.

Два інших типи: контекстуальні та просторові, – не відповідають характеристикам розглянутих бізнес-процесів. Зокрема, точність даних нівелює їхнє різне трактування в залежності від контексту, що унеможлиблює появу контекстуальних аномалій, а географічні дані є зазвичай нетиповими для цієї предметної області.

Методи виявлення аномальних ситуацій можна розділити на певні групи у відповідності до їхнього механізму роботи [8]:

- імовірнісні: базуються на використанні статистичних моделей для

визначення наборів даних, що відхиляються від звичних показників; такі методи припускають, що нормальна поведінка системи може бути представлена у вигляді імовірнісного розподілу, у якому аномальні ситуації вважаються рідкісними;

- лінійні моделі: ці методи, власне, і базуються на лінійних моделях, які припускають, що дані можуть бути представлені у вигляді лінійних залежностей між вхідними змінними;

- методи на основі наближення (Proximity Based): базуються на визначені відстаней між даними, де аномальні ситуації розташовуються якнайдалі від основного набору;

- групові методи (Outlier Ensembles): комбінація декількох методів виявлення аномальних ситуацій, що дозволяє покращити точність процесу виявлення аномалій;

- нейронні мережі: нейронні мережі у методах виявлення аномальних ситуацій застосовуються для виявлення даних, що мають високий рівень помилок під час роботи нейронних мереж або ж взагалі не підходять до натренованої моделі.

Існує велика кількість методів виявлення аномальних ситуацій, що належать до вище зазначених типів. Зокрема, доволі відомими методами є наступні:

- Empirical-Cumulative-distribution-based Outlier Detection (ECOD) [9]: заснований на використанні емпіричних функцій кумулятивного розподілу даних, які для кожної ознаки визначають чи виходить її значення за межі допустимого;

- Copula-Based Outlier Detection (COPOD) [10]: для знаходження відхилень у множині даних використовує копули – функції, що зв'язують багатовимірний розподіл випадкових величин з їх одновимірним представленням;

- One-Class Support Vector Machine (OCSVM) [11]: призначений для побудови певної поверхні у просторі даних, що охоплюватиме лише нормальні дані, а усі інші вважатимуться аномальними;

- Local Outlier Factory (LOF) [12]: оцінює щільність розподілу сусідніх областей на просторі даних з припущенням, що щільність нормальних даних завжди більше, ніж в аномальних;
- Cluster-Based Local Outlier Factor (CBLOF) [13]: використовує такий самий підхід, що і LOF, але виявлення аномалій відбувається не на всьому просторі даних, а окремо в кожному попередньо визначеному кластері;
- Lightweight Online Detector of Anomalies (LODA) [14]: використовує підхід зниження розмірності вхідних даних задля аналізу на одновимірному просторі замість багатовимірною, що знижує обчислювальну складність методу;
- Isolation Forest [15]: з використанням випадково згенерованих порогів рекурсивно перевіряє кількість кроків, що необхідна для досягнення певної точки у просторі даних, чим менше кроків – тим більша умовірність того, що точка є аномалією;
- Single-Objective Generative Adversarial Active Learning (SO_GAAL) [16]: базується на використанні генеративних змагальних мереж з одноцільовою функцією;
- Single-Objective Generative Adversarial Active Learning (MO_GAAL) [16]: також використовує генеративні змагальні мережі, проте застосовує багатоцільові функції, що дозволяє одночасно визначати аномалії за багатьма ознаками.

На рисунку 1.3 наведено вище описані методи виявлення аномальних ситуацій з групуванням за відповідними типами.

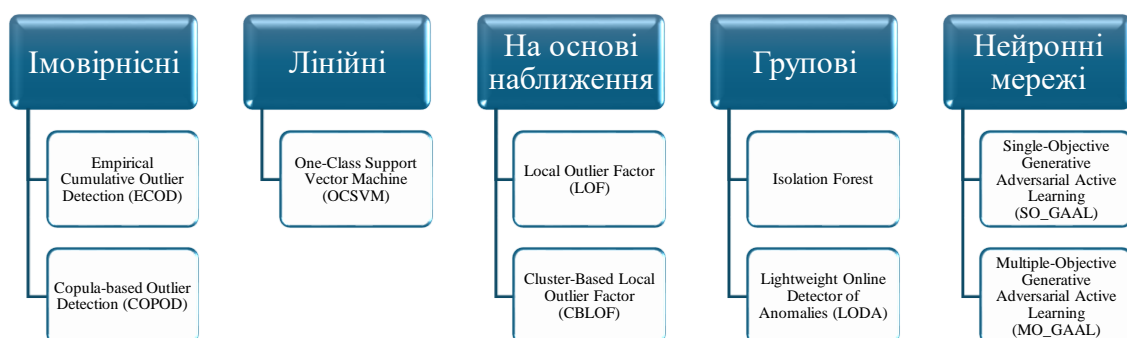


Рисунок 1.3 – Приклади методів виявлення аномальних ситуацій за типами

Додатково було проведено детальний аналіз одного з представлених вище методів, а саме методу ECOD. Цей метод виявлення аномальних ситуацій, або ж відхилень (outliers), заснований на використанні емпіричних кумулятивних функцій розподілу. Як зазначалось раніше, цей метод відноситься до статистичних, бо функції розподілу, що використовуються ним, базуються саме на відомих наборах даних з використанням їхніх статистик.

Відповідні емпіричні кумулятивні функції розподілу відображають накопичену ймовірність для кожного з вимірів даних у багатовимірних наборах та визначають ті значення у наборі даних, що значно відхиляються у меншу або у більшу сторону від основної «нормальної» вибірки даних, що досліджується.

Далі функції використовуються для розрахунку окремих ймовірностей для кожного хвостового (tail) – правого та лівого, – значення по відношенню до окремих точок у наборі даних. Таким чином визначається ступінь «екстремальності» відхилення даних по відношенню до середнього їх значення в окремому вимірі.

Останнім кроком виступає агрегація отриманих значень ймовірностей хвостових значень для отримання загальної оцінки того, що визначені точки з набору даних є аномаліями, або ж відхиленнями. Відповідна оцінка тим більша, чим більша ймовірність того, що обрана точка є аномалією.

Відповідно, алгоритм цього методу складається з трьох основних кроків:

- визначення функцій розподілу для кожного з вимірів даних;
- розрахунок ймовірностей хвостових значень з використанням функцій;
- агрегування отриманих ймовірностей та отримання оцінки приналежності окремих точок до аномалій (викидів).

Перевагами цього методу виявлення аномальних ситуацій є:

- відсутність параметрів: спрощує можливість застосування цього методу на різноманітних наборах даних завдяки відсутності необхідності налаштовувати специфічні для предметної області параметри;
- пристосованість до багатовимірних даних: завдяки розрахунку функцій для окремих вимірів даних цей метод може застосовуватись для наборів з будь-

якою кількістю вимірі;

– легкість в реалізації: метод загалом доволі простий у розумінні та має багато готових реалізацій.

Недоліки методу:

– залежність від розміру набору даних: для дуже великої кількості вимірів та точок у наборі даних може потребувати досить багато часу на проведення відповідних розрахунків;

– припущення про незалежність вимірів: метод робить таке припущення для спрощення розрахунків для окремих вимірів даних, проте ймовірності відповідних даних між вимірами дуже рідко є повністю незалежними один від одного.

Загалом розглянутий метод пошуку аномальних ситуацій (відхилень) є доволі простим у реалізації та застосуванні.

Відповідно, з урахуванням раніше наведених прикладів було сформовано таблицю 1.3, у якій вказано переваги та недоліки відповідних методів.

Таблиця 1.3 – Приклади методів виявлення аномальних ситуацій за типами

Метод	Переваги	Недоліки
1	2	3
ECOD	Простота реалізації, масштабованість	Чутливість до шуму, відсутність контексту між даними
COPOD	Незалежність від розподілу даних, стійкість до шуму	Продуктивність на багатовимірних даних
OCSVM	Продуктивність на багатовимірних даних	Чутливість до вибору параметрів, чутливість до шуму
LOF	Виявлення локальних аномалій, низька параметризація	Чутливість до параметру кількості сусідів, обчислювальна складність
CBLOF	Підтримка кластеризації, виявлення локальних та глобальних аномалій	Залежність від алгоритму кластеризації, чутливість до параметрів
Isolation Forest	Продуктивність, масштабованість	Чутливість до параметрів, залежність від розмірності даних

Кінець таблиці 1.3

1	2	3
LODA	Продуктивність, масштабованість, низька обчислювальна складність	Чутливість до вибору параметрів, обмеження інтерпретованості
SO_GAAL	Генеративний підхід, виявлення локальних та глобальних аномалій	Обчислювальна складність, залежність від параметрів
MO_GAAL	Багатоцільовий, стійкість до шуму	Збільшена обчислювальна складність, складність навчання

Відповідні методи використовуються для виявлення аномальних ситуацій саме на множині даних, що умовно представляються множиною багатовимірних точок у якомусь визначеному просторі, вісі якого представляють відповідні атрибути даних, за якими і відбувається пошук аномалій. Проте ці методи не враховують послідовність процесів, які виконуються у рамках ведення діяльності медичних закладів, які у такому ж форматі представити неможливо.

1.4 Постановка задачі дослідження

Існуючі методи виявлення аномальних ситуацій орієнтовані на виявлені відхилень або у множині даних процесу, або у порядку його робіт, проте відсутнє їхнє поєднання у рамках єдиного підходу. Ця проблема є особливо актуальною в медичній сфері, де критично важливими вважаються як і дані, такі як час очікування пацієнтів, час проведення оглядів та досліджень, так і послідовності виконання бізнес-процесу, зокрема порядок виконання процедур пацієнтами.

Об'єктом дослідження є бізнес-процеси медичних закладів.

Предметом дослідження є методи виявлення аномальних ситуацій у бізнес-процесах медичних закладів.

Метою роботи є дослідження методів виявлення аномальних ситуацій у бізнес-процесах медичних закладів та розробка удосконаленого методу з

урахуванням як відповідності послідовності процесів їхнім моделям, так і нормальності їхніх даних.

Для досягнення зазначеної мети в роботі вирішуються наступні задачі:

- аналіз бізнес-процесів медичних закладів та методів виявлення аномальних ситуацій;

- удосконалення методу виявлення аномальних ситуацій у бізнес-процесах на основі процесного аналізу з використанням систем нечіткого виведення;

- розробка інформаційної технології виявлення аномальних ситуацій на основі процесного аналізу для подальшого впровадження;

- експериментальна перевірка удосконаленого методу виявлення аномальних ситуацій у медичних бізнес-процесах.

2 РОЗРОБКА МЕТОДУ ВИЯВЛЕННЯ АНОМАЛЬНИХ СИТУАЦІЙ У МЕДИЧНИХ БІЗНЕС-ПРОЦЕСАХ З ВИКОРИСТАННЯМ МЕТОДОЛОГІЇ PROCESS MINING

2.1 Розробка підходу до виявлення аномальних ситуацій у бізнес-процесах медичних закладів

Через зростання витрат на обслуговування та забезпечення суспільства якісними та високотехнологічними медичними послугами критично важливим є своєчасне виявлення незвичних дій, аномальних ситуацій, у процесах медичних закладів. Зокрема, саме у клінічних бізнес-процесах виявлення аномальних ситуацій є найбільш критичним та важливим через наявність великих ризиків впливу помилок, що можуть виникати у процесі надання послуг, на стан здоров'я пацієнтів.

