

О.Ю. Пшеничний, Н.Б. Шаховська
Національний університет “Львівська політехніка”,
кафедра інформаційних систем та мереж

ІНФОРМАЦІЙНА СИСТЕМА АНАЛІЗУ ЗАХВОРЮВАНOSTI НА ПАНКРЕАТИТ

© Пшеничний О.Ю., Шаховська Н.Б., 2011

Описано архітектуру програмної системи, що дасть змогу автоматизувати збір, аналіз даних та формування звітів та прогнозів захворюваності на панкреатит.

Ключові слова: панкреатит, статистика, аналіз, моделювання, програмна система.

This paper describes architecture of software system that would be able to collect, process and predict needed statistic values in disease spreading model.

Key words: pancreatitis, statistics, analyzes, modeling, software system.

Вступ

Панкреатит – це група захворювань та синдромів, за яких спостерігається запалення підшлункової залози [2]

За характером перебігу панкреатит поділяється на:

- гострий;
- гострий рецидивуючий;
- хронічний;
- загострення хронічного.

Останніми роками в Україні та світі спостерігається різке зростання випадків гострого панкреатиту, а також летальних результатів захворювання, незважаючи на можливості сучасної медицини.

Панкреатит входить до списку захворювань, що жорстко контролюються Міністерством охорони здоров'я України. Але варто відзначити, що дані, які збираються панкреатологічними центрами [1], мають доволі мало критеріїв градації, які не повністю описують структуру та можливі причини захворювання.

Панкреатит має чіткі симптоми та перевірені часом методи діагностування, до того ж це хірургічна хвороба, яка часто закінчується оперативним втручанням, де помилки діагностування неприпустимі. Тому сумнівів у точності діагностування захворювання не виникає.

Водночас з наявністю доволі точних річних звітів зі статистичними даними по кількості хворих на панкреатит різних форм в Україні проводиться вкрай мало аналітичних досліджень цих даних. У результаті планування роботи панкреатологічних центрів виділення коштів на їх утримання та лікування хворих здійснюється плановими методами, спираючись на попередні роки. В умовах динаміки теперішньої ситуації з захворюваності панкреатитом такий метод надто неточний. Окрім того, відсутність детального аналізу зібраних даних не дає змоги виявити вплив соціальних, природних, екологічних, фізичних та інших чинників на імовірність виникнення захворювання, а отже, і розроблення ефективних засобів попередження хвороби є неможливою. Сьогодні статистика ведеться на річній базі (узагальнені звіти за рік) тільки по кількості хворих та кількості померлих залежно від комбінації критеріїв: причина загострення хвороби, форма захворювання, спосіб лікування. Визначення додаткових критеріїв та чинників, що впливають на виникнення та перебіг захворювання є важливим завданням галузі, яке сьогодні не виконується. Тому проведення детального аналізу статистики є актуальним та необхідним для галузі.

Постановка завдання

Ця робота аналізує отримані дані з Львівського міського панкреатологічного центру за допомогою інформаційної системи, виявляє важливі залежності даних предметної галузі, що не описуються параметрами, зібраними до сьогоднішнього дня, для проведення подальших детальних досліджень причин цих явищ.

Дослідження виявлення залежностей даних займалися Поспелов [4], Пінскер [5], Шефе [6] та ін. Проте особливістю даних, які необхідно проаналізувати для цієї предметної галузі, є те, що структури даних та призначення атрибутів наперед невідомі. Тому ставиться завдання попереднього визначення структур даних вхідних джерел та приведення їх до одного виду.

Оскільки важливість знайдених залежностей може бути визначена виключно експертно, то подання залежностей здійснюватиметься графічно.

Опис архітектури системи моніторингу та аналізу даних по захворюваності на панкреатит.

Структура бази даних

Дані, що були надані Львівським міським панкреатологічним центром, представлені у вигляді роздрукованих звітів, і для обробки та програмного аналізу їх було необхідно перевести у цифровий вигляд. Як засіб збереження був обраний СУБД MS SQL Server 2008. Представлення інформації у базі даних дає змогу отримувати легкий доступ та потужні засоби фільтрування та обробки даних.