Основною ж особливістю у виявленні аномальних ситуацій у бізнес-процесах медичних закладів є висока динамічність самих процесів і їх велика залежність від конкретних випадків у захворюваннях пацієнтів. До того ж, знання і підходи до виявлення захворювань та методи їх лікування залежать від рівня знань та наявного досвіду кожного конкретного медичного спеціаліста, а також наявного матеріального забезпечення, зокрема необхідного медичного обладнання та лікувальних препаратів, в медичних закладах. Усі ці фактори загалом ускладнюють підхід до виявлення аномальних ситуацій.

Також окремими критеріями оцінки аномальних ситуацій є їх класифікація, що базується на причинах їх виникнення [17]:

- виключна ситуація: незвичний порядок виконання процесів під час обслуговування пацієнтів, який загалом може підтримуватись бізнесом;
- шум: помилки у роботі програмного забезпечення (ПЗ) інформаційних систем (ІС) медичних закладів, що може бути пов'язаний з неправильною роботою системи або невірно введеними користувачами системи даними;
- спроби шахрайства: нетиповий спосіб використання системи, який може

нашкодити як пацієнтам, так і медичних закладам, зокрема їхній репутаційній та економічній діяльності.

Якщо виключні ситуації ще можуть відноситись до прийнятних сценаріїв виконання процесів у медичних закладах, то шум та спроби шахрайства вже є небажаними та потребують як попереднього виявлення, так і проведення їхнього аналізу задля виправлення помилок у випадку проблем із системами медичних закладів, або ж притягнення до відповідальності людей, що замішані у спробах шахрайства.

Відповідно, для ефективної роботи систем медичних закладів виникає необхідність розробки та впровадження дієвих механізмів виявлення вищезгаданих типів аномальних ситуацій, а також своєчасне інформування про них усіх зацікавлених сторін окремих бізнес-процесів, у рамках яких такі аномальні ситуації виникають.

Загалом, для виявлення усіх трьох вищезазначених типів аномальних ситуацій можна використати підхід процесного аналізу [17-24], що базується на використанні логів (logs) ІС медичних закладів для побудови моделей їхніх бізнес-процесів та подальшому їх використанні для виявлення аномальних ситуацій – таких логів, що не відповідають визначеній моделі бізнес-процесу та, відповідно, вважаються нетиповими.

У такому випадку логи – це набір певних трас (trase), що характеризуються собою послідовність активностей (activities), які є окремими складовими бізнес-процесів медичних закладів, як от реєстрація пацієнта, очікування прийому до лікаря, проведення прийому та інші.

Формули (2.1), (2.2) та (2.3) відображають відповідні залежності між активностями, трасами та логами

$$A = \{a_1, a_2, \dots, a_n\}, \quad (2.1)$$

$$T = \{t_1, t_2, \dots, t_m\}, t_i \in A^*, i = \overline{1, m}, \quad (2.2)$$

$$L \subseteq T'^2, T' \subseteq T, \quad (2.3)$$

де A – множина усіх активностей $a_i, i = \overline{1, n}$;

A^* – множина усіх можливих комбінацій послідовностей активностей a_i ;

T – множина усіх трас t_i ;

L – лог, множина декартового добутку підмножини трас T' .

На рисунку 2.1 наведено загальний підхід до виявлення аномальних ситуацій з використанням процесного аналізу, що складається з п'яти етапів: фільтрація логів (scoring), застосування алгоритмів виявлення процесів, фільтрація отриманих моделей бізнес-процесів, вибір найбільш підходящої моделі, розділення трас логів (splitting) – фактично виконання виявлення аномальних ситуацій в окремо узятому бізнес-процесі.

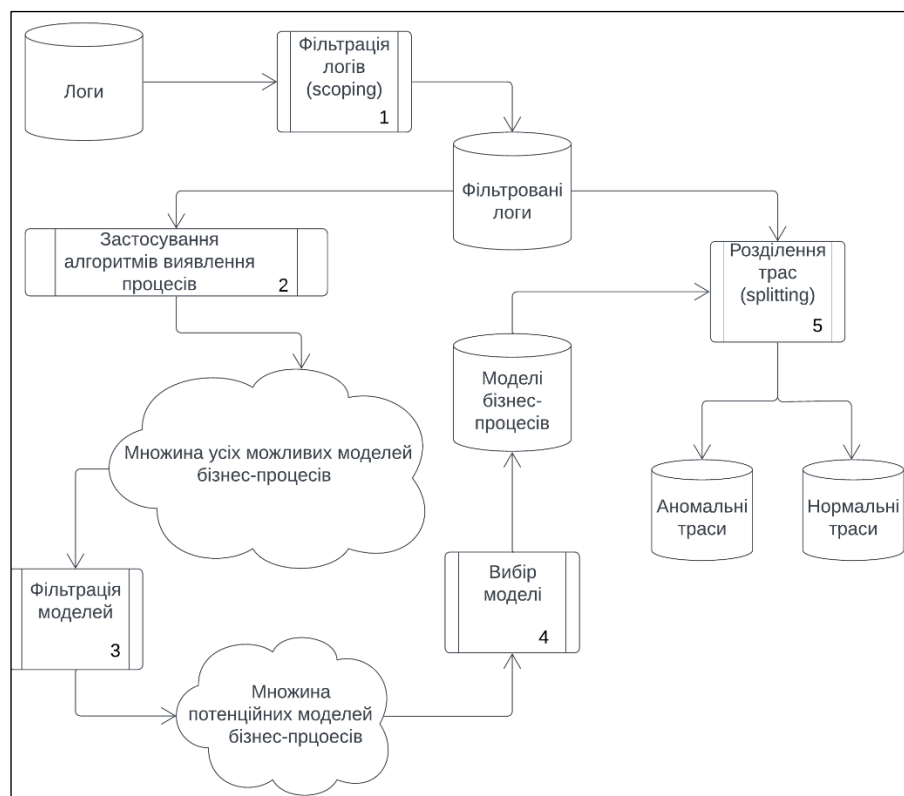


Рисунок 2.1 – Схема загального підходу до виявлення аномальних ситуацій на основі процесного аналізу

Рисунок 2.1 надає розуміння загального підходу до виявлення аномальних ситуацій з використанням процесного аналізу та надає визначає основні етапи

цього методу:

- фільтрація логів (scoring): виконується фільтрація початкових логів з використанням знань з предметної області бізнес-процесів, зокрема, видаляються активності, що вважаються неважливими, та траси, які почались або завершилися за межами часових рамок досліджуваних логів;

- застосування алгоритмів виявлення процесів: на цьому етапі виконується побудова множини усіх можливих моделей бізнес-процесів на основі отриманих фільтрованих логів;

- фільтрація моделей: множина усіх можливих моделей бізнес-процесів потенційно може бути нескінченною, тому виконується їхня фільтрація з урахуванням параметру допасованості (fitness parameter), який визначає мінімальний відсоток трас, що повинні відповідати визначеній моделі бізнес-процесу;

- вибір моделі: з-поміж відфільтрованої множини потенційних моделей обирається та, яка найбільше задовольняє вимогами та особливості конкретного бізнес-процесу;

- розділення трас (splitting): виконання фактичного виявлення аномальних ситуацій у вигляді аномальних трас, що відділяються від початкової множини.

Проте поточний метод виявлення аномальних ситуацій базується лише на процесній послідовності активностей без урахування даних та користувачів, що беруть участь у відповідних бізнес-процесах. Це, як було зазначено раніше, доволі критичний аспект бізнес-процесів медичних закладів, а тому потребує додаткового врахування під час побудови відповідних моделей, що викликає необхідність розширення базового підходу у методі та додавання нового етапу до наявних п'яти, а саме етапу оцінки розділених трас, зокрема нормальних, для виявлення аномалій серед даних, що використовуються відповідними процесами.

З урахуванням впровадження стандартизації та наявності загально прийнятих методик обслуговування пацієнтів у медичних закладах новий етап оцінки розділених трас може базуватись на цих правилах та використовувати

метод оцінки, заснований на нечіткій логіці, що спростить як саму побудову методу, так і подальше його впровадження в ІС медичних закладів завдяки можливості динамічної модифікації правил, що дозволить швидко пристосовуватись до ймовірних змін у майбутньому.

2.2 Удосконалення методу виявлення аномальних ситуацій на основі процесного аналізу бізнес-процесів медичних закладів

Відповідний нечіткий метод оцінки розділених трас, призначений для виявлення аномалій серед трас, що вважаються нормальними, стане останнім етапом у методі виявлення аномальних ситуацій на основі процесного аналізу.

На рисунку 2.2 наведено відповідну оновлену схему з додаванням нового етапу.

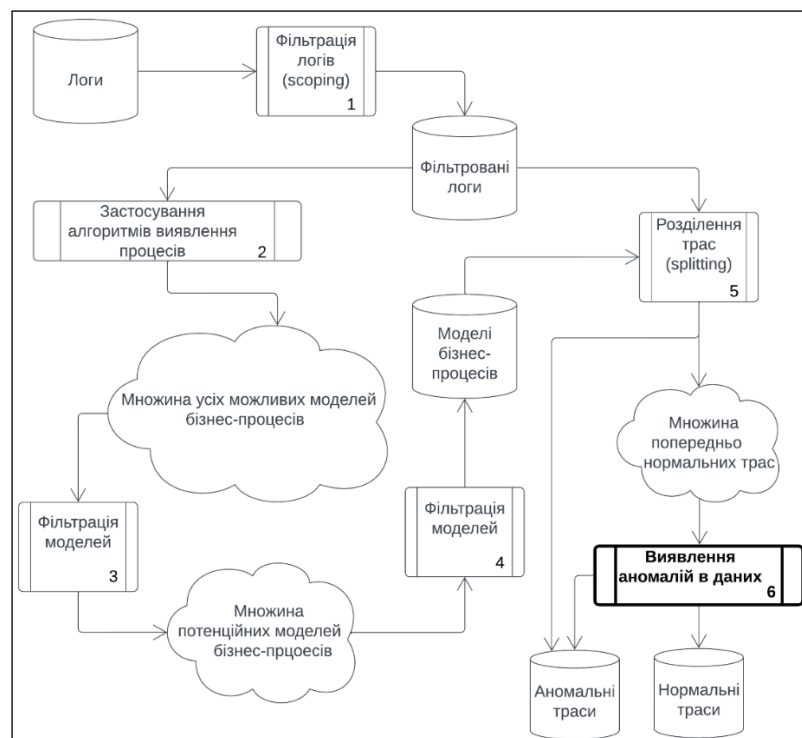


Рисунок 2.2 – Оновлена схема методу виявлення аномальних ситуацій на основі процесного аналізу

Доданий етап оцінки розділення нормальних трас використовуватиме попередньо отриману множину нормальних трас та застосуватиме нечітку систему виведення з правилами, що визначені відповідно до поточного бізнес-процесу, для якого виконується виявлення аномальних ситуацій.

У результаті буде отримано фінальну множину аномальних та нормальних трас як з урахуванням послідовності активностей у трасах відповідно до отриманої моделі бізнес-процесу, так і даних, що брали участь у відповідних процесах.

Матимемо наступну послідовність виконуваних етапів:

фільтрація початкових логів на основі їхньої частоти та відомостей про кожний окремий бізнес-процес медичних закладів;

– виявлення найкращої моделі цільових бізнес-процесів з використанням генетичного алгоритму: виконується групування трьох окремих етапів в один – застосування алгоритмів виявлення процесів, фільтрація моделей та вибір моделі;

– розділення трас з використанням відібраної моделі бізнес-процесів;

– оцінка розділення нормальних трас з використанням нечітких систем виведення: новий доданий етап, призначений для виявлення аномалій серед використаних процесами даних.

На рисунку 2.3 наведено остаточну схему удосконаленого методу виявлення аномальних ситуацій на основі процесного аналізу бізнес процесів медичних закладів.

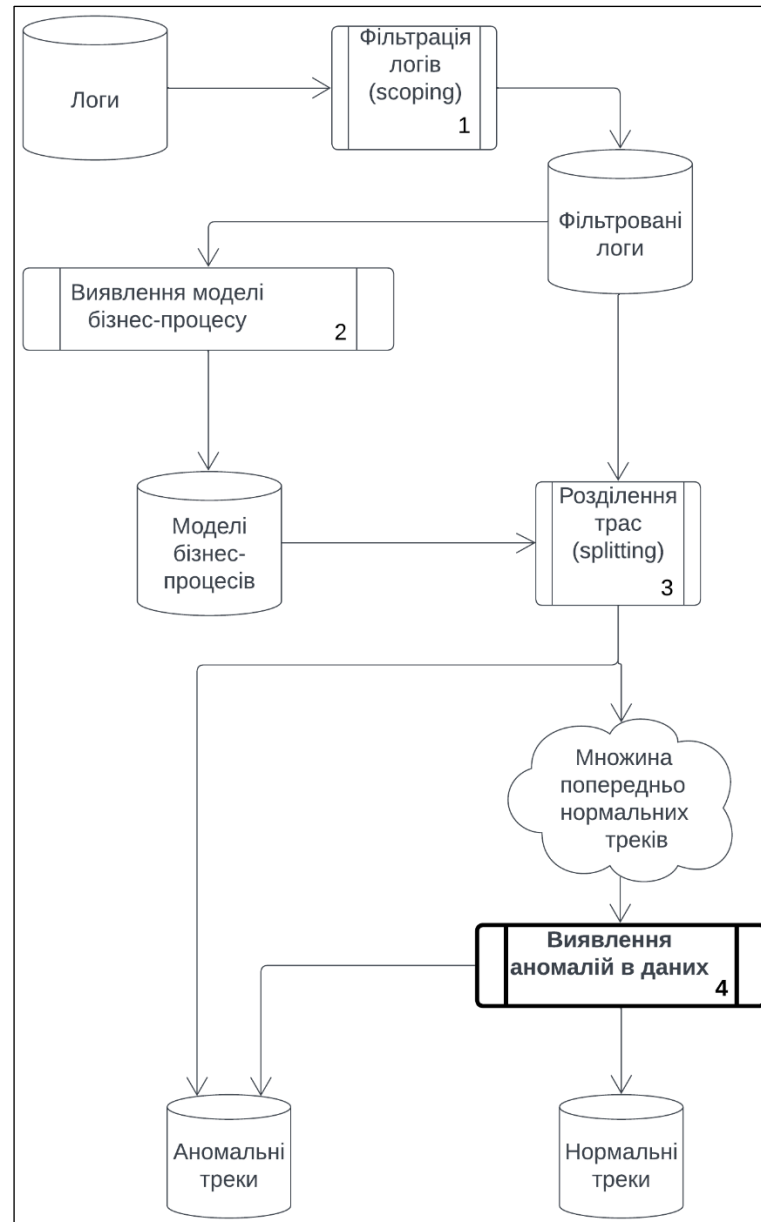


Рисунок 2.3 – Схема удосконаленого методу виявлення аномальних ситуацій на основі процесного аналізу

Фільтрація логів на основі знань про кожний окремий бізнес-процес виконуватиметься із залученням носіїв цих знань – працівників медичних закладів, що залучені у відповідні процеси. Окрім цього, траси, що не повністю відображені у логах, зокрема через відсутність їхнього початку або кінця у визначеному наборі логів, також будуть видалені з результуючих відфільтрованих логів. Цей етап дозволить позбутись активностей, що є неважливими, незмістовними та неповноцінними у логах.

Формула (2.4) відображає відповідне відношення між фільтрованими й початковими логами та їхніми трасами і активностями.

$$L^S = \{t_1, t_2, \dots, t_n\}, t_i \in A^{*S}, i = \overline{1, n}, \quad (2.4)$$

де L^S – фільтровані логи, множина трас t_i , що мають початок і кінець;

A^{*S} – множина комбінацій відфільтрованих активностей.

Наступний етап виявлення типової моделі бізнес-процесу використовуватиме генетичний алгоритм [25], що на основі отриманих фільтрованих логів знайде цільову модель. Загалом цей алгоритм складається з трьох ключових кроків:

а) створення початкової популяції;

б) розрахунок значення функції допасованості: у випадку задоволення мінімального значення допасованості або досягнення максимальної величини, або вичерпання циклів виконання алгоритму отримується модель з найкращим показником – у поточному дослідженні це відсоток трас, що відповідають моделі;

в) застосування генетичних операцій:

1) селекція: відбір найкращих представників початкової популяції, що потраплять у наступний цикл-покоління;

2) схрещування: отримання нових представників популяції (нащадків) на основі схрещування двох представників поточної популяції (батьків);

3) мутація: випадкова зміна в отриманих нащадках.

Кроки б) та в) повторюються доти, доки не буде виконана умова кроку б).

Враховуючи особливості процесного аналізу, а саме наявність каузальних залежностей між процесами, до цього алгоритму додаються попередні кроки виявлення цих залежностей на основі фільтрованих логів зі вже існуючими трасами, що описують певні послідовності активностей.

Відповідно, на рисунку 2.4 наведено остаточну схему послідовності кроків генетичного алгоритму, що використовуватиметься для виявлення найкращої моделі бізнес-процесів.

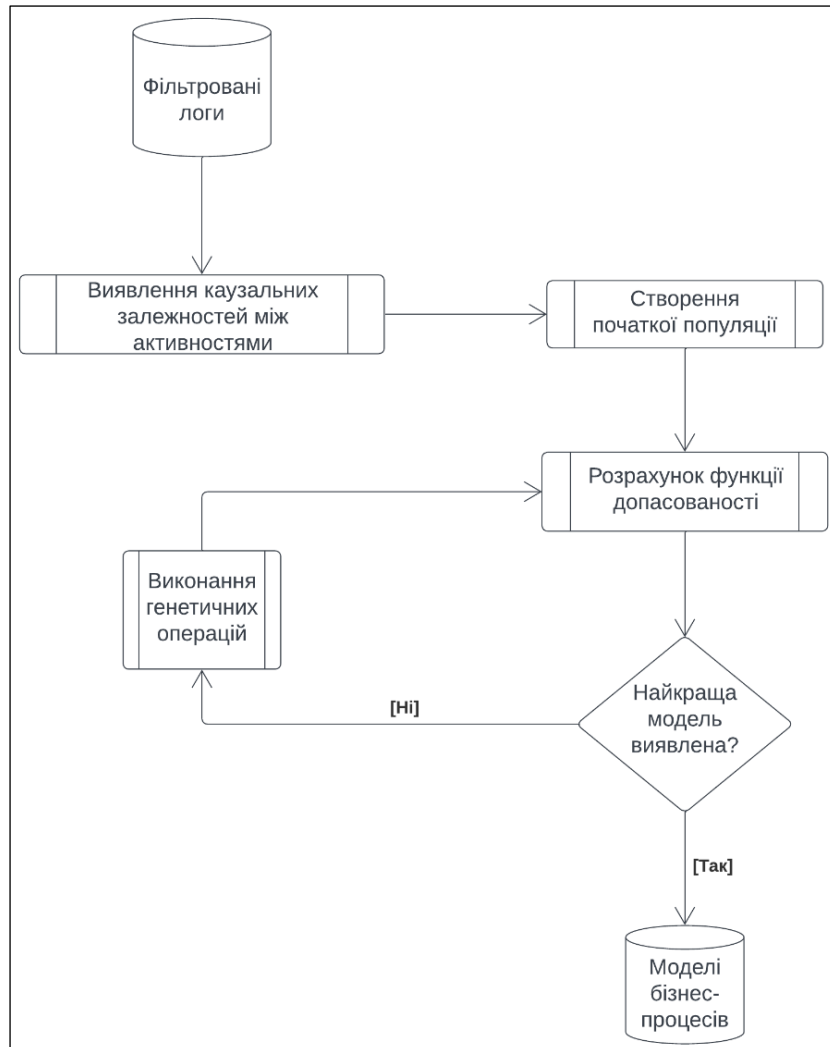


Рисунок 2.4 – Схема послідовності кроків генетичного алгоритму виявлення типової моделі бізнес-процесу

У деталях цей метод пошуку типової моделі бізнес-процесу формує наступний алгоритм:

- а) побудова каузальної матриці [25] для кожної з трас логу;
- б) визначення частот появи послідовності однакових пар активностей у трасах логу;
- в) формування випадковим чином початкової популяції, що складається із

заданої кількості індивідів – моделей бізнес-процесу;

г) модифікація початкової популяції шляхом перестановки активностей в отриманих моделях бізнес-процесів з урахуванням раніше визначених частот появи одбнакових пар активностей – якщо частота вище заданого порогу, то така послідовність визначається сталою;

д) розрахунок функції пристосованості для кожного представника початкової популяції, у якій враховуються такі показники:

- 1) кількість правильно проаналізованих трас;
- 2) кількість повністю завершених трас;

е) виконання селекції шляхом збереження заданої частки представників поточної популяції, значення функцій пристосованості яких є найбільшими – ці індивіди перейдуть у наступне покоління одразу;

ж) виконання схрещування:

1) з поточної популяції випадковим чином обирається задана кількість індивідів;

2) за попередньо розрахованим значенням функції пристосованості обираються два індивіди з найбільшим значенням – вони стають «батьками»;

3) послідовності активностей «батьків» копіюються та копії стають «нащадками»;

4) на множині активностей «нащадків» випадковим чином обирається одна;

5) для обраної активності виконується обмін вхідними та вихідними активностями між обома «нащадками»;

7) виконується нормалізація каузальної матриці з урахуванням змінених послідовностей активностей «нащадків»;

и) в отриманих після схрещування «нащадках» за умови, що випадкове число менше або дорівнює заданому порогу, виконується операція мутації: входи та виходи раніше визначеної випадкової активності в «нащадках» розділяються (зв'язка «I» стає «АБО») або поєднуються (зв'язка «АБО» стає «I»);

к) кроки е-и повторюються доти, доки популяція нвоого покоління не буде повністю заповнена;

л) кроки д-к повторюються доти, доки не буде виконана одна з умов:

1) отримано модель з максимально можливим значенням функції пристосованості;

2) досягнуто максимальної кількості ітерацій, що попередньо задається;

3) максимальне значення функції пристосованості вже довгий час не змінюється – це значення також попередньо задається;

м) модель бізнес-процесу останньої популяції з максимальним значенням функції пристосованості обирається як типова;

н) на основі каузальної матриці знайденої типової моделі бізнес-процесу будується відповідна їй мережа Петрі для графічного відображення моделі.

Приклад задання каузальної матриці для множини активностей $\{A, B, C, D\}$ наведено у вигляді таблиці 2.1.