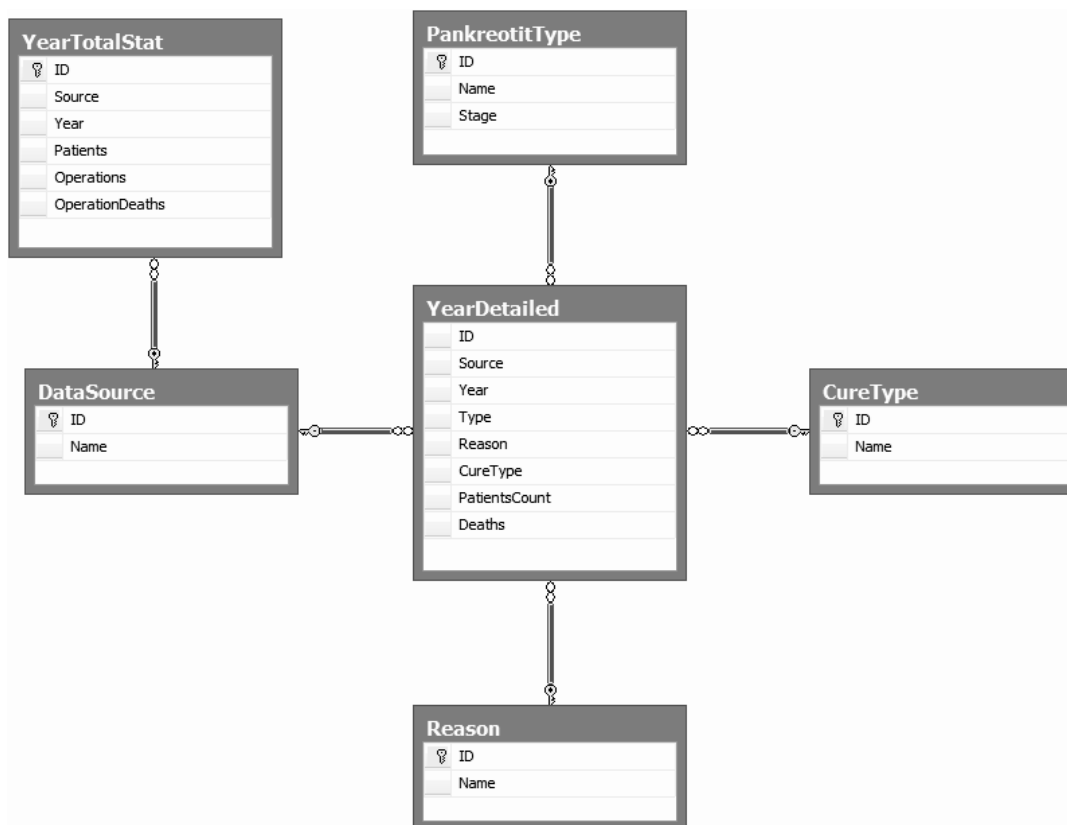


Рис. 1. Діаграма бази даних статистики захворювань на панкреатит

Схема бази даних сьогодні доволі проста, оскільки адаптована до єдиного сьогодні джерела даних – Львівського міського панкреатологічного центру і включає тільки критерії, за якими в ньому ведеться статистика. У майбутньому схема буде ускладнюватись додаванням нових параметрів ведення статистики.

Архітектура системи

Для реалізації завдань моніторингу, аналізу та прогнозування панкреатологічної ситуації була спроектована інформаційна система, архітектуру якої показано на рис. 2.