Таблиця 2.1 – Приклад задання каузальної матриці

Активність	A	B	C	D	Вхідні активності	Вихідні активності
A	0	1	1	0	-	B АБО C
B	0	0	0	1	A	D
C	0	0	0	1	A	D
D	0	0	0	0	B АБО C	-

Рядки цієї матриці відображаються наявність зв'язків між кожною активністю. Зокрема, для активності А наявні зв'язки з активностями В і С, що заповнюються 1. В комірках між активностями з відсутніми зв'язками встановлюються 0. До того ж ця матриця у рядках показує саме послідовну залежність зв'язку: активності в колонках виконуються після активності на початку рядка за умови існуючого зв'язка.

Останні дві колонки фіналізують отримані вхідні та вихідні активності для відповідних активностей у першій колонці. Зокрема, ми маємо, що для

активності А відсутні вхідні активності, бо вона є початковою. Для активності D немає вихідної активності, бо вона кінцева. За умови наявності декількох вхідних чи вихідних активностей для них встановлюються відповідна операція «І» чи «АБО». Ці операції можуть визначатись на основі знань про предметну область, зокрема про відповідний бізнес-процес, ждя логу якого застосовуються процесний аналіз.

Також для спрощеного вигляду цю матрицю можна представити у вигляді таблиці 2.2, де для кожної усніючої активності визначається множи на вхідних та вихідних активностей. Зокера, для початкої активності вхідна множина визначається як пуста, а для кінцевої – вихідна множина є пустою. Для зв'язки «АБО» усі активності належать одній підмножині. Якщо ж існує декілька підмножин, то між ними встановлється зв'язка «І».

Таблиця 2.2 – Спрощений формат подання каузальної матриці у вигляді вхідних та вихідних множин активності

Активність	Вхідна множина активностей	Вихідна множина активностей
A	{}	{{B, C}}
B	{{A}}	{{D}}
C	{{A}}	{{D}}
D	{{B, C}}	{}

На рисунку 2.5 наведено графічне представлення відповідної моделі процесу, що був наведений у прикладі каузальної матриці, у вигляді мережі Петрі. Вона загалом представляє собою направлений граф, вершинами якого є вузли між активностями, а самі активності входять до складу множини ребер – переходи в термінології мережі Петрі. Якщо перехід виходить з одного вузла, то повинен зійтись також в одному визлі – це визначає вибір між активностями, що будуть виконуватись, тобто операцію «АБО». Ця операція зображена у наведеному прикладі. Операція «І» в мережі Петрі визначається переходом з однієї вершини в декілька інших, які сходяться на єдиному переході.

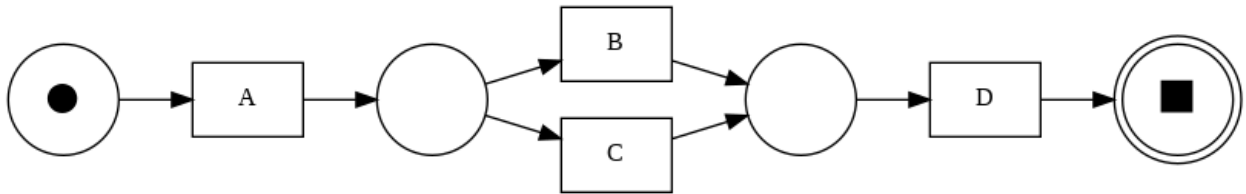


Рисунок 2.5 – Графічне представлення прикладу каузальної матриці у вигляді мережі Петрі

На третьому етапі удосконаленого методу виконуватиметься розділення фільтрованих логів на аномальні та нормальні траси з урахуванням визначеної моделі бізнес-процесу. До аномальних трас відноситимуться ті, які не відповідають моделі, до нормальних – усі інші. У даному випадку критерієм аномальності виступає множина активностей та переходів між ними у трасі, що не відтворюються на визначеній типовій моделі бізнес-процесу.

Останнім етапом, власне, стає оцінка розділення нормальних трас із застосуванням системи нечіткого виведення на основі правил предметної області, що задані для відповідних бізнес-процесів медичних закладів. Як було попередньо зазначено, цей етап дозволить виявити аномалії серед даних, що були залучені під час виконання відповідних процесів. Таким чином будуть визначені аномальні траси, які відповідають визначеній моделі бізнес-процесу, проте мають аномалії в даних.

Загалом система нечіткого виведення складається з наступних компонентів:

- формування бази правил;
- фазифікація входних змінних;
- агрегація підумов;
- активізація підвисновків;
- акумуляція;
- дефазифікація вихідних змінних.

В залежності від алгоритму, що використовується системою нечіткого виведення, кожний з компонентів може мати власні покращення та особливості.

У цій роботі використовуватиметься алгоритм Мамдані, який покладається на базову теоретичну основу систем нечіткого виведення.

Формат правил для формування бази правил системи наведено у (2.5), (2.6)

$$\text{ПРАВИЛО } i: \text{ЯКЩО умова}_i \text{ ТО висновок}_i, i = \overline{1, n}, \quad (2.5)$$

$$\text{умова}_i = \text{висновок}_i = a_j \in b_j \text{ ОП } a_{j+1} \in b_{j+1}, j = \overline{1, m} \quad (2.6)$$

де a_j – лінгвістична змінна, приклад: «час очікування», «вартість прийому»;

b_j – терм, лінгвістичне значення змінної, приклад для «час очікування»: «короткий», «середній», «довгий»;

ОП – деяка бінарна операція, приклад: «І», «АБО».

Фазифікація має на меті встановлення відповідності між конкретним значенням вхідної лінгвістичної змінної та значенням функції приналежності для кожного з її термів. На вході ми маємо множину вхідних значень (2.7), що відповідають кожній з визначених лінгвістичних змінних. У результаті отримуємо множину значень функцій приналежності усіх задіяних термів (2.8).

$$V = \{v_1, v_2, \dots, v_n\}, v_j \in X_j, j = \overline{1, n}, \quad (2.7)$$

$$W = \{w_1, w_2, \dots, w_n\}, w_j = \mu(v_j), j = \overline{1, n}, \quad (2.8)$$

де X_i – універсум (множина усіх можливих значень) лінгвістичної змінної a_j ;

$\mu(v_j)$ – функція приналежності терму b_j , визначає ймовірність приналежності вхідної лінгвістичної змінної до відповідного терму.

У процесі агрегації підумов отримані в (2.8) значення функцій приналежності термів застосовуватимуться як аргументи для отримання значень ступенів істинності умов (2.6) з використанням нечіткої кон'юнкції (2.9) для операцій «І» та нечіткої диз'юнкції (2.10) для операцій «АБО»

$$p_i(w_j \wedge w_{j+1}) = \min\{w_j, w_{j+1}\}, j = \overline{1, n-1}, \quad (2.9)$$

$$p_i(w_j \vee w_{j+1}) = \max\{w_j, w_{j+1}\}, j = \overline{1, n-1} \quad (2.10)$$

У результаті буде отримано множину значень ступенів істинності умов системи (2.11) для визначеної бази правил (2.5).

$$P = \{p_1, p_2, \dots, p_n\} \quad (2.11)$$

У рамках активізації визначається множина ступенів ймовірностей висновків (2.12) бази правил (2.5), які загалом дорівнюють значенням ступенів істинності їхніх умов (2.11).

$$C = \{c_1, c_2, \dots, c_n\} = P \quad (2.12)$$

На основі отриманої множини значень (2.12) визначаються функції приналежності кожного з висновків з використанням модифікованого методу нечіткої композиції, зокрема *min*-активізації (2.13).

$$\mu'_i(x) = \min\{c_i, \mu_i(x)\}, i = \overline{1, n}, \quad (2.13)$$

де $\mu_i(x)$ – функція приналежності терму, який є значенням деякої вихідної лінгвістичної змінної, у даному дослідженні визначені такі терми висновків для лінгвістичної змінної «імовірність аномальності траси»: «нульова», «низька», «середня», «висока».

На етапі акумуляції з використанням формули (2.14) та результатів (2.13) виконується знаходження функції приналежності для кожної лінгвістичної змінної висновків визначеної бази правил, зокрема, як було зазначено вище, у дослідженні використовується одна лінгвістична змінна висновків «імовірність аномальності траси».

$$\mu(x) = \max\{\mu'_1(x), \mu'_2(x), \dots, \mu'_n(x)\} \quad (2.13)$$

На етапі дефазифікації знаходиться якесь дійсне число, що у даному дослідженні виступає відсотком імовірності аномальності кожної окремої траси з логів. Для цього в алгоритмі Мамдані використовується метод центру тяжіння (Centre of Gravity – COG), відображений у (2.14).

$$y = \frac{\int_{min}^{max} x\mu(x)dx}{\int_{min}^{max} \mu(x)dx}, \quad (2.13)$$

де y – дійсне числове значення, отримане у результаті дефазифікації;

x – змінна, що відповідає лінгвістичній змінній;

$\mu(x)$ – функція приналежності лінгвістичної змінної, визначена в (2.12);

min, max – інтервал значень лінгвістичної змінної.

На основі отриманої в (2.13) імовірності того, що траса з логів є аномальною, можна робити остаточний висновок про його аномальність на основі деякого порогу, що може задаватись окремо для кожного процесу в моделі бізнес-процесів, або ж загалом для всіх

Відповідно, траса з логів є аномальною, якщо виконується умова (2.14).

$$y > p, \quad (2.14)$$

де y – отримане у результаті дефазифікації значення імовірності аномальності траси;

p – заданий поріг максимально допустимого значення імовірності аномальності трас у моделі бізнес-процесів.

Загалом за допомогою запропонованого методу на основі системи нечіткого виведення на останньому етапі можна аналізувати будь-яку множину даних процесів в залежності від застосованої бази правил, проте найчастіше

виконуватиметься виявлення аномалій саме серед часових даних як таких, що є обов'язковими атрибутами кожної активності усіх траси в логах ІС медичних закладів.

Відповідно, незалежно від обраної цільової бази правил фінальним результатом використання методу виявлення аномальних ситуацій на основі процесного аналізу стає розділення трас з множини фільтрованих логів на аномальні та нормальні з урахуванням як їхньої відповідності виявленій моделі бізнес-процесів медичних закладів, так і їхніх даних.

3 РОЗРОБКА ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВИЯВЛЕННЯ АНОМАЛЬНИХ СИТУАЦІЙ У МЕДИЧНИХ БІЗНЕС-ПРОЦЕСАХ

Запропонований метод виявлення аномальних ситуацій у бізнес-процесах медичних закладах на основі процесного аналізу та системи нечіткого виведення у процесі впровадження загалом складається з шести компонентів:

- визначення несуттєвих активностей;
- формування бази правил системи нечіткого виведення – підготовча складова нового доданого етапу;
- фільтрація журналу логів;
- побудова типової моделі бізнес-процесу;
- формування множин аномальних та нормальних трас на основі послідовності дій процесу;
- виявлення аномальних трас з використанням даних в активностях – новий доданий етап.

Перший етап виконується із залученням медичних експертів, які на основі знань з предметної області визначають множину несуттєвих активностей із журналу логів, що можна видалити та не використовувати у подальших етапах процесного аналізу.

Другий етап є підготовчим для новододаного шостого. У його рамках на основі знань про предметну область медичні експерти формують базу правил для системи нечіткого виведення. Формат правил визначається у відповідності до заданого системою. Зокрема, у вдосконаленому методі формат правил був визначений формулами (2.5) та (2.6) в другому розділі роботи. Зберігатись ці правила можуть у будь-якій формі – письмовій чи електронній.

Третій етап включає у себе фільтрацію вхідного журналу логів ІС медичних закладів. Формат подання логу було раніше визначено відношеннями (2.1), (2.2) та (2.3), наведеними у другому розділі роботи. Процес фільтрації виконується на стороні платформи для процесного аналізу та складається з двох

етапів:

- із журналу логів видаляються активності, що на першому етапі були визначені несуттєвими;

- на основі часових міток початку та завершення кожної траси визначаються та видаляються ті, які повністю не потрапляють у досліджувані часові рамки.

У результаті виконання третього етапу ми отримуємо множину фільтрованих логів, що використовуються у наступних етапах.

Четвертий етап на основі отриманої множини відфільтрованих логів зосереджується на пошуку типової моделі бізнес-процесу. Для цього використовується відповідний генетичний алгоритм. Уся множина потенційно типових моделей задається з використанням моделі представлення каузальних залежностей, зокерма, мережі Петрі. Критерієм типовості у цьому випадку виступає відсоток трас з множини логів, послідовність дій яких від початку до кінця може бути відтворена на отриманій моделі бізнес-процесу. У результаті визначається типова модель бізнес-процесу, яка використовується на наступному етапі.

П'ятий етап застосовує знайдену типову модель бізнес-процесу для виявлення аномальних трас. Критерієм аномальності у цьому випадку стає наявність вершин та ребер в моделі траси, що не можуть бути відтворені на типовій моделі. Для отримання множини аномальних та нормальних трас виконується відповідний аналіз раніше визначеної множини відфільтрованих логів.

Попередньо отримана множина нормальних трас використовується на останньому шостому етапі, що був доданий у рамках удосконалення методу виявлення відхилень. На цьому етапі з використанням сформованої бази правил системи нечіткого виведення виконується її побудова та виконання. У результаті застосування відповідного методу виявлення аномалій серед даних активностей (дій процесу) ми отримуємо додаткову множину аномальних трас.

Відповідно, у процесі повного виконання заданих етапів на виході ми

отримуємо три множини трас:

- аномальні за послідовністю активностей;
- аномальні за даними активності;
- нормальні за обома складовими.

Обидві множини аномальних трас загалом повністю визначають траси з відхиленням, що у подальшому можуть бути використані в інших модулях в ІС медичних закладів.

У таблиці 3.1 наведено коротку хаарктеристику вище описаних процесів.

Таблиця 3.1 – Характеристика процесів інформаційної технології виявлення аномальних ситуацій

Процес технології	Вхідні дані	Методи, моделі, алгоритми	Виконавець	Вихідні дані
1	2	3	4	5
Визначення несуттєвих активностей	Знання про предметну область, журнал логів ІС медичних закладів	-	Експерти з медичних закладів	Множина несуттєвих активностей
Формування бази правил системи нечіткого виведення	Знання про предметну область	Представлення правил системи нечіткого виведення	Експерти з медичних закладів	База правил системи нечіткого виведення
Фільтрація логів	Множина несуттєвих активностей, журнал логів ІС медичних закладів	-		Фільтровані логи
Побудова моделі бізнес-процесу	Фільтровані логи	Модель представлення каузальних залежностей процесу, генетичний алгоритм	Платформа процесного аналізу	Модель бізнес-процесу

Кінець таблиці 3.1

1	2	3	4	5
Виявлення аномальних та нормальних трас	Фільтровані логи, модель бізнес-процесу	-	Платформа процесного аналізу	Множина аномальних трас, множина попередньо нормальних трас
Виявлення аномальних трас на основі аналізу даних	База правил системи нечіткого виведення, множина попередньо нормальних трас	Метод виявлення аномальних ситуацій на основі системи нечіткого виведення	Платформа процесного аналізу	Доповнення до множини аномальних трас, множина нормальних трас

Задля формалізованого подання інформаційної технології на основі удосконаленого методу виявлення аномальних ситуацій у бізнес-процесах медичних закладів було побудовано відповідну декомповану модель технології у нотації IDEF0, що зображена на рисунку 3.1.

У результаті отримана інформація технологія виявлення аномальних ситуацій у медичних бізнес процесах може бути програмно реалізована та у подальшому впроваджена в існуючі інформаційні системи медичних закладів у вигляді додаткового модулю для платформи процесного аналізу.

Завдяки цьому буде підвищено якість моніторингу бізнес-процесів медичних закладів за рахунок аналізу двох раніше визначених аспектів процесу: послідовності його дій та його даних.

Виявлення аномалій у цьому випадку дозволить підвищити якість надання медичних послуг клієнтам медичних закладів – пацієнтам, – за рахунок своєчасного інформування про нетиповість виконуваних ними процедур, несвоєчасність їхнього виконання, невідповідність стандартам тощо.

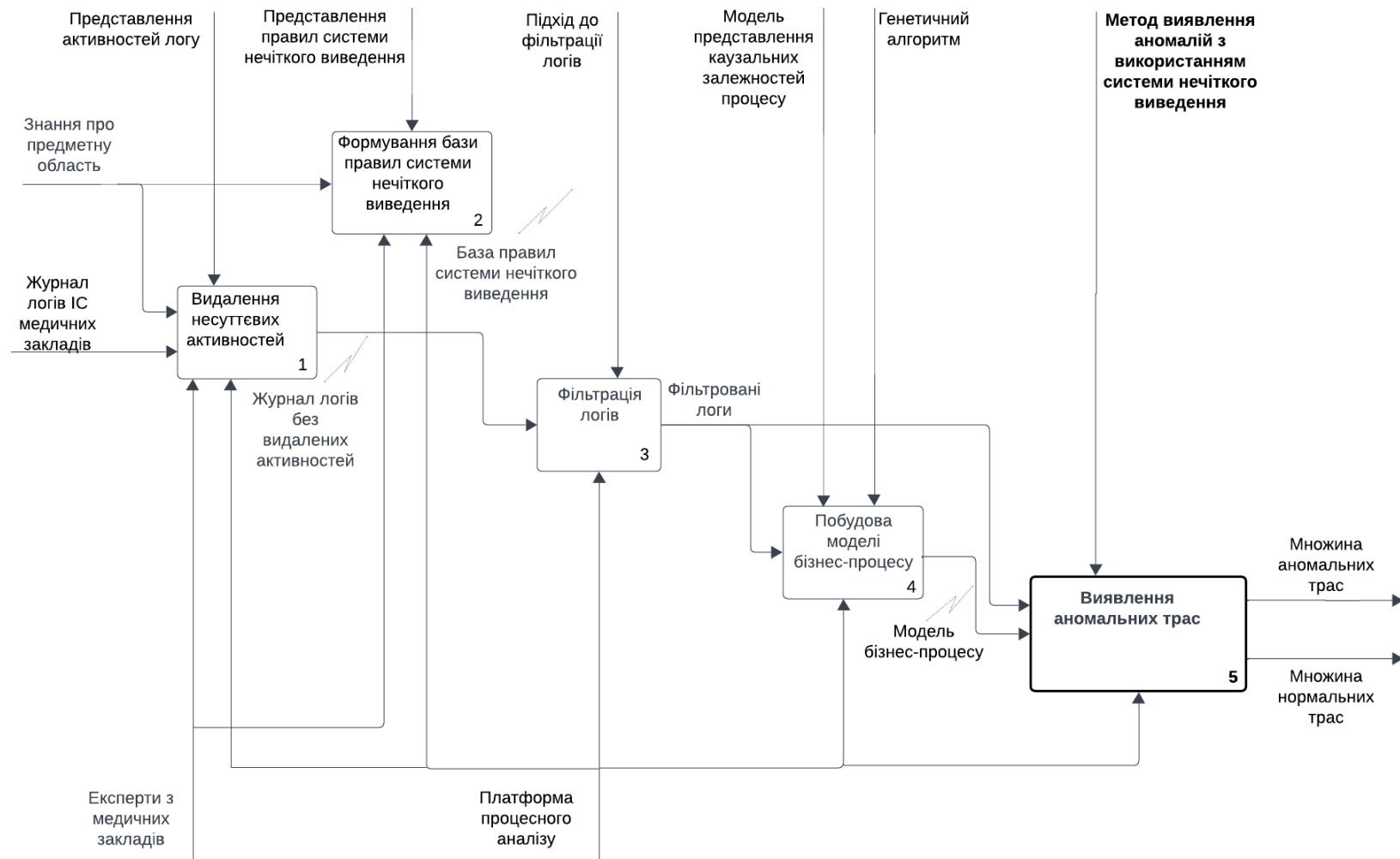


Рисунок 3.1 – Модель інформаційної технології виявлення аномальних ситуацій

4 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА УДОСКОНАЛЕНОГО МЕТОДУ ВИЯВЛЕННЯ АНОМАЛЬНИХ СИТУАЦІЙ

4.1 Програмна реалізація модулю виявлення аномалій з використанням системи нечіткого виведення

Враховуючи, що процес виявлення аномальних ситуацій у бізнес-процесах медичних закладів, або ж загалом у медичних бізнес-процесах, є складовою процесу дотримання безпекових заходів, відповідна програмна реалізація удосконаленого методу належить до системи безпеки інформаційних систем медичних закладів.

До того ж головним результатом роботи стали розробка та додавання нового етапу до вже існуючого підходу виявлення аномалій у послідовностях активностей трас (дій у процесі), який є основною частиною підсистеми виявлення аномальних ситуацій у системі безпеки ІС медичних закладів.

Відповідно, беручи до уваги послідовність процесів у попередньо розробленій інформаційній технології, програмна реалізація буде виконуватись на основі існуючого рішення з додаванням нового модуля, який відповідає новому доданому етапу виявлення аномалій у даних бізнес-процесу на основі системи нечіткого виведення.

Розроблений модуль має на меті бути такою частиною програмного коду, що легко інтегрується в існуючі рішення та швидко замінюється у випадку появи нових більш досконалих підходів до виявлення аномальних ситуацій.

Задля досягнення цієї мети, модуль створюється з урахуванням існуючих вимог до інформаційних систем медичних закладів, зокрема щодо формату подання логів до підсистеми виявлення аномальних ситуацій.

На рисунку 4.1 наведено приклад структури даних логів у вигляді таблиці, де кожний рядок це певна активність, а кожна колонка це її атрибут.

	patient	datetime	action	resource
0	patient 0	2017-01-02 11:40:11	First consult	Dr. Anna
1	patient 0	2017-01-02 12:47:33	Blood test	Lab
2	patient 0	2017-01-02 12:53:50	Physical test	Nurse Jesse
3	patient 0	2017-01-02 16:21:06	Second consult	Dr. Anna
4	patient 0	2017-01-05 13:23:09	Surgery	Dr. Charlie

Рисунок 4.1 – Приклад структури даних логів у вигляді таблиці

У цьому форматі тип активності визначається атрибутом «action», а групуються вони, відповідно, за атрибутом «patient». До того ж дані пацієнтів анонімізовані і слугують лише для ідентифікації повних трас.

Усі інші атрибути, зокрема у даному випадку це «datetime» та «resource», використовуватимуться як джерело відповідних даних, у яких виявлятимуться відхилення. Зокрема, з використанням атрибуту «datetime» визначатиметься час очікування між активностями.