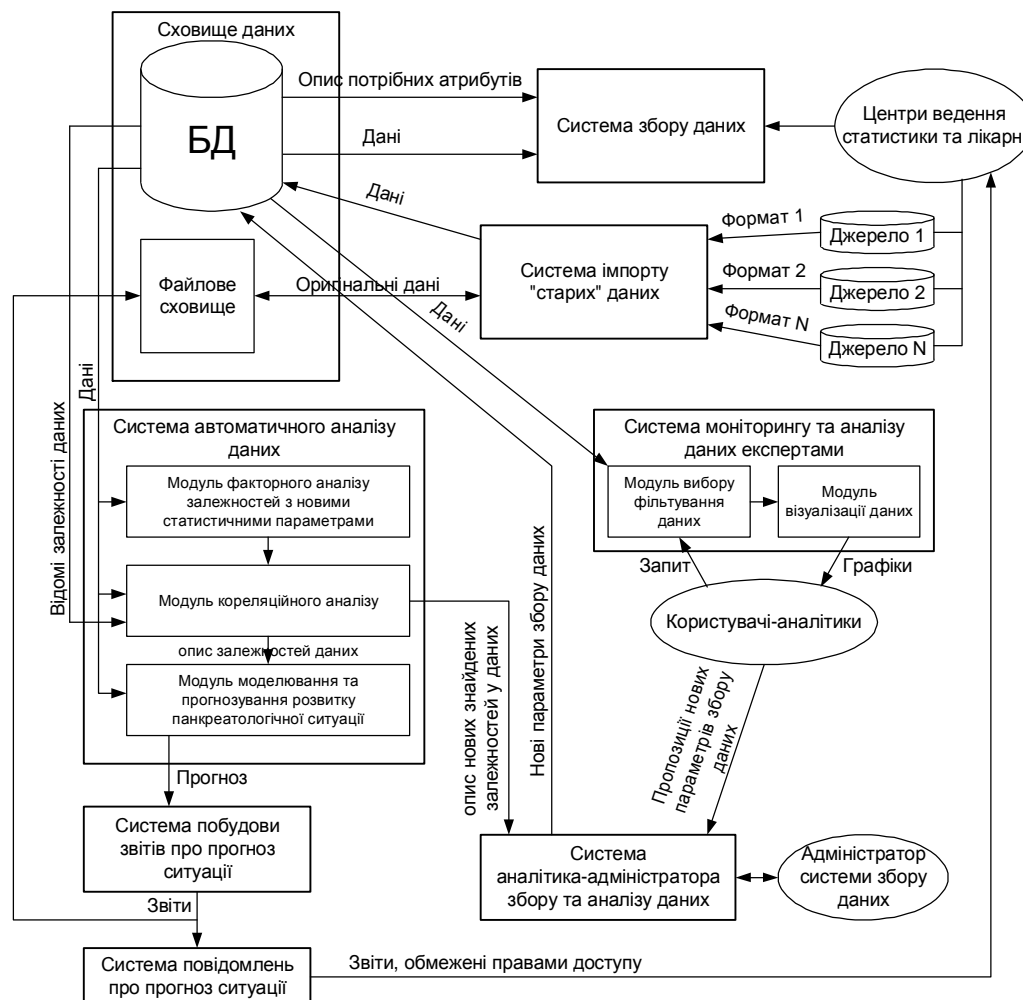


Рис. 2. Архітектура системи аналізу захворюваності на панкреатит

Сьогодні вже зібраний великий обсяг корисних даних по кількості хворих на панкреатит різних типів та форм, способах лікування та смертності у вигляді річних звітів. Ця інформація повинна бути імпортована в базу даних з рис. 1. Якщо звіти панкреатологічних центрів інших міст та регіонів мають додаткову інформацію, яку сьогодні система не використовує для аналізу, то усі ці документи повинні бути збережені у файловому сховищі для подальшого можливого використання.

Система збору даних покликана за заданим набором необхідних параметрів описання захворювання побудувати діалог введення даних користувачем – працівником відділу статистики лікарні та зберегти зібрану інформацію у БД.

Система моніторингу та аналізу даних експертами дає можливість вибрати будь-які розрізи статистики з БД та побудувати потрібні графіки. Виявляючи непередбачувані залежності, аналітики повинні повідомити адміністратора системи збору даних про знайдені залежності, а той – перевірити їх та за потреби додати параметри збору даних для можливості детальнішого дослідження впливу знайденого фактора. За подальших додавань записів про хворих працівники статистичних відділів лікарень повинні будуть ввести дані і за новим критерієм.

Дещо аналогічною до попередньої є підсистема автоматичного аналізу даних. Вона, використовуючи факторний аналіз, визначає залежності між новими зібраними наборами даних (які

ще не мають чітко дослідженого впливу на захворюваність панкреатитом). Потім знайдені нові залежності об'єднуються з відомими і передаються у модуль кореляційного аналізу та адміністратору системи збору даних. Модуль кореляційного аналізу призначений для визначення форми та сили залежностей і формування опису моделі розвитку панкреатологічної ситуації. За цією моделлю модуль прогнозування ситуації здійснює обчислення можливих варіантів розвитку стану захворюваності та їх імовірностей і формує прогноз розвитку панкреатологічної ситуації по регіонах, конкретних центрах, що надали дані для аналізу, та країні загалом. Роботу цієї підсистеми зображено на рис. 3.