Для програмної реалізації вирішено використовувати мову програмування Python як таку, що вже стала стандартом для роботи з великими даними, їх аналізом та має велику кількість різноманітних готових базових методів та алгоритмів, які доволі просто інтегрувати у нові рішення.

Зокрема, для роботи з нечіткими системами виведення існує бібліотека SciKit-Fuzzy [26], яка є складовою технологічного стеку SciPy [27]. Ця бібліотека надає множину готових реалізацій алгоритмів для роботи з нечіткими числами, множинами та нечіткою логікою, у тому числі зручний інтерфейс для побудови систем нечіткого виведення, який складається з функцій:

- задання універсумів лінгвістичних змінних;
- створення функцій приналежності різних типів: трикутних, трапецієвидних та інших;

- побудови окремих причинно-наслідкових правил, що об'єднуються у єдину базу правил;
- створення системи нечіткого виведення на основі цих правил;
- виконання системи задля рішення відповідних задач, зокрема, у дані роботі це виявлення імовірності аномалій у даних процесів.

У лістингу 4.1 наведено відповідну програмну реалізацію функцій побудови системи нечіткого виведення з використанням вище описаних можливостей бібліотеки SciKit-Fuzzy.

Лістинг 4.1 – Програмна реалізація функцій для побудови системи нечіткого виведення

```

1 def define_antecedent(antecedent_name, u_start, u_end, u_step):
2     return ctrl.Antecedent(np.arange(u_start, u_end, u_step),
3                             antecedent_name);
4 def define_consequent(consequent_name, u_start, u_end, u_step):
5     return ctrl.Consequent(np.arange(u_start, u_end, u_step),
6                             consequent_name);
7 def define_trimfn(variable, term, tridots):
8     variable[term] = fuzz.trimf(variable.universe, tridots);
9 def define_rule(condition, conclusion):
10    return ctrl.Rule(condition, conclusion);
11 def define_system(rules):
12    system = ctrl.ControlSystem(rules);
13    system_simulation = ctrl.ControlSystemSimulation(system);
14    return system_simulation;
15 def execute(system, input_dictionary, output_name):
16    for key, value in input_dictionary.items():
17        system.input[key] = value;
18    system.compute();
19    return system.output[output_name];

```

Відповідні шість функцій виступають обгортками над функціями бібліотеки, що у подальшому, за необхідності, дозволить без проблем модифікувати код, зокрема, задля написання власної логіки та позбавлення залежності від сторонньої бібліотеки.

Для вирішення, чи є отримане значення імовірності аномалії також створено окрему функцію, наведену в лістингу 4.2.

Лістинг 4.2 – Функція визначення чи є отримане значення імовірності аномалії допустимим

```
1 def is_anomaly(y):
2     p = 15; # порогове значення ймовірності аномалії
3     return y > p;
```

У цьому випадку також використовується окрема функція задля можливості подальшого розширення критеріїв аномальності на основі результату, що отримується в системі нечіткого виведення. Поточна реалізація заснована на рівнянні (2.14), що було наведено раніше у розділі 2.

Задля уніфікації використання систем нечіткого виведення, що будуються для різних складових медичних бізнес-процесів, також було створено єдиний вхідний метод, наведений в лістингу 4.3.

Лістинг 4.3 – Вхідний метод для пошуку аномалій на основі побудованої системи нечіткого виведення

```
1 def find_anomalies(system, action_data_rows, input_name):
2     anomalies = [];
3     for _, row in action_data_rows.iterrows():
4         input_dict = {
5             input_name: row['prev_time_diff_s']
6         };
7         y = execute(system, input_dict, 'anomaly_probability');
8         if is_anomaly(y):
9             anomalies.append(row);
10    return pd.DataFrame(anomalies,
11                        columns=action_data_rows.columns);
```

Поточна реалізація зосереджується на пошуку аномалій серед зафіксованих часових міток активностей у логах, що дозволяє визначити час очікування між ними у рамках однієї траси. Відповідне значення «prev_time_diff_s» попередньо розраховується для кожної активності у рамках своєї траси за допомогою коду, наведеному в лістингу 4.4. Цей код складається з низки послідовних кроків розрахунку часу в секундах.

Лістинг 4.4 – Розрахунку відносного часу «prev_time_diff_s» між активностями логу

```

1 # Розрахунок відносного часу між активностями окремих трас
2 patients['prev_datetime'] =
patients.groupby('patient')['datetime'].shift(1);
3 patients['prev_time_diff'] = patients['datetime'] -
patients['prev_datetime'];
4 # Встановлення значення 0 для першої активності у трасі
5 patients['prev_time_diff'] =
patients['prev_time_diff'].fillna(pd.Timedelta(0));
6 # Конвертація у секунди для стандартизованих розрахунків
7 patients['prev_time_diff_s'] =
patients['prev_time_diff'].dt.total_seconds();

```

У даному випадку масив «patients» містить у якості свої елементів активності логу з раніше вказаними атрибутами, зокрема, атрибутом «datetime», який використано у цих розрахунках.

Відповідно, з урахуванням вище наведеної додаткової обробки вхідних даних логу, ми маємо розширену структуру даних (див. рис. 4.2), що використовується розробленим модулем виявлення аномалій.

	patient	datetime	action	resource	prev_datetime	prev_time_diff	prev_time_diff_s
13	patient 2	2017-01-04 10:02:49	First consult	Dr. Anna	NaT	0 days 00:00:00	0.0
14	patient 2	2017-01-06 09:05:01	Physical test	Nurse Jesse	2017-01-04 10:02:49	1 days 23:02:12	169332.0
15	patient 2	2017-01-10 08:05:47	X-ray scan	Team 2	2017-01-06 09:05:01	3 days 23:00:46	342046.0
16	patient 2	2017-01-10 09:12:09	Blood test	Lab	2017-01-10 08:05:47	0 days 01:06:22	3982.0
17	patient 2	2017-01-12 16:32:00	Second consult	Dr. Anna	2017-01-10 09:12:09	2 days 07:19:51	199191.0
18	patient 2	2017-01-13 10:32:51	Medicine	Pharmacy	2017-01-12 16:32:00	0 days 18:00:51	64851.0
19	patient 2	2017-01-17 11:54:48	Final consult	Dr. Ben	2017-01-13 10:32:51	4 days 01:21:57	350517.0

Рисунок 4.2 – Приклад розширеної структури вхідних даних

Додані колонки «prev_datetime», «prev_time_diff» та «prev_time_diff_s» відповідають різним представленням часу активностей та допомагають у подальших розрахунках системи нечуткого виведення.

4.2 Експериментальна перевірка удосконаленого методу виявлення аномальних ситуацій

Експериментальна перевірка удосконаленого методу виявлення аномальних ситуацій у медичних бізнес-процесах, зокрема у типовому процесі лікування пацієнтів, проводиться у відповідності до наступних етапів:

- попередня обробка вхідних даних;
- задання універсумів лінгвістичних змінних та функцій приналежності їхніх термів для активностей логу;
- побудова бази правил та системи нечіткого виведення з використанням визначених причин та наслідків у вигляді лінгвістичних змінних;
- виконання системи з використанням попередньо оброблених вхідних даних;
- оцінка ефективності виконання пошуку аномалій серед даних процесу з використанням системи нечіткого виведення.

Для експериментальної перевірки удосконаленого методу було використано набір даних з сайту kaggle [28], що відповідає раніше заданому формату вхідних даних.

Цей набір даних представляє собою лог одного медичного бізнес-процесу лікування пацієнтів та був згенерований вручну його автором. Він загалом складається з 8 активностей, 100 трас та усього 690 записів. Додатково було додано 10 нових трас з нетиповим порядком активностей задля перевірки їхнього впливу на отриманий метод.

Відповідно, попередня обробка вхідних даних включає у себе розрахунок різниці часу між активностями логу у межах окремих трас, що у подальшому використовуватиметься системою нечіткого виведення. Цей етап виконується з використанням попередньо реалізованого програмного коду, по був описаний у попередньому підрозділі.

У подальших розрахунках використовуватиметься новостворений атрибут

вхідних даних «prev_time_diff_s».

Крім того, для наглядності зв'язку між активностями, зокрема їхнього порядку, типову модель медичного бізнес-процесу лікування пацієнтів, що отримується у попередніх етапах існуючого підходу до виявлення аномалій серед послідовностей дій процесу, було представлено у вигляді мережі Петрі.

Відповідну модель зображено на рисунку 4.3.

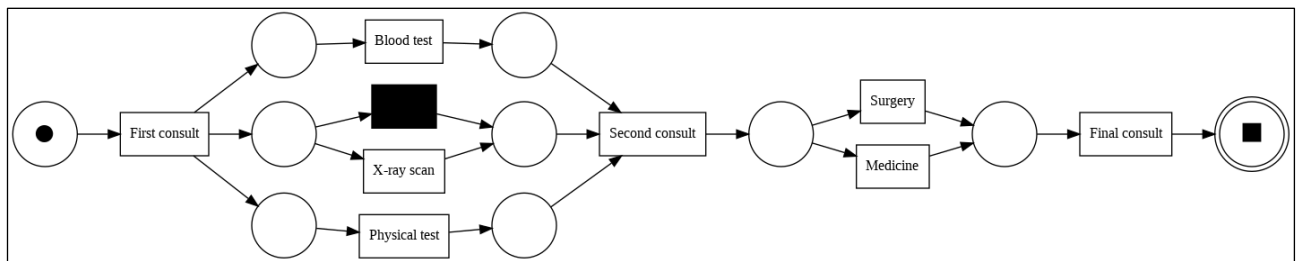


Рисунок 4.3 – Отримана типова модель медичного бізнес-процесу

У цій моделі ми маємо усі вісім активностей з набору даних:

- First consult: перша консультація з лікарем;
- Blood test: аналіз крові пацієнта;
- X-ray scan: рентгенівський знімок;
- Physical test: фізичні тести;
- Second consult: друга консультація з лікарем;
- Surgery: проведення хірургічної операції;
- Medicine: прийом лікувальних засобів;
- Final consult: остання консультація з лікарем.

У процесі виявлення аномалій саме у послідовності дій процесу було визначено відповідну множину аномальних трас, що не відповідають отриманій типовій моделі. На рисунку 4.4 наведено приклад моделі однієї з виявлених трас у вигляді мережі Петрі.

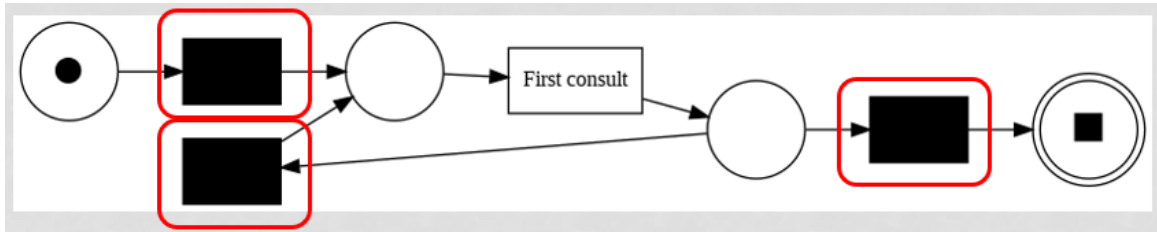


Рисунок 4.4 – Модель виявленої аномальної траси

На цьому рисунку ми маємо виділені червоним трикутником переходи, що не відображені в типовій моделі бізнес-процесу, а тому їх неможливо відтворити для проходження процесу від початку до кінця, що і є критерієм аномальності траси.