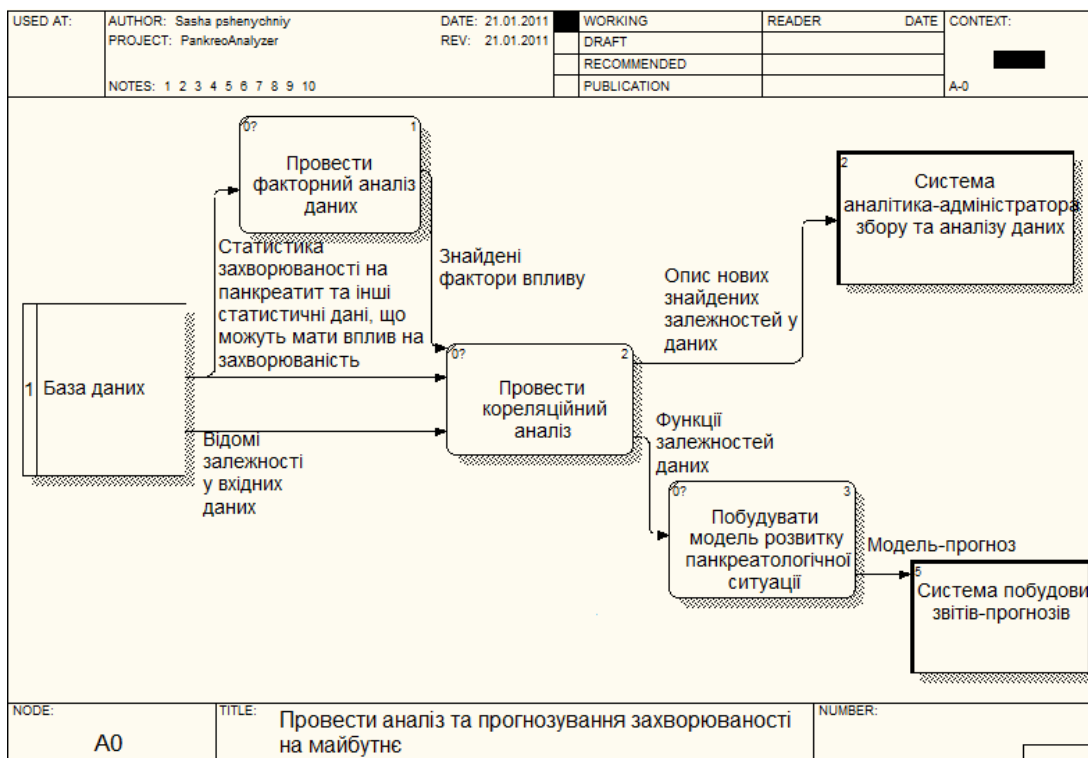


Рис. 3. DFD рівня 1 системи прогнозування захворюваності на панкреатит

Наступним етапом роботи системи є перетворення прогнозу, з яким працює програмна система, у прийнятний для людини вигляд. Цим займається підсистема побудови звітів про прогноз ситуації.

Сформовані звіти повинні бути розіслані у відповідні статистичні центри, центри прогнозування та лікарні, враховуючи їхні права доступу до даних.

Попередній аналіз даних

Насамперед шукалася залежність даних між кількістю пацієнтів та способами перебігу хвороби (рис. 4).

У результаті цього вдалося отримати, що кількість пацієнтів та оперативних втручань зросла у 2.3 раза, кількість смертей – у 1.5 раза. У 2010 р. спостерігалось значне зниження захворюваності порівняно з 2007 р., але відсоток смертності при цьому зріс.

Загалом смертність за досліджуваний період коливалась від 2.1 до 4.7 % з загальною тенденцією до зменшення. Але цей показник все ж підтверджує актуальність та необхідність проведення глибоких досліджень причин цього захворювання.

Експерти предметної галузі виділяють дві основні причини загострення панкреатиту: вживання алкоголю та жовчно-кам'яна хвороба, відповідно випадки захворювання поділяються на етанольні та біліарні і більшість наявної деталізованої статистики має цю градацію.