Для побудови системи нечіткого виведення було використано наступні лінгвістичні змінні та їхні терми:

- «час очікування активності»: «малий», «середній», «довгий»;
- «імовірність аномалії»: «нульова», «низька», «середня», «висока».

Відповідні універсуми та функції приналежності було побудовано з використанням попередньо реалізованої програмної частини.

Для відображення графіків функцій приналежності термів на заданих універсумах лінгвістичних змінних було використано код, наведений в лістингу 4.5.

Лістинг 4.5 – Код для відображення графіків функцій приналежності термів лінгвістичних змінних.

```

1 anomaly_probability = define_consequent('anomaly_probability',
0, 100, 0.01)
2 define_trimfn(anomaly_probability, 'none', [0, 0, 5]);
3 define_trimfn(anomaly_probability, 'low', [2.5, 15, 30]);
4 define_trimfn(anomaly_probability, 'medium', [20, 40, 60]);
5 define_trimfn(anomaly_probability, 'high', [50, 100, 100]);
6 # Відображення графіків функцій приналежності
7 plt.figure(figsize=(10, 6))
8 plt.plot(anomaly_probability.universe,
anomaly_probability['none'].mf, label='нульова')
10 plt.plot(anomaly_probability.universe,
```

```

anomaly_probability['low'].mf, label='низька')
11 plt.plot(anomaly_probability.universe,
anomaly_probability['medium'].mf, label='середня')
12 plt.plot(anomaly_probability.universe,
anomaly_probability['high'].mf, label='висока')
13 plt.title('Функції приналежності для імовірності аномалії')
14 plt.xlabel('Імовірність аномалії (%)')
15 plt.ylabel('Ступінь приналежності')
16 plt.legend()
17 plt.show()

```

На рисунку 4.5 наведено графіки функцій приналежності термів лінгвістичної змінної «імовірність аномалії».

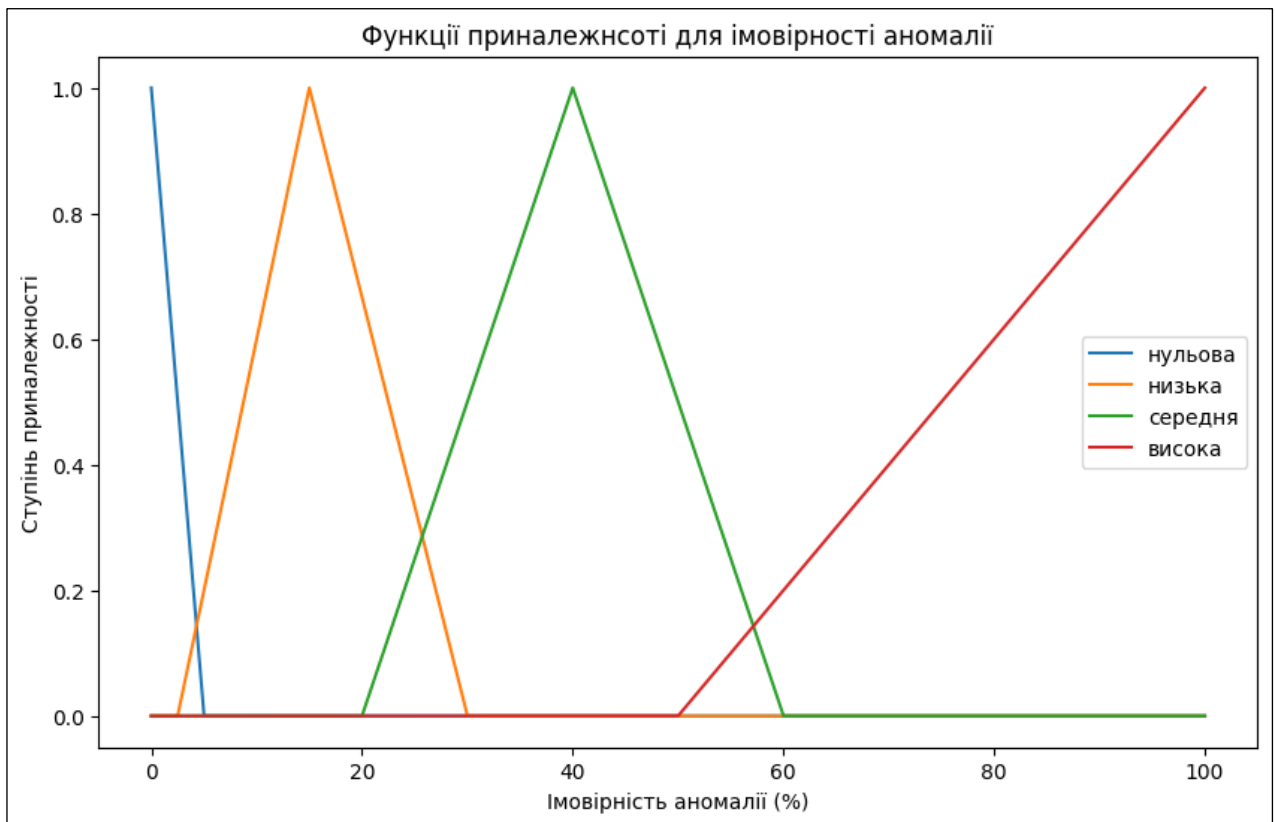


Рисунок 4.5 – Графіки термів лінгвістичної змінної «імовірність аномалії»

Також для двох окремих активностей «Blood test» та «Medicine» було визначено свої функції приналежності, зображені на рисунку 4.6 та рисунку 4.7 відповідно.

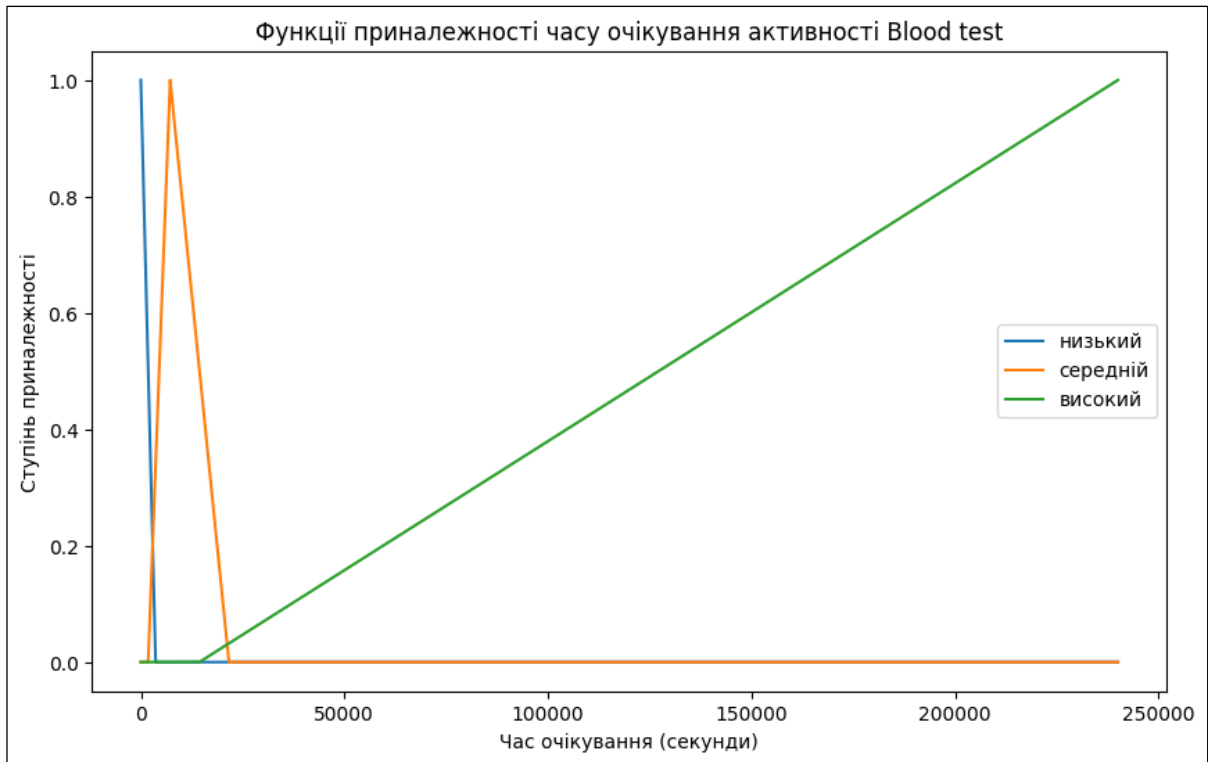


Рисунок 4.6 – Функції приналежності часу очікування активності «Blood test»

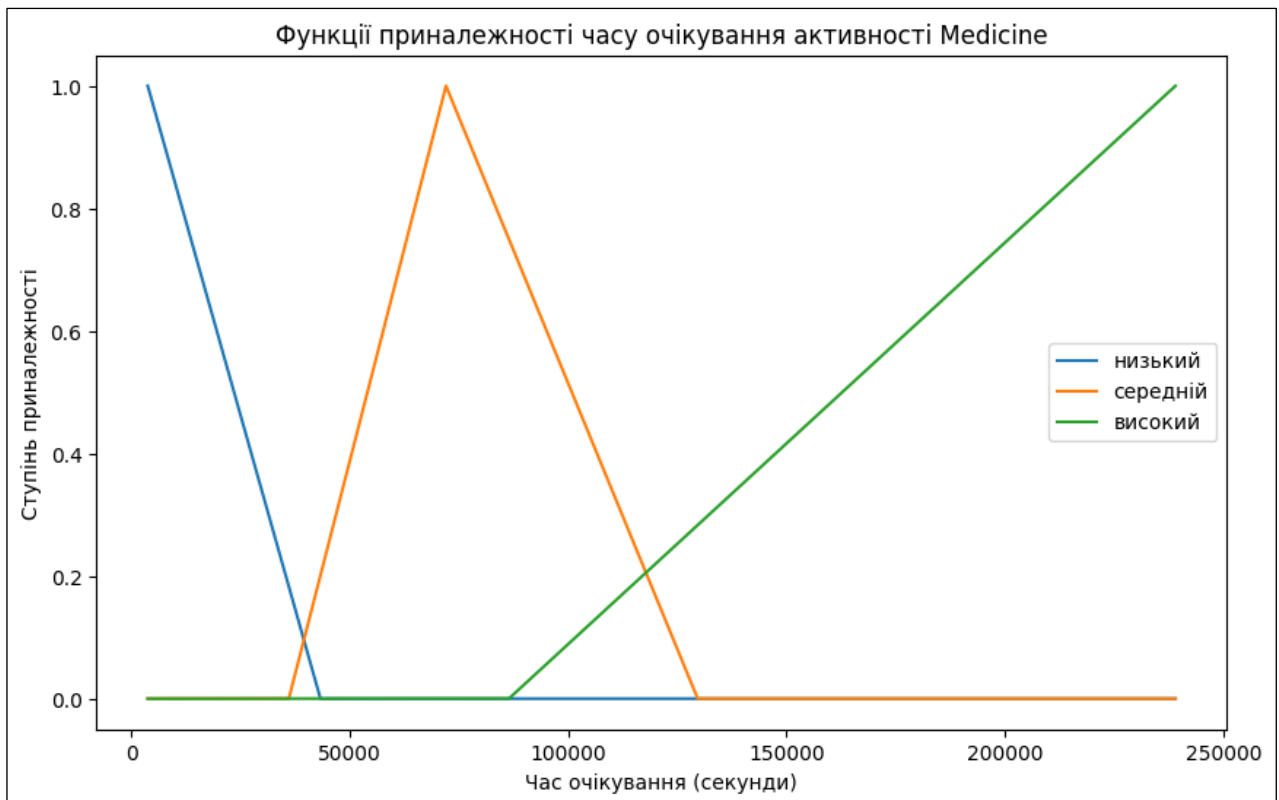


Рисунок 4.7 – Функції приналежності часу очікування активності «Medicine»

База правил та система нечіткого виведення також будується на основі раніше наведеної програмної реалізації. Зокрема, для лінгвістичної змінної причини «час очікування активності» та лінгвістичної змінної наслідку «імовірність аномалії» було сформульовано наступні два правила:

– ЯКЩО «час очікування активності малий» АБО «час очікування активності середній» ТО «імовірність аномалії нульова»;

– ЯКЩО «час очікування активності високий» ТО «імовірність аномалії висока».