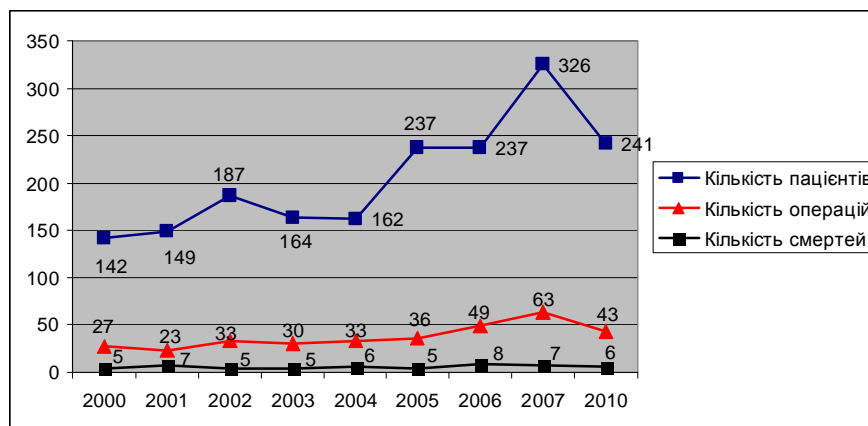


Рис. 4. Статистика захворювань на гострий панкреатит по м. Львову за даними Львівського міського панкреатологічного центру за 2000–2010 рр.

Далі визначалося, чи не збільшився відсоток смертності за останні роки (рис. 5).

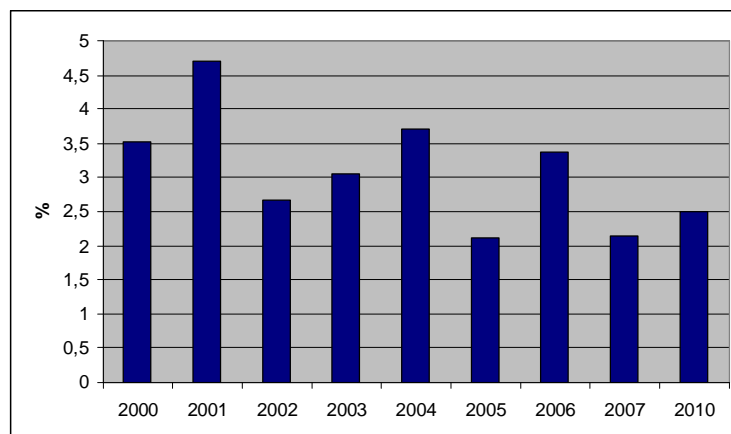


Рис. 5. Смертність хворих на панкреатит у м. Львові

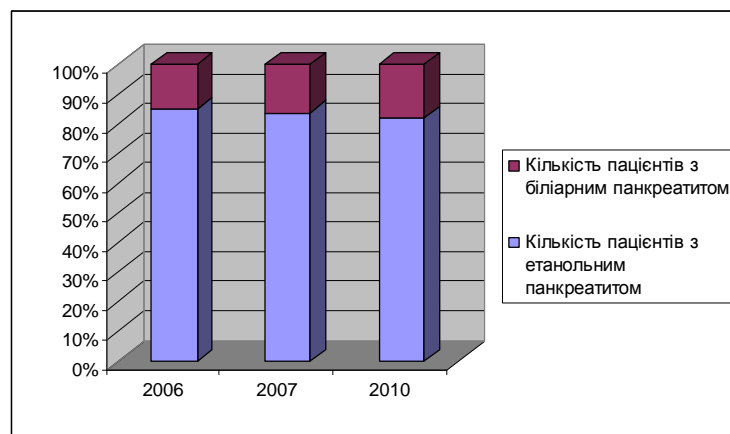


Рис. 6. Частки етанольний та біліарних панкреатитів у м. Львові за 2006–2010 рр.

Хоча кількість хворих на панкреатит змінюється доволі швидко, але частки етанольних та біліарних панкреатитів залишаються фактично незмінними і спостерігається тільки дуже повільне зменшення відсотка етанольних панкреатитів з 85 до 82 % за останні чотири роки.

Вдалося встановити залежність між статтю та типом захворюваності (близько 80 % хворих на етанольний панкреатит є чоловіками; щодо біліарної форми, то статевий розподіл протилежний – близько 80 % хворих є жінками).

Система аналізу даних, описана вище, передбачає адаптивну структуру бази даних з можливістю додавання критеріїв та градацій ведення статистики вже на етапі роботи. Це дасть змогу оперативно реагувати на виявлені нові залежності в даних та відразу розпочати збір необхідної інформації для детальнішого аналізу та уточнення форми знайденої залежності. Керувати процесом додавання критеріїв збору даних повинен адміністратор даних, який аналізує усі знайдені потенційні залежності (як автоматичним, так і ручним аналізом) та за потреби підтверджує введення нового атрибуту. Якщо дозволити додавання атрибутів напряму з різних підсистем, то це може призвести до дуже заплутаної структури таблиць бази даних, а також до їхнього надмірного розростання в ширину.

Оскільки деталізовані звіти про статистику захворювань вдалось отримати тільки за три роки, то дані доволі бідні, навіть для початкового аналізу. Можна констатувати факт різкого зростання кількості випадків усіх типів панкреатиту, окрім псевдокісти.

Визначно зміни часток хворих по формах панкреатиту (рис. 6).

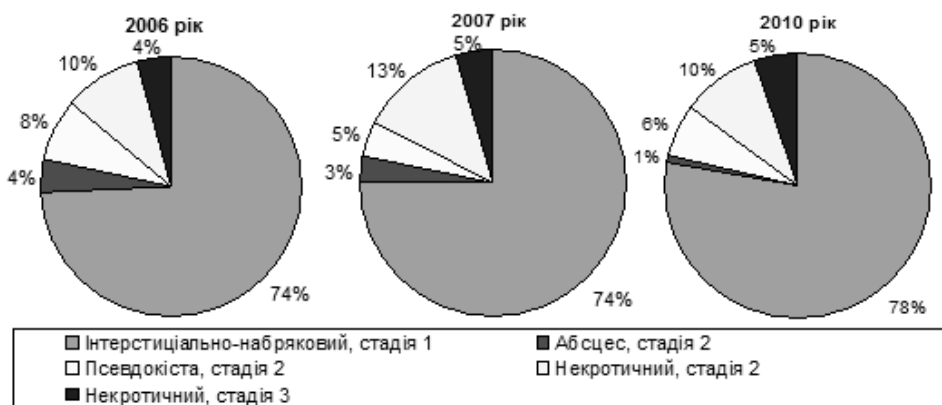


Рис. 6. Порівняння часток форм етанольного панкреатиту за останні роки по м. Львову

Провівши аналіз навіть стосовно невеликого обсягу даних, помітна їх різноманітність як в структурі та обмеженнях, так і в типах можливих залежностей. Багато параметрів є дискретними і їх комбінації утворюють групи, що мають значні кореляції даних всередині однієї групи, але зв'язки між даними різних груп доволі малі. Для розв'язання задачі пошуку залежностей у таких умовах пропонується застосовувати апарат грубих множин у підсистемі автоматичного аналізу даних.

Як бачимо, лікування абсцесу не змінилось і єдиним варіантом є операційне втручання.

Щодо найважчої форми захворювання – некротичного панкреатиту 3-ї стадії, то з'явилися безопераційні методи лікування. Щоправда, їх застосування доволі рідкісне (лише 10 % випадків).

Найбільші зміни прослідковуються у лікуванні псевдокісти – частка операцій у лікуванні зменшилась від 67 у 2007 р. до 17 % у 2010 р.

Єдиною формою захворювання, яка збільшила частку операційних хворих за досліджуваній період, є некротична 2-ї стадії.

Градація кількості смертей від форми захворювання та способу лікування (консервативне чи оперативне) не приводиться, оскільки фактично усі летальні випадки настали під час оперування хворих на некротичний панкреатит 3-ї стадії. Варто зазначити, що смертність цієї форми захворювання становила 58–75 % навіть за сучасного рівня розвитку медицини:

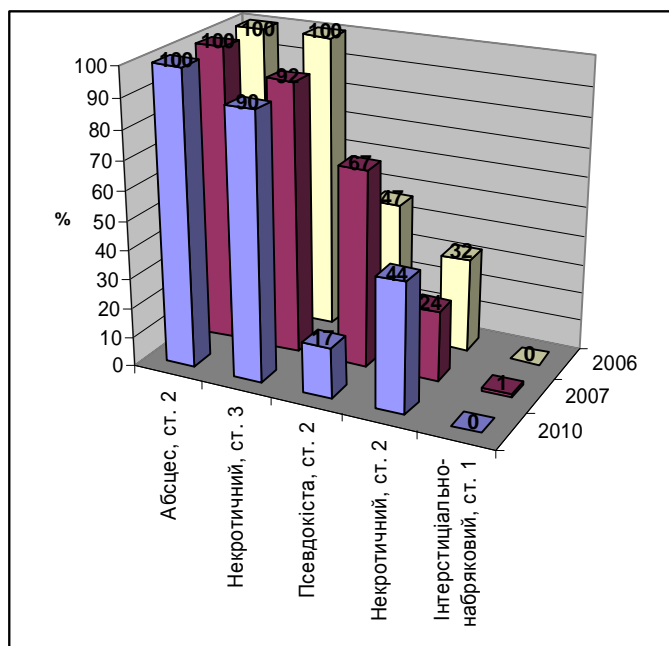


Рис. 7. Статистика частки операцій під час лікування залежно від форми етанольного панкреатиту по роках



Рис. 8. Смертність хворих на некротичний етанольний панкреатит 3-ї стадії залежно від способу лікування у м. Львові

Зіставивши дані з рис. 7 та 8, можна зробити висновок, що впровадження безопераційних методів лікування було неефективним – усі хворі померли. Щоправда, вибірка не є репрезентативною, оскільки містить лише по одному випадку за 2007 та 2010 рр. і жодного – за 2006 р. Можна також констатувати факт зменшення операційної смертності за некротичного панкреатиту 3-ї стадії.