Для побудови цих правил використовується відповідна програмна реалізація. У лістингу 4.6 наведено фрагмент коду, який формує правила, створює об'єкт системи нечіткого виведення та виконує її. На виході ми отримуємо множину аномальних треків.

Лістинг 4.6 – Фрагмент коду створення та виконання системи нечіткого виведення

```

1 rule1 = define_rule(blood_test_time['low']
2                   | blood_test_time['medium'],
3                   anomaly_probability['none']);
4 rule2 = define_rule(blood_test_time['high'],
5                   anomaly_probability['high']);
6 system = define_system([rule1, rule2]);
7 blood_test_data = actions_data['Blood test'][0];
8 df1 = execute_for_action(system,
9                          blood_test_data,
10                         'blood_test_time');
```

У таблиці 4.1 наведено результат проведеної експериментальної перевірки удосконаленого методу виявлення аномальних ситуацій у медичних бізнес-процесах, зокрема у процесі лікування пацієнтів.

У рамках цієї перевірки було використано дані часу очікування усіх восьми наявних в логу активностей. На основі цих даних та з використанням вище наведеного коду проводилось визначення, чи є відповідна траса аномальною.

Таблиця 4.1 – Результати експериментальної перевірки удосконаленого методу виявлення аномальних ситуацій

Виявлено з використанням		Кількість виявлених аномалій	Номери аномальних трас
послідовності активностей		5	100, 101, 102, 103, 109
даних активностей	Blood test	4	3, 6, 28, 29
	Medicine	4	18, 44, 69, 70
ЗАГАЛОМ:		13	3, 6, 18, 28, 29, 44, 69, 70, 100, 101, 102, 103, 109

Як наведено у таблиці, за результатами виконання удосконаленого методу виявлення аномалій на основі процесного аналізу логу бізнес-процесу лікування пацієнтів було виявлено усього 13 аномальних трас. Серед них:

- 5 було виявлено за аналізом послідовності активностей з використанням знайденої типової моделі бізнес-процесу;
- 8 було виявлено на новому етапі, призначеному для пошуку аномалій серед даних активностей трас (дій процесу).

Відповідно, завдяки удосконаленню методу виявлення аномальних ситуацій на основі процесного аналізу шляхом додавання нового етапу для виявлення аномалій у даних активностей було підвищено точність виявлення трас з відхиленнями.

ВИСНОВКИ

У рамках виконання кваліфікаційної роботи було розглянуто проблему відсутності єдиного підходу до виявлення аномальних ситуацій у бізнес-процесах медичних закладів як з урахуванням послідовності дій процесу, так і його даних. Вирішення цієї проблеми є однією зі складових впровадження поліпшених засобів моніторингу та управління медичними бізнес-процесами, що зосереджені на лікуванні пацієнтів – головному бізнес-процесу лікувально-профілактичних медичних закладів, що розглядались в роботі.

Було проведено аналіз бізнес-процесів медичних закладів та існуючих методів виявлення аномальних ситуацій. Визначено класифікацію медичних закладів та методів виявлення аномальних ситуацій. Сформовано відповідну постановку задачі за результатами проведеного аналізу.

Удосконалено існуючий метод виявлення аномальних ситуацій на основі процесного аналізу шляхом додавання нового етапу, що базується на системі нечіткого виведення. Відповідний етап призначений для виявлення відхилень у даних бізнес-процесу з використанням бази причинно-наслідкових правил.

Розроблено інформаційну технологію, що реалізує удосконалений метод виявлення аномальних ситуацій у медичних бізнес-процесах та призначена для подальшого впровадження в існуючі інформаційні системи медичних закладів.

Виконано програмну реалізацію удосконаленого методу та проведено експериментальну перевірку, у ході якої було доведено практичну цінність методу – підвищено точність виявлення аномальних треків у журналі логів бізнес-процесу.

Кваліфікаційну роботу було оформлено згідно з ДСТУ 3008:2015 [29] та ДСТУ 8302:2015 [30].

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Про затвердження переліків закладів охорони здоров'я, лікарських посад, посад фармацевтів, посад фахівців з фармацевтичною освітою, посад професіоналів у галузі охорони здоров'я, посад фахівців у галузі охорони здоров'я, посад професіоналів з вищою немедичною освітою у закладах охорони здоров'я та посад професіоналів з вищою немедичною освітою, залучених до надання реабілітаційної допомоги у складі мультидисциплінарних реабілітаційних команд в закладах охорони здоров'я: Наказ Міністерства охорони здоров'я України від 28.10.2002 р. № 385. URL: <https://zakon.rada.gov.ua/laws/show/z0892-02#Text> (дата звернення: 15.01.2025).
2. Lenz R., Reichert M. IT Support for Healthcare Processes – Premises, Challenges, Perspectives. *Data & Knowledge Engineering*. 2007. Vol. 61(1). P. 39–58.
3. Poulymenopoulou M., Malamateniou F., Vassilacopoulos G. Specifying Workflow Process Requirements for an Emergency Medical Service. *Journal of medical systems*. 2003. Vol. 27(4). P. 325–335.
4. Ляшенко Р. В. Основні підходи до управління в бізнесі. Молодий вчений. 2018. Т. 12(64). С. 296–301.
5. Buttigieg S., Dey P., Gauci D. Business process management in health care: current challenges and future prospects. *Innovation and Entrepreneurship in Health*. 2016. Vol. 3. P. 1–13.
6. Fernández A., Fernández. D., García Y. Business Process Management for optimizing clinical processes: A systematic literature review. *Health Informatics Journal*. 2020. Vol. 26(2). P. 1305–1320.
7. Goldstein M., Uchida S. A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLOS ONE*. 2016. Vol. 11(4). 31 p.
8. Baldan Lozano F. J., García-Gil D. Benchmarking Anomaly Detection

Methods: Insights From the UCR Time Series Anomaly Archive. *Expert Systems*. 2024. 11 p.

9. Ecod: Unsupervised Outlier Detection Using Empirical Cumulative Distribution Functions / Li Z. et al. *IEEE Transactions on Knowledge and Data Engineering*. 2022. Vol. 35. P. 12181–12193.

10. Copod: Copula-Based Outlier Detection / Li Z. et al. *IEEE International Conference on Data Mining (ICDM)*. 2020. P. 1118–1123.

11. Estimating the Support of a High-Dimensional Distribution / Schölkopf B. et al. *Neural Computation*. 2001. Vol. 13. P. 1443–1471.

12. Lof: Identifying Density-Based Local Outliers / Breunig M. M. et al. *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*. 2000. P. 93–104.

13. He Z., Xu X. Deng S. Discovering Cluster-Based Local Outliers. *Pattern Recognition Letters*. 2003. Vol. 24. P. 1641–1650.

14. Liu F. T., Ting K. M., Zhou Z.-H. Isolation-Based Anomaly Detection. *ACM Transactions on Knowledge Discovery From Data (TKDD)*. 2012. Vol. 6. P. 1–39.

15. Pevny T. Loda: Lightweight On-Line Detector of Anomalies. *Machine Learning*. Vol. 102. P. 275–304.

16. Deep One-Class Classification / Ruff L. et al. *International Conference on Machine Learning*. 2018. P. 4393–4402.

17. Generative Adversarial Active Learning for Unsupervised Outlier Detection / Liu Y. et al. *IEEE Transactions on Knowledge and Data Engineering*. 2019. Vol. 32. P. 1517–1528.

18. Чалий С.Ф., Моцейко Т.М. Представлення аномальних послідовностей робіт у бізнес-процесах з використанням темпоральних правил. ТОВ «Друкарня Мадрид». 2024. С. 113–114.

19. Чалий С.Ф., Полозов М.О. Виявлення бізнес-правил на основі темпоральних залежностей з використанням технології інтелектуального аналізу процесів. Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління : тез. доп. дванадцятій міжнародної науково-технічної

конференції, 27-28 квітня 2022 р. Т. 2. Баку–Харків–Жиліна, 2022. С. 84.

20. Chalyi S.F., Bogatov I.O. Technology of automated construction of a business process prototype model based on pre-processing of event logs. Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies. 2020. Vol. 2(4), P. 57–63.

21. Чалий С.Ф., Прибильнова І.Б. Construction of a situational representation of knowledge based on the analysis of logs. Bulletin of National Technical University "KhPI". Series: System Analysis, Control and Information Technologies. 2017. Vol. 28, P. 70–73.

22. Чалий С.Ф., Левикін І.В. Виявлення інтервалів очікування в бізнес-процесах на основі аналізу послідовностей подій. Технологічний аудит і резерви виробництв. 2016. № 5. С. 71–76.

23. Чалий С.Ф., Левикін І.В. Метод побудови інтервальної моделі процесу розв’язування задачі у складі прецеденту на основі аналізу журналу подій. Наукові праці Вінницького національного технічного університету. 2016. № 4. URL: <https://praci.vntu.edu.ua/index.php/praci/article/view/485/484> (дата звернення: 15.01.2025).

24. Чалий С. Ф., Буцукіна І. Б. Вилучення прецедентів з використанням технології інтелектуального аналізу процесів. Системи обробки інформації. 2015. № 9. С. 79-82.

25. Aalst W., Medeiros A, Weijters A. Genetic Process Mining. An experimental valuation, Data Mining and Knowledge Discovery. 2005. Vol. 14. P. 48–69.

26. SciKit Fuzzy. URL: <https://pythonhosted.org/scikit-fuzzy/> (дата звернення: 15.01.2025).

27. SciPy. URL: <https://scipy.org/> (дата звернення: 15.01.2025).

28. Event Log of Hospital. URL: <https://www.kaggle.com/datasets/ujjwal15/event-log-of-hospital/data> (дата звернення: 15.01.2025).

29. ДСТУ 3008:2015. Інформація та документація. Звіти у сфері науки і техніки. Структура та правила оформлювання. Чинний від 2017-07-01. – Київ:

ДП «УкрНДНЦ», 2016. – 31 с.

30. ДСТУ 8302:2015. Інформація та документація. Бібліографічне посилання. Загальні положення та правила складання. Чинний від 2016-07-01. – Вид. офіц. Київ : УкрНДНЦ, 2016. – 16 с.