З рис. 8 та 9 зрозуміло, що не усі залежності мають чітко виражений характер – деякі можуть бути просто винятками, спричиненими непередбачуваними обставинами. Виявлення таких значень у даних можливе після проведення кореляційного аналізу в спроектованій системі (рис. 3) та виявлення форми основної компоненти залежності. Після цього можливе додаткове вивчення знайдених відхилень експертами з використанням спеціалізованої підсистеми, або подальші ітераційні спроби знайти локальний вплив деяких параметрів середовища на ситуацію.

Аналіз захворюваності на біліарний панкреатит за 2000–2010 рр. у м. Львові

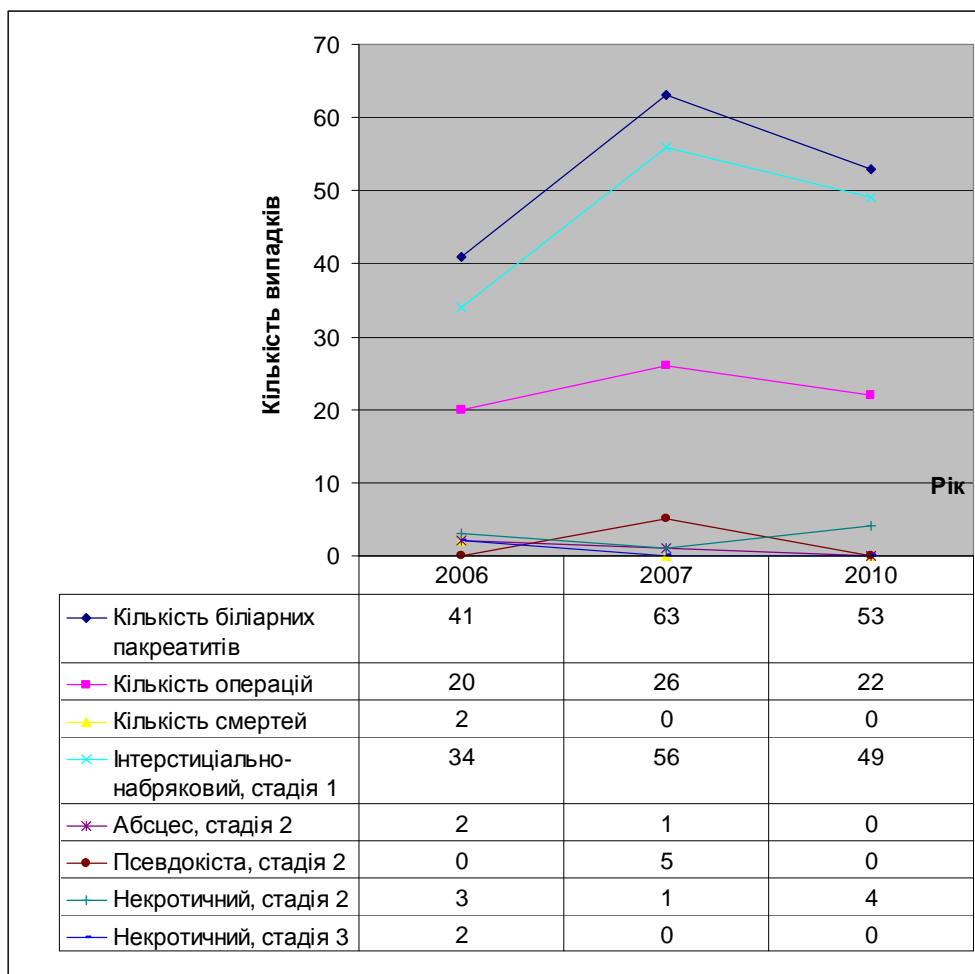


Рис. 9. Деталізована статистика по типах біліарних панкреатитів у м. Львові за даними Львівського міського панкреатологічного центру

Чинники можливого впливу на захворюваність панкреатитом

У ході проведення дослідження проводились консультації з експертом-хірургом Львівської міської клінічної лікарні швидкої допомоги, на основі програмно отриманих графіків та залежностей було висунуто багато гіпотез та напрямів, які необхідно вивчити у подальших дослідженнях. Наведемо деякі з них.

Кількість етанольних панкреатитів має періодичність на річній основі на відміну від біліарних.

Імовірність захворювання на етанольний панкреатит та його форма залежать від:

- 1) зайнятості населення;
- 2) статі;
- 3) віку;
- 4) професії людини;
- 5) рівня життя;
- 6) регіону та країни.

Імовірність захворювання на біліарний панкреатит та його форма залежать від:

- 1) статі;
- 2) віку;
- 3) рівня народжуваності.

Частка важких форм захворювання та смертність залежать від рівня довіри до медицини.

Використання спроектованої системи дасть змогу детально перевірити усі ці припущення, а також виявити нові залежності в даних як в глобальних, так і в певних межах зміни визначених атрибутів, якщо наявні локальні впливи чи значні зміни впливу одного (чи кількох) параметра моделі на інші.

Висновки

У цій роботі розроблено систему для аналізу захворюваності на панкреатит у м. Львові за 2000–2010 рр. Виявлено багато фактів, що не можуть бути пояснені у межах отриманих даних.

Подано тільки найважливіші факти, що були виявлені під час дослідження, і виявлення причин яких може допомогти у профілактиці захворювання. Тому усі вони потребують подальшого детального дослідження та пояснення.

Достовірно невідома кількість можливих впливів на виникнення захворювання, тому важливим завданням автоматизації аналізу цих даних є пошук недосліджених атрибутів, що впливають на статистику захворюваності на панкреатит. Завдяки модулю факторного аналізу спроектована система зможе виявляти невідомі раніше чинники впливу. Експерти предметної області отримуватимуть звіти про такі невідомі атрибути та зможуть запропонувати додавання нових критеріїв збору даних, а також вносити нові дані в базу, не обмежуючись фіксованою схемою. Зібрані дані мають різні типи (абсолютні значення, відсотки, множини фіксованих значень тощо), і аналіз повинен враховувати ці особливості, тому спроектована система не змушує навести усю вхідну інформацію в одній таблиці фіксованого вигляду в базі даних. Застосування апарата грубих множин дасть змогу виявити локальні залежності у певних межах зміни груп атрибутів.

У співпраці з експертом-хірургом Львівської міської клінічної лікарні швидкої допомоги було висунуто багато гіпотез та напрямів, які необхідно вивчити у подальших дослідженнях. Для вищої точності результатів необхідне створення системи детального моніторингу захворюваності з можливістю додавання потрібних параметрів та градацій ведення статистики.

Навіть початковий аналіз даних з одного джерела є доволі трудомістким завданням, а вивчення закономірностей виникнення захворювання у межах країни є фактично неможливим без застосування потужних обчислювальних систем, тому було прийнято рішення розробити таку систему аналізу даних.

Розроблено архітектуру системи моніторингу та аналізу даних по захворюваності на панкреатит, спроектовано та створено базу даних для цієї системи.

У майбутніх дослідженнях планується детально вивчити наведені у роботі результати та гіпотези, розробити систему моніторингу та аналізу даних по захворюваності на панкреатит та за її допомогою виявити чинники, що впливають на імовірність виникнення захворювання, і в такий спосіб допомогти медичним спеціалістам розробити заходи профілактики цього небезпечного захворювання.

1. Річні звіти Львівського міського панкреатологічного центру про статистику захворювань на гострий панкреатит за 2000–2010 рр. 2. Енциклопедія “Вікіпедія”. Панкреатит [електронний ресурс]. Режим доступу: <http://ru.wikipedia.org/wiki/%D0%9F%D0%B0%D0%BD%D0%BA%D1%80%D0%B5%D0%B0%D1%82%D0%B8%D1%82>. 3. Фаулер М. Архитектура корпоративных программных приложений / М. Фаулер. – СПб: Вильямс, 2007. – 544 с. 4. Поспелов Д.А. Нечеткие множества в моделях управления и искусственного интеллекта. – М., 1986. 5. Пинскер И.Ш. Поиск зависимости и оценка погрешности. – М.: Наука, 1985. – 148 с. 6. Шеффе Г. Дисперсионный анализ / пер. с англ. – М., 1963